PADERBORN UNIVERSITY

First-order methods and gradient dynamical systems for multiobjective optimization

Von der Fakultät für
Elektrotechnik, Informatik und Mathematik
der Universität Paderborn
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
– Dr. rer. nat. –

genehmigte Dissertation von

Konstantin Heinrich Sonntag

Paderborn 2025

Gutachter: Prof. Dr. Sina Ober-Blöbaum
Prof. Dr. Sebastian Peitz (TU Dortmund)
Univ.-Prof. Dr. Radu Ioan Boţ (Universität Wien)

Tag der mündlichen Prüfung: 24.11.2025

Acknowledgements

I am sincerely grateful to all the people from both academic and personal circles whose continuous support, encouragement, and assistance were essential to the completion of this dissertation.

First and foremost, I wish to express my sincere gratitude to my supervisors, Prof. Dr. Michael Dellnitz and Prof. Dr. Sina Ober-Blöbaum. I am grateful to Prof. Dr. Sina Ober-Blöbaum for her valuable guidance throughout my PhD and her dedicated support in completing this thesis. I sincerely thank Prof. Dr. Michael Dellnitz for the opportunity to pursue this thesis within his research group at the Chair of Applied Mathematics and the freedom he gave me during my research.

I would like to extend my deepest thanks to Prof. Dr. Sebastian Peitz from TU Dortmund University, who acted as a dedicated mentor and supporter throughout my doctoral studies. His wide-ranging knowledge in applied mathematics, data science, and machine learning profoundly shaped my path. The successful cooperation and joint research projects we carried out together significantly contributed to my academic growth and development.

Special thanks go to Univ.-Prof. Dr. Radu Ioan Boţ from the University of Vienna for the ongoing exchange during my PhD research and his insightful advice. The foundation of my research was greatly influenced by my master's studies under his supervision, a period that continues to inspire this thesis. Moreover, I am especially grateful for the opportunity to conduct a research visit to Vienna, which led to a joint project that is part of this thesis.

I would like to express my gratitude to Prof. Dr. Stefan Volkwein from the University of Konstanz for the collaboration and for the enjoyable and productive research visit to Konstanz. In connection with this collaboration, I also wish to thank Dr. Bennet Gebken from TU Munich for sharing his insights into multiobjective and nonsmooth optimization, as well as Dr. Georg Müller for the valuable discussions on PDE-constrained optimization during my research visits to Heidelberg.

My appreciation also goes to my past and present colleagues. From the Institute of Mathematics in Paderborn I thank: Dr. Khaled Belgacem, Dr. Katharina Bieker, Dr. Bennet Gebken, Dr. Raphael Gerlach, Dr. Sören von der Gracht, Marianne Kalle, Dr. Lukas Lanza, Dr. Sofya Maslovskaya, Dr. Christian Offen, Karin Senske, and Dr. Boris Wembe. From the former Data Science for Engineering group, which moved to TU Dortmund University, I thank: Junaid Akhter, Augustina Amakor, Jannis Becktepe, Manuel Berkemeier, Septimus Boshoff, Paul Fährmann, Hans Harder, Sèdjro Hotegni, Elisabeth Schlatt, Jan Stenner, Michael Skowronek, Diana Voth, Stefan Werner and Dr. Christian Zagabe. The shared collaborations and conversations made the work not only productive but truly enjoyable.

Beyond academia, this thesis owes its existence to the constant support of my family and friends. In particular, I am grateful to my parents, Katharina and Joachim, who made this achievement possible in the first place, and to my brothers Alexander and Maximilian and my sister Anastasia for their constant encouragement. I would also like to thank my friends in Salzkotten and elsewhere for reminding me that there is a world beyond mathematics.

Finally, I am profoundly grateful to my wife, Louisa, for her unwavering support and everlasting patience throughout this journey.

Abstract

This dissertation contributes to the field of multiobjective optimization, with a focus on unconstrained problems formulated in a general Hilbert space setting under varying regularity assumptions on the objective functions.

For the class of multiobjective optimization problems with locally Lipschitz continuous objective functions, we define a multiobjective ε-subdifferential, which we analyze for the first time in the context of general Hilbert spaces. Building on these theoretical investigations, we present a descent method in which, at each iteration, a descent direction is determined via a numerical approximation of the multiobjective ε-subdifferential. To the best of our knowledge, this is the first method for infinite-dimensional, nonsmooth multiobjective optimization that does not require either a prior discretization of the infinite-dimensional Hilbert space or a scalarization of the objective functions.

In the setting of convex, continuously differentiable objective functions with Lipschitz continuous gradients, we introduce a family of inertial gradient dynamical systems that generalize well-known continuous-time systems from scalar optimization. This work affirms the feasibility of extending such dynamics to the multiobjective setting. We present three novel systems: one with constant damping, one with asymptotic vanishing damping, and one combining vanishing damping with time-dependent Tikhonov regularization. For each, we establish improved convergence results that align with the best-known rates in the scalar case. Notably, our analysis employs merit functions, which are commonly used in multiobjective optimization, for the first time in the study of continuous-time systems.

Building on the investigation of the novel gradient dynamical systems, we develop an accelerated gradient method for multiobjective optimization via discretization of the multiobjective gradient system with asymptotic vanishing damping. The proposed method retains the favorable convergence properties of the continuous system while achieving faster convergence than standard approaches, such as the classical multiobjective steepest descent method. In the scalar case, our algorithm recovers Nesterov's accelerated gradient method, and we observe convergence rates consistent with known results. These findings highlight the potential of gradient dynamical systems to derive efficient gradient methods for multiobjective optimization.

Zusammenfassung

Diese Dissertation enthält Beiträge zum Bereich der Mehrzieloptimierung mit einem Fokus auf unbeschränkten Problemen, die auf einem allgemeinen Hilbertraum definiert sind, unter verschiedenen Regularitätsannahmen an die Zielfunktionen.

Für die Klasse der Mehrzieloptimierungsprobleme mit lokal Lipschitz-stetigen Zielfunktionen definieren wir zunächst ein multikriterielles ε-Subdifferential, das wir erstmals im Kontext allgemeiner Hilberträume analysieren. Aufbauend auf diesen theoretischen Untersuchungen präsentieren wir ein Abstiegsverfahren, bei welchem in jeder Iteration eine Abstiegsrichtung mittels einer numerischen Approximation des multikriteriellen ε-Subdifferentials bestimmt wird. Nach unserem Kenntnisstand handelt es sich dabei um das erste Verfahren für unendlichdimensionale, nichtglatte Mehrzieloptimierungsprobleme, das sowohl ohne vorherige Diskretisierung des unendlichdimensionalen Hilbertraums als auch ohne Skalarisierung der Zielfunktionen auskommt.

Im Kontext konvexer, stetig differenzierbarer Zielfunktionen mit Lipschitz-stetigen Gradienten, führen wir eine Familie von dynamischen Gradientensystemen mit Trägheitsterm ein, die bekannte kontinuierliche Systeme aus der skalaren Optimierung verallgemeinern. Diese Arbeit zeigt, dass solche Systeme in den multikriteriellen Fall übertragen werden können. Wir stellen drei neue Systeme vor: eines mit konstanter Dämpfung, eines mit asymptotisch abnehmender Dämpfung und eines, das zusätzlich eine zeitabhängige Tikhonov-Regularisierung beinhaltet. Für jedes dieser Systeme zeigen wir verbesserte Konvergenzeigenschaften, die mit den besten bekannten Raten im skalaren Fall übereinstimmen. Bemerkenswert ist, dass in unserer Analyse erstmals sogenannte Meritfunktionen, die in der Mehrzieloptimierung häufig verwendet werden, im Kontext kontinuierlicher dynamischer Systeme eingesetzt werden.

Aufbauend auf den Untersuchungen der neuen dynamischen Gradientensysteme, entwickeln wir ein beschleunigtes Gradientenverfahren zur Mehrzieloptimierung, das auf einer Diskretisierung des multikriteriellen Gradientensystems mit asymptotisch abnehmender Dämpfung beruht. Das hergeleitete Verfahren bewahrt die günstigen Konvergenzeigenschaften des kontinuierlichen Systems und erreicht gleichzeitig eine schnellere Konvergenz als klassische Ansätze wie das Verfahren des steilsten Abstiegs zur Mehrzieloptimierung. Im skalaren Fall entspricht unser Verfahren dem bekannten beschleunigten Gradientenverfahren von Nesterov, wobei wir übereinstimmende Konvergenzraten nachweisen. Diese Ergebnisse unterstreichen das Potenzial gradientenbasierter dynamischer Systeme zur Herleitung effizienter Gradientenverfahren zur Mehrzieloptimierung.

Contents

Chapter 1

Introduction

In real-world problems, both in everyday life and technical applications, there is often more than one objective to consider. For example, when purchasing a residential home, you typically want it to have the right size and perhaps a large garden. It should be located in a beautiful city, be well situated, ideally close to your workplace as well as near family and friends. Additionally, the house should not be too old, should have good energy efficiency, and be as affordable as possible. In this example, it is clear that no house will perfectly satisfy all these objectives simultaneously.

Similar challenges arise in technical domains. In optimal transport, for example, a goal is to allocate resources from suppliers to demanders in a manner that is not only fast and cost-effective, but also robust to disruptions and sustainable in the long term. In machine learning applications, a key objective is to design intelligent agents that solve specific tasks with high performance, while also being data-efficient, energy-efficient, resource-efficient, environmentally sustainable, robust, and secure. The presence of multiple often conflicting objectives necessitates a careful balancing act to achieve the best possible overall outcome by means of a suitable compromise.

These introductory examples clearly demonstrate that in the presence of multiple objectives, identifying a single optimal solution is inherently difficult. This highlights the importance of investigating methods that can yield effective solutions nonetheless. Many problems in technical domains can be modeled mathematically and formulated as optimization problems. In such cases, objectives are expressed as functions to be minimized, such as minimizing total cost or processing time. This reformulation allows the application of algorithms specifically developed for function minimization. Classic optimization theory primarily focuses on optimizing a single objective function, possibly subject to constraints defined by additional functions or geometric conditions that restrict the feasible solution space. As the earlier examples illustrate, this approach is limited and often inadequate for problems that naturally involve multiple, conflicting objectives. Multiobjective optimization offers a structured framework to overcome these limitations.

In multiobjective optimization, we seek to simultaneously minimize multiple objective functions. As the introductory example illustrates, these objectives are often mutually conflicting. This shifts the focus from identifying a single optimal solution to determining a set of trade-off solutions, each representing a different balance among the competing objectives. For instance, returning to the introductory problem of purchasing a residental house, consider two houses, one being less expensive, the other more energy efficient. There is no a priori way to determine which is superior. The choice depends on the preferences of a decision-maker. However, a rational decision-maker would not choose a house that is both more expensive and less energy efficient. This concept of preference is formally captured by the notion of Pareto optimality, attributed to Pareto [189]. In multiobjective settings, a solution is considered Pareto optimal if no other solution exists that is at least as good in all objectives and strictly better in at least one. This is a fundamental concept used to model decision-making and rational choice behavior.

Optimization, also known as mathematical programming or operations research, is one of the most prominent branches in applied mathematics and is widely used across an extensive variety of fields to solve an ever-growing number of problems. As a result, there is a non-ending thrive to develop stronger optimization methods, being straightforward to implement, fast and resource efficient. Unfortunately, we cannot apply optimization theory to solve all real-world problems, but only those for which we can construct a reasonable mathematical model that fits within the optimization framework. In such cases, we aim to minimize a mathematical objective function, possibly subject to additional constraints, which can be represented by constraint functions or defined through geometric conditions. In this thesis, we use mathematical optimization techniques to solve multiobjective optimization problems.

As described above, mathematical structure is essential to solve a problem using optimization techniques. In this thesis, we work in the context of Hilbert spaces and consider objective functions under different regularity assumptions. Choosing Hilbert spaces enables the inclusion of a wide range of classical problems, typically stated in Euclidean spaces, as well as many applications from optimal control, inverse problems and PDE-constrained optimization. Additionally, we need to impose some structure on the objective functions. Without any structure, an optimization problem is generally unsolvable. Imposing structure on the problem requires a trade-off. While assuming more structure allows for the design of efficient optimization techniques, it simultaneously restricts applicability to problems that satisfy the imposed assumptions. In this thesis we consider the two following problem classes. We solve infinite-dimensional, unconstrained multiobjective optimization, with objective functions fulfilling one of the following assumptions:

- The objective functions are nonconvex and locally Lipschitz continuous;
- The objective functions are convex and continuously differentiable with Lipschitz continuous gradients.

These two problem classes differ significantly, and their analysis highlights the previously discussed trade-off between generality and structure. The first class is highly general. In continuous optimization, it is rarely possible to further relax this assumption on the objective functions. Conversely, the second class is more restrictive, yet it encompasses many important applications. Tackling simpler problems first is crucial, as it enables a step-by-step increase in complexity while refining and enhancing the methods. This progression also illustrates the benefits of additional structure. The method we propose for the first class guarantees convergence to points satisfying necessary optimality conditions. In contrast, the methods developed for the second class exhibit stronger theoretical properties, including improved convergence rates and quantitative complexity bounds, which provide insight into how computational effort scales with the desired solution accuracy.

Following this general introduction on multiobjective optimization we want to explain the title of this thesis to the reader. In the following, we explain what we understand under first-order methods and gradient dynamical systems for multiobjective optimization, respectively.

First-order methods for multiobjective optimization

The solution methods investigated in this thesis are iterative methods, i.e., starting from an initial point these methods generate a sequence of iterates that converges to an optimal solution or to a point that satisfies a necessary optimality condition. An optimization method is called a first-order method, if it relies only on problem information in terms of objective function and gradient evaluations. Hence, the method has only access to local information of the problem. When the functions under consideration are nonsmooth and classical gradients do not exist, we allow appropriate generalizations of the gradients, typically in the form of subgradients. The methods we study, update the current iterate by computing first an appropriate search direction and then a suitable step length.

Gradient dynamical systems for multiobjective optimization

Gradient dynamical systems, related to optimization problems, are systems that depend on the gradient information of the objective function. When appropriately designed, these systems exhibit favorable properties in the context of optimization and are closely connected to first-order methods: the continuous limit of a first-order method often corresponds to a gradient dynamical system, and discretizations of gradient dynamical systems yield first-order methods. These connections often lead to shared asymptotic behavior between the continuous and discrete formulations. The analysis of continuous systems is often more tractable, as differentiation and integration can be applied directly, whereas discrete systems typically require more involved arguments. Therefore, it can be beneficial to begin by studying the gradient dynamical system and then transfer the obtained results to the analysis of the corresponding discrete method.

The results of this thesis are presented in two independent parts. The first part is formed by Chapter 3 which deals with nonconvex locally Lipschitz continuous multiobjective optimization problems. The second parts encompasses Chapters 4 and 5 and is concerned with convex and smooth multiobjective optimization. Chapter 4 provides a discussion of gradient dynamical systems, and Chapter 5 introduces a first-order method for multiobjective optimization. In the following, we present an overview of the development of first-order methods and corresponding gradient dynamical systems in multiobjective optimization, starting with the case of scalar optimization.

The simplest gradient dynamical system associated with an optimization problem with a scalar objective function is the steepest descent dynamical system or (sub)gradient flow, which was studied by Bruck [56] in the Hilbert space setting. Using techniques from monotone operator theory it can be shown that trajectories of the steepest descent dynamical system converge weakly to solutions of the optimization problem. In the smooth case, an explicit discretization of this system leads to the steepest descent method which dates back to Cauchy [62] and is analyzed more generally in the context of optimization by Polyak [199]. In the nonsmooth case, the same discretization yields the subgradient method which is attributed to Shor [198]. An implicit discretization of the steepest descent dynamical system gives rise to proximal point methods. The proximal point method was introduced by Martinet [165] and later analyzed by Rockafellar [203] and Güler [119]. Combinations of gradient and proximal point methods lead to proximal gradient methods. These methods belong to the broader class of splitting methods described by Lions & Mercier [152]. They can be obtained from semi-implicit discretizations of the subgradient flow for structured optimization problems.

The described methods have the advantage that they are highly general and can readily be applied to a large class of optimization problems. They are backed by a mature theory, have convergence guarantees and there exist complexity bounds that describe the numerical effort required to obtain an approximate solution. On the downside, the described methods can suffer from slow convergence, especially for ill-conditioned problems. For example, when we consider the steepest descent method for smooth convex optimization problems with a constant step size, it is known that the function values converge to the optimal function value at a sublinear rate in the general case. For smooth and strongly convex problems the iterates converge to the optimal solution with a linear rate that depends on the condition number of the problem. In the mathematical optimization literature, numerous methods have been proposed to improve upon the plain steepest descent method. The most prominent adaptions are linear and nonlinear conjugate gradient methods [99, 130, 195], gradient methods using more sophisticated step size rules, like exact line search [62], backtracking line search and Armijo conditions [8], Wolfe conditions [183, 236], Barzilai-Borwein step size rules [30] or trust region approaches [77, 175]. An alternative class of optimization methods are higher-order methods like Newton or quasi-Newton methods [55, 100, 112, 184, 211]. The initial ideas of the listed approaches are contained in most text books on nonlinear optimization [38, 101, 155, 178, 180, 184, 197].

In this thesis, we follow a different strategy to accelerate the convergence of first-order methods. A general way to speed up the convergence of an iterative scheme is proposed by Polyak in [196]. This approach involves introducing a momentum term that utilizes information from previous iterates to accelerate the convergence. Simultaneously, in [196] an analogous gradient dynamical system with an inertial term is proposed. The inertial gradient dynamical system introduced in this work is further analyzed in the context of optimization by different authors. In [19], Attouch, Goudou & Redont derive the intertial gradient system by modeling a ball rolling down the graph of a function and therefore denote it by the heavy ball with friction dynamical system. A more efficient way to accelerate the steepest descent method is proposed by Nesterov [182]. In this paper a method with a non-constant momentum parameter is proposed to derive improved convergence rates for solving smooth convex optimization problems. These ideas where adapted to accelerate the proximal point method by Güler [118] and to accelerate proximal gradient methods by Beck & Teboulle [33]. In [218], Su, Boyd & Candès derive a gradient dynamical system with asymptotic vanishing damping which is the continuous counterpart to the accelerated gradient method, and which shares the improved asymptotical convergence properties. The strong connection between gradient dynamical system and optimization methods sparked active research on accelerated gradient methods, proximal point methods and more general splitting schemes. By now, there is a rapidly growing literature investigating these ideas in the context of optimization and related applications, like constrained optimization by means of primal-dual dynamical systems [48, 125, 124, 241], min-max problems [64, 121, 46], monotone inclusions [4, 5, 238, 60, 22] and variation inequalities [209, 45]. Additionally, researchers strive to derive dynamical systems and optimization methods with better properties by using higher order information in form of Hessian-driven damping to obtain Newton-like methods [6, 25], Tikhonov regularization to obtain better convergence properties in infinite-dimensional spaces [134, 144, 145, 146, 147] and time-scaling to enforce even faster convergence rates [20, 47].

In multiobjective optimization, the connection between dynamical systems and optimization algorithms remains significantly underexploited. Although, foundational contributions on gradient dynamical systems in the multiobjective setting exist, the field is still developing. One of the pioneering works in this area is due to Smale [214], who employed smooth continuous-time trajectories exhibiting a common descent property with respect to all objective functions to characterize Pareto critical points. This seminal research was motivated by models in full exchange economies and grounded in the mathematical frameworks of global analysis and Morse theory. The first generalization of the steepest descent dynamical system for multiobjective optimization was proposed by Henry [127] in the context of economics and investigated by Cornet [78, 79] in resource allocation problems. In the context of multiobjective optimization, the multiobjective steepest descent system is analyzed by Schäffler, Schultz & Weinzierl [207] and by Miglerina [170], where it is shown that cluster points of the trajectory satisfy a necessary optimality condition. In the context of convex, infinite-dimensional problems the multiobjective steepest descent dynamical system is further examined by Attouch & Goudou [18]. Generalization to the nonsmooth case were obtained by Attouch, Garrigos & Goudou [17]. The first inertial multiobjective gradient system in the spirit of the heavy ball with friction dynamical system is proposed by Attouch & Garrigos [16] which is also part of the PhD thesis by Garrigos [105]. Using a constant damping parameter they propose a system for which convergence of trajectories to Pareto optimal points can be shown. However, it is not clear if this system improves the multiobjective steepest descent system or if a first-order method with theoretical convergence guarantees can be derived from this system. In particular, Attouch & Garrigos [16] identified the challenge of incorporating time-dependent damping into inertial multiobjective gradient systems and deriving accelerated multiobjective gradient methods from such continuous-time dynamics.

The multiobjective steepest descent method was initially introduced by Mukai [177] and independently by Fliege & Svaiter [104]. They define a multiobjective steepest descent direction incorporating gradient information of all objective functions, simultaneously. Analogous to developments in scalar optimization, numerous adaptations of the multiobjective steepest descent method have been proposed, aiming to enhance its algorithmic performance and convergence properties. The first attempt to incorporate acceleration into the multiobjective steepest descent framework, inspired by Nesterov's accelerated gradient method [182], was made by El Moudden & El Moutasim [94]. Their approach demonstrates improved convergence rates, although only under restrictive assumptions. Notably, their proof primarily consists of applying Nesterov's original accelerated gradient method to a weighted sum scalarization of the multiobjective problem, rather than directly addressing the multicriterial nature of the problem. A significant advancement was the introduction of the first proximal gradient method for multiobjective optimization that does not rely on an a priori scalarization of the objectives. This method was proposed by Tanabe, Fukuda & Yamashita [225] and further analyzed in their subsequent works [223]. This method is not motivated from a gradient dynamical system but derived from the concept of merit functions. In multiobjective optimization merit functions are functions that use objective functions values to characterize Pareto optimality. In general, for a problem a merit function is a function that is positive everywhere and vanishes only at a solution. In multiobjective optimization, merit functions where first introduced for convex problems with linear constraints by Chen [66] and later investigated in more general settings [89, 153, 220]. Using this concept Fukuda, Yamashita & Tanabe propose a method for multiobjective optimization which generalizes the accelerated gradient method from scalar optimization [220, 221, 222].

In this thesis, we address the gap between inertial gradient dynamical systems and accelerated first-order methods in multiobjective optimization. On the one hand we positively answer the question proposed in [16] whether it is possible to define fast gradient dynamical systems for multiobjective optimization, by presenting multiple novel gradient dynamical systems. On the other hand, we show that our approach is suitable to derive accelerated gradient methods for multiobjective optimization and we point out the relation to recently discovered fast gradient methods. The analysis of the proposed systems and methods uses the concept of merit functions which was not applied for gradient dynamical systems so far. Furthermore, we show that our approach is strong enough to generalize to more involved gradient dynamical systems. We are convinced that this approach constitutes a foundational step toward designing novel, efficient algorithms capable of addressing multiobjective optimization problems with increasingly intricate structures. We conclude the introduction with the outline of this thesis.

In Chapter 2, we present the theoretical background of this thesis. We introduce the most important concepts from functional analysis, such as Hilbert spaces, essential concepts from convex analysis, various notions of the derivative and a subdifferential which generalizes the derivative to nonsmooth functions. Furthermore, we provide an overview on differential equations and inclusions, summarizing key existence results as well as important differential and integral inequalities. The final part of Chapter 2 is reserved for the introduction of the multiobjective optimization problem. We formally define Pareto optimal points and describe necessary optimality conditions. Additionally, we introduce a merit function which is an important measure for optimality in multiobjective optimization. Finally, we discuss the multiobjective steepest descent method which serves as a starting point for the more elaborate first-order methods and gradient dynamical systems examined in the following chapters.

Chapter 3 is dedicated to a descent method for multiobjective optimization with nonconvex locally Lipschitz continuous objective functions. This method is based on a gradient sampling scheme and is the first to address nonsmooth, infinite-dimensional multiobjective optimization problems, without discretizing the infinite dimensional space or scalarizing the multiple objectives beforehand. Before defining our method, we introduce a generalization of the Goldstein ε-subdifferential to multiobjective optimization problems. We investigate the main theoretical properties of the multiobjective ε-subdifferential, which are important for deriving necessary optimality conditions and for proving convergence of the proposed algorithm. Prior to formulating the descent method, we describe how to numerically approximate the multiobjective ε-subdifferential and show that we can obtain a suitable descent direction from this approximation. Subsequently, we use this direction to define a descent method using a backtracking line search. To validate the applicability of the proposed method, we apply it to a multiobjective optimal control problem and demonstrate its capability to efficiently compute Pareto optimal solutions.

In Chapter 4, we present gradient dynamical systems associated with smooth convex multiobjective optimization problems. An introductory section motivates continuous-time approaches in scalar optimization and outlines the general approach of the analysis. The initial treatment of scalar optimization problems allows to highlight difficulties arising when shifting to the multiobjective setting. Then, we review existing gradient systems for multiobjective optimization. An extensive discussion of the multiobjective steepest descent dynamical system serves as a starting point for the more advanced systems presented later. Relevant results from the literature are also thoroughly presented. In the main part of this chapter, we present a total of three dynamical gradient systems for multiobjective optimization. Before introducing these systems, an existence result for a generalized differential inclusion is stated, which will be used to prove existence of solutions for the novel systems. In Section 4.4, we introduce an inertial multiobjective gradient system and show that trajectories of this system converge weakly to weakly Pareto optimal points. This system is improved in Section 4.5 by including asymptotically vanishing damping. It is proven that this yields fast convergence rates for the function values while trajectories achieve weak convergence to weakly Pareto optimal points. Numerical experiments verify the theoretical convergence rates. Further improvements are made in Section 4.6 by including vanishing Tikhonov regularization. First, we generalize Tikhonov regularization for multiobjective optimization. This extension yields strong convergence to Pareto optimal points satisfying a minimum norm property. Finally, we show that strong convergence is indeed obtained and verify the theoretical findings by numerical experiments.

An accelerated gradient method for convex smooth multiobjective optimization is developed in Chapter 5. To provide a solid foundation, we begin with a concise overview of Nesterov's accelerated gradient method for scalar optimization. Building on this, our proposed method is rigorously derived through a discretization of the multiobjective gradient system with asymptotically vanishing damping, introduced in Chapter 4. Moreover, we discuss its relation to other existing first-order methods in multiobjective optimization. The main theoretical contributions are presented in the section dedicated to the asymptotic analysis of the algorithm. Here, we establish fast convergence rates for the merit function values and prove weak convergence of the generated iterates to weakly Pareto optimal points. Notably, our convergence guarantees align with the optimal rates known from scalar optimization theory. To complement the theoretical developments, the chapter concludes with multiple numerical experiments. These include both finite-dimensional problems and an infinite-dimensional problem in a Hilbert space framework, illustrating the broad applicability of our approach. Overall, our findings demonstrate that fast multiobjective optimization methods can be systematically derived from gradient dynamical systems, validating the foundational perspective introduced earlier.

Chapter 6 presents the conclusion of this thesis by summarizing the main findings, discussing their implications, and highlighting open questions for future research.

Previous publications

The content of this thesis is based on the following publications. References to these publications are provided at the beginning of each chapter and section where the corresponding results are presented.

[49] Boţ, R. I. and Sonntag, K. *Inertial dynamics with vanishing Tikhonov regularization for multiobjective optimization.* In: *Journal of Mathematical Analysis and Applications* 554 (2) (2025). doi: 10.1016/j.jmaa.2025.129940.

[215] Sonntag, K., Gebken, B., Müller, G., Peitz, S., and Volkwein, S. *A descent method for nonsmooth multiobjective optimization in Hilbert spaces.* In: *Journal of Optimization Theory and Applications* 203 (1) (2024), pp. 455–487. doi: 10.1007/s10957-024-02520-4.

[216] Sonntag, K. and Peitz, S. *Fast convergence of inertial multiobjective gradient-like systems with asymptotic vanishing damping.* In: *SIAM Journal on Optimization* 34 (3) (2024), pp. 2259–2286. doi: 10.1137/23M1588512.

[217] Sonntag, K. and Peitz, S. *Fast Multiobjective Gradient Methods with Nesterov Acceleration via Inertial Gradient-Like Systems.* In: *Journal of Optimization Theory and Applications* 201 (2024), pp. 539–582. doi: 10.1007/s10957-024-02389-3.

Chapter 2

Theoretical background

In this chapter, we present the theoretical background that forms the foundation of this thesis. In each section, we provide relevant literature to put our analysis into context, and we cite the most significant results that will be used later.

In Section 2.1, we introduce various topics related to infinite-dimensional analysis. Since this is a broad field, we do not attempt to provide a comprehensive introduction to all underlying concepts. Instead, we establish some essential notation and clarify the most important notions to our work. We introduce general Hilbert spaces and their topological dual spaces. Additionally, we introduce important concepts from convex analysis. We examine differentiability of functions defined on a Hilbert space and discuss extensions of the derivative for nonsmooth functions.

Section 2.2 is dedicated to differential equations and inclusions. We review the most important existence results for differential equations. Furthermore, we introduce key elements of set-valued analysis and state existence results for differential inclusions. The final part of this section covers essential differential and integral inequalities that will be applied in the analysis of certain dynamical systems.

In Section 2.3, we introduce the multiobjective optimization problem which is central to this thesis. We provide a rigorous definition of Pareto optimal points for a general multiobjective optimization problem and discuss necessary optimality conditions for both smooth and nonsmooth objective functions. In preparation for the asymptotic analysis of gradient dynamics and first-order methods, we introduce the concept of merit functions, which are defined using a suitable scalarization of the objective functions to quantify optimality. We conclude this section with the introduction of the multiobjective steepest descent method, which we later improve in subsequent parts of this thesis.

2.1 Functional analysis

2.1.1 Hilbert spaces

In this subsection, we introduce the basic concepts of Hilbert spaces, focusing on the notation for inner products, norms and the topological dual space. Additionally, we present two variants of Opial's Lemma which is an important tool for proving weak convergence of sequences in infinite-dimensional spaces. We do not present all results from Hilbert space theory used throughout this thesis. For a comprehensive treatment of the theory, we refer to [50, 206, 239].

In this thesis, $\mathcal{H}$ is a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$.

$\mathcal{H}$ is equipped with the strong topology induced by the norm $\|\cdot\|$. The topological dual of $\mathcal{H}$ is denoted by $\mathcal{H}^* := \{x^* : \mathcal{H} \to \mathbb{R} \ : \ x^* \text{ is linear and continuous}\}$. The dual space $\mathcal{H}^*$ together with the dual norm $\|x^*\|_* := \sup_{x \in \mathcal{H} \setminus \{0\}} \frac{|x^*(x)|}{\|x\|}$, for $x^* \in \mathcal{H}^*$, forms a Banach space. There exists a natural embedding of $\mathcal{H}$ into $\mathcal{H}^*$ given by $\mathcal{R} : \mathcal{H} \to \mathcal{H}^*$, $x \mapsto \langle x, \cdot \rangle$, where $\mathcal{R}(x)(y) := \langle x, y \rangle$ for all $x, y \in \mathcal{H}$. This embedding is linear and bounded as a straightforward computation shows, i.e., $\|\mathcal{R}\| := \sup_{x \in \mathcal{H} \setminus \{0\}} \frac{\|\mathcal{R}x\|_*}{\|x\|} \leq 1$. The relation between $\mathcal{H}$ and $\mathcal{H}^*$ is actually stronger. This observation is described by the Riesz–Fréchet representation Theorem which we recite in the following [32, Fact 2.24].

Theorem 2.1.1. *Let $x^* \in \mathcal{H}^*$. Then there exists a unique vector $x \in \mathcal{H}$ such that for all $y \in \mathcal{H}$ it holds that $x^*(y) = \langle x, y \rangle$. Moreover, $\|x^*\|_* = \|x\|$.*

This result has strong implications. From Theorem 2.1.1 it follows that $\mathcal{R} : \mathcal{H} \to \mathcal{H}^*$ is not just an embedding but in fact an isometric isomorphism, which we call the *Riesz operator*. Using the inverse $\mathcal{R}^{-1}$ of the Riesz operator, we can define the following inner product on $\mathcal{H}^*$. Define $\langle \cdot, \cdot \rangle_* : \mathcal{H}^* \times \mathcal{H}^* \to \mathbb{R}$, $(x^*, y^*) \mapsto \langle x^*, y^* \rangle_* := \langle \mathcal{R}^{-1}(x^*), \mathcal{R}^{-1}(y^*) \rangle$. For all $x^* \in \mathcal{H}^*$, it holds that $\sqrt{\langle x^*, x^* \rangle_*} = \sqrt{\langle \mathcal{R}^{-1}(x^*), \mathcal{R}^{-1}(x^*) \rangle} = \|\mathcal{R}^{-1}(x^*)\| = \|x^*\|_*$, and in fact, $\mathcal{H}^*$ together with the inner product $\langle \cdot, \cdot \rangle_*$ forms a Hilbert space with induced norm $\|\cdot\|_*$.

Because of the strong relation between a Hilbert space $\mathcal{H}$ and its topological dual $\mathcal{H}^*$, one often identifies $\mathcal{H}^*$ with $\mathcal{H}$ without making a distinction between these spaces in notation. In this thesis, we choose to differentiate between $\mathcal{H}$ and $\mathcal{H}^*$ and make use of $\mathcal{R}$ in Chapter 3. For the implementation of the algorithm, we develop in this chapter, it is beneficial to distinguish between $\mathcal{H}$ and $\mathcal{H}^*$. In the remaining parts of the thesis, we do not work with the dual space $\mathcal{H}^*$ explicitly.

In the following, we introduce the notion of strong and weak convergence in $\mathcal{H}$. Let $(x^k)_{k \geq 0} \subset \mathcal{H}$ be a sequence and let $x^\infty \in \mathcal{H}$. If $\lim_{k \to +\infty} \|x^k - x^\infty\| = 0$, we say that x^k converges (strongly) to x^∞ and we write $x^k \to x^*$ as $k \to +\infty$. Similarly for a function $x : [t_0, +\infty) \to \mathcal{H}$, $t \mapsto x(t)$ and $x^\infty \in \mathcal{H}$, if $\lim_{t \to +\infty} \|x(t) - x^\infty\|$ we say $x(\cdot)$ converges (strongly) to x^∞ and denote this by $x(t) \to x^\infty$ as $t \to +\infty$. If a sequence $(x^k)_{k \geq 0}$ and an element $x^\infty \in \mathcal{H}$ satisfy $\lim_{k \to +\infty} \langle x^k - x^\infty, y \rangle \to 0$, for all $y \in \mathcal{H}$, or equivalently $\lim_{k \to +\infty} x^*(x^k - x^\infty) = 0$, for all $x^* \in \mathcal{H}^*$, we say that x^k converges weakly to x^∞ and write $x^k \rightharpoonup x^\infty$ as $k \to +\infty$. Similarly, for a function $x : [t_0, +\infty) \to \mathcal{H}$, $t \mapsto x(t)$ and $x^\infty \in \mathcal{H}$, with $\lim_{t \to +\infty} \langle x(t) - x^\infty, y \rangle = 0$, for all $y \in \mathcal{H}$, or equivalently $\lim_{t \to +\infty} x^*(x(t) - x^\infty) = 0$, for all $x^* \in \mathcal{H}^*$, we say that $x(\cdot)$ converges weakly to x^∞ and write $x(t) \rightharpoonup x^\infty$ as $t \to +\infty$. Additionally, at certain points we work in the weak*-topology on $\mathcal{H}^*$. For a sequence $\left(\xi^k\right)_{k \geq 0} \subset \mathcal{H}^*$ and an element ξ^∞, that satisfy $\lim_{k \to +\infty} (\xi^k - \xi^\infty)(x) = 0$ for all $x \in \mathcal{H}$, we say that ξ^k weak*-converges to ξ^∞ and we write $\xi^k \rightharpoonup^* \xi^\infty$, as $k \to +\infty$.

In this context we fix the following notation.

Definition 2.1.2. *We define the interior and closure of subsets of a Hilbert space and the open and closed ball.*

i) Let $A \subset \mathcal{H}$ *and* $B \subset \mathcal{H}^*$.

a) The interior of A is defined as $\operatorname{int}(A) := \bigcup_{\substack{U \subset A,\\ U \text{ open}}} U$.

b) The closure of A is defined as $\overline{A} := \bigcap_{\substack{A \subset K \subset \mathcal{H},\\ K \text{ closed}}} K$.

c) The weak-closure of B is defined as* $\overline{B}^* := \bigcap_{\substack{B \subset K \subset \mathcal{H}^*,\\ K \text{ weak}^*\text{-closed}}} K$.

ii) Let $x \in \mathcal{H}$ *and* $\varepsilon > 0$.

a) The open ball centered at x with radius ε *is* $B_\varepsilon(x) := \{y \in \mathcal{H} : \|x - y\| < \varepsilon\}$.

b) As a consequence the closure of the open ball centered at x with radius ε *is* $\overline{B_\varepsilon(x)} = \{y \in \mathcal{H} : \|x - y\| \leq \varepsilon\}$.

In the following, we recall the Banach-Alaoglu Theorem [206, Section 3.15] and the Eberlein-Šmulian Theorem [234] which are fundamental results in functional analysis and which are used implicitly in various parts of this thesis. While these theorems hold more generally in Banach spaces, we stay in the Hilbert space setting for the sake of consistency throughout this thesis.

Theorem 2.1.3. *The set* $B := \{\xi \in \mathcal{H}^* : \|\xi\|_* \leq 1\}$ *is weak*-compact.*

Theorem 2.1.4. *Let* $A \subset \mathcal{H}$. *Then, the following are equivalent:*

i) Each sequence in A has a subsequence that is weakly convergent in $\mathcal{H}$;

ii) Each sequence of elements in A has a weak cluster point in $\mathcal{H}$;

iii) The weak closure of A is weakly compact.

In multiple parts of this thesis, to prove weak convergence of a sequence or a function, we use Opial's Lemma. We recite a discrete and a continuous version of this lemma here. The discrete version of Opial's Lemma can be found in [186].

Lemma 2.1.5. *Let* $S \subseteq \mathcal{H}$ *be a nonempty set and let* $(x^k)_{k\geq 0} \subset \mathcal{H}$ *be a sequence satisfying the following conditions:*

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

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

Then, x^k *converges weakly to an element in S, i.e.,* $x^k \rightharpoonup x^\infty \in S$ *as* $k \to +\infty$.

For the continuous version of Opial's Lemma we refer to [13, Lemma 5.7].

Lemma 2.1.6. *Let $S \subseteq \mathcal{H}$ be a nonempty set and let $x : [t_0, +\infty) \to \mathcal{H}$ be a function satisfying the following conditions:*

i) For every $z \in S$, $\lim_{t\to+\infty}\|x(t) - z\|$ exists;

ii) Every weak sequential cluster point of $x(\cdot)$ belongs to S.

Then, $x(\cdot)$ converges weakly to an element in S, i.e., $x(t) \rightharpoonup x^\infty \in S$ as $t \to +\infty$.

2.1.2 Differentiability

In this subsection, we introduce different notions of the derivative of a function $f : \mathcal{H} \to \mathbb{R}$. We define the directional derivative, the Gâteaux derivative, the Fréchet derivative and the gradient. Furthermore, we present two lemmas to locally bound the function f using the gradient. The content of this subsection is contained in any book featuring analysis in normed spaces, see e.g., [75, 233].

Definition 2.1.7. *Let $f : \mathcal{H} \to \mathbb{R}$, $x \mapsto f(x)$ be a function and let $x, v \in \mathcal{H}$. We say that f is directionally differentiable at x in direction v, if the limit*

$$f'(x; v) := \lim_{t \searrow 0} \frac{f(x + tv) - f(x)}{t},$$

exists. The limit $f'(x, v)$ is called the directional derivative of f at x in direction v.

Let $x \in \mathcal{H}$. If there exists a linear and bounded operator $\mathrm{D}f(x) : \mathcal{H} \to \mathbb{R}$ such that for all $v \in \mathcal{H}$ it holds that

$$f'(x, v) = \mathrm{D}f(x)(v),$$

we say that f is Gâteaux differentiable in x and we call $\mathrm{D}f(x)$ the Gâteaux derivative. Moreover, f is called Gâteaux differentiable if it is Gâteaux differentiable in every $x \in \mathcal{H}$.

Definition 2.1.8. *Let $f : \mathcal{H} \to \mathbb{R}$, $x \mapsto f(x)$ be Gâteaux differentiable in $x \in \mathcal{H}$. Then, by Theorem 2.1.1 there exists a unique vector $\nabla f(x) = \mathcal{R}^{-1}(\mathrm{D}f(x)) \in \mathcal{H}$ with*

$$f'(x; v) = \mathrm{D}f(x)(v) = \langle \nabla f(x), v\rangle \quad \textit{for all} \quad v \in \mathcal{H}.$$

We call $\nabla f(x)$ the gradient of f in x.

Definition 2.1.9. *Let $f : \mathcal{H} \to \mathbb{R}$, $x \mapsto f(x)$ be a function. We say that f is Fréchet differentiable in $x \in \mathcal{H}$, if there exists a linear and bounded operator $A_x : \mathcal{H} \to \mathbb{R}$, with*

$$\lim_{v \to 0} \frac{|f(x + v) - f(x) - A_x(v)|}{\|v\|} = 0.$$

If f is Fréchet differentiable in x it is also Gâteaux differentiable in x and it holds that $A_x = \mathrm{D}f(x)$, and we call $\mathrm{D}f(x)$ the Fréchet derivative of f in x. Moreover, f is called Fréchet differentiable if it is Fréchet differentiable in every $x \in \mathcal{H}$.

Remark 2.1.10. *In this thesis, when we say that a function $f : \mathcal{H} \to \mathbb{R}$, $x \mapsto f(x)$ is (continuously) differentiable, we mean that it is (continuously) differentiable in the Fréchet sense.*

We close this subsection stating two lemmas which give local upper bounds on the function values of f close to a point $x \in \mathcal{H}$ using the gradient $\nabla f(x)$.

Lemma 2.1.11. *Let $f : \mathcal{H} \to \mathbb{R}$ be Fréchet differentiable and let $v \in \mathcal{H}$ with $\langle \nabla f(x), v\rangle < 0$. Then, there exists $\bar{t} > 0$ such that for all $t \in (0, \bar{t})$*

$$f(x + tv) < f(x).$$

Proof. Assume this is not the case. Then, there exists $(t_k)_{k\geq 0}$ with $t_k \searrow 0$ and $f(x+t_k v) \geq f(x)$ for all $k \geq 0$. From this we conclude

$$0 \leq \lim_{k\to+\infty} \frac{f(x + t_k v) - f(x)}{t_k} = f'(x)(v) = \langle \nabla f(x), v\rangle < 0,$$

which is a contradiction. □

The following lemma is known as the Descent Lemma [31, 38].

Lemma 2.1.12. *Let $f : \mathcal{H} \to \mathbb{R}$ be continuously Fréchet differentiable with L-Lipschitz continuous gradient ∇f, i.e., for all $x, y \in \mathcal{H}$, it holds that*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Then, for all $x, y \in \mathcal{H}$, it holds that

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2.$$

Proof. Let $x, y \in \mathcal{H}$ and define the function

$$\phi : [0, 1] \to \mathbb{R}, \quad t \mapsto \phi(t) := f(x + t(y - x)).$$

By definition $\phi(0) = f(x)$ and $\phi(1) = f(y)$. Further, $\phi(\cdot)$ is continuous on $[0, 1]$ and continuously differentiable on $(0, 1)$ as the composition of a continuously Fréchet differentiable and an affine linear function. Then, by the fundamental theorem of calculus, the Cauchy–Schwarz inequality and the Lipschitz continuity of ∇f, we obtain

$$\begin{aligned}
f(y) = \phi(1) = \phi(0) &+ \int_0^1 \frac{d}{dt}\phi(t)\,dt \\
&= f(x) + \int_0^1 \langle \nabla f(x + t(y - x)), y - x\rangle\,dt \\
&= f(x) + \langle \nabla f(x), y - x\rangle + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle\,dt \\
&\leq f(x) + \langle \nabla f(x), y - x\rangle + \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\|\,\|y - x\|\,dt \\
&\leq f(x) + \langle \nabla f(x), y - x\rangle + \int_0^1 Lt\|y - x\|^2\,dt \\
&= f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2.
\end{aligned}$$

□

2.1.3 Convex analysis

In this subsection, we present fundamental concepts from convex analysis. As this is a well-established field, we do not provide a comprehensive overview of all key results. Instead, we present a selection and introduce essential notation. For a comprehensive presentation of the mentioned topics, we refer to [3, 32, 93, 96, 205]. In the following, we primarily focus on the projection operator and summarize three lemmas concerning convex projections. Furthermore, we introduce convex functions, define the convex subdifferential and collect several results related to these objects.

Definition 2.1.13. *A set $C \subset \mathcal{H}$ is called convex, if for all $x, y \in C$ and all $\lambda \in [0,1]$ it holds that*

$$\lambda x + (1-\lambda)y \in C.$$

Before we discuss properties of convex sets, we introduce the positive unit simplex and the convex hull.

Definition 2.1.14. *Let $n \geq 1$. The n-dimensional positive unit simplex is defined as*

$$\Delta^n := \left\{ \theta \in \mathbb{R}^n \,:\, \sum_{i=1}^{n} \theta_i = 1 \quad \textit{and} \quad \theta_i \geq 0 \quad \textit{for} \quad i = 1, \ldots, n \right\}.$$

Definition 2.1.15. *Let $A \subset \mathcal{H}$ be an arbitrary set. The convex hull of a set is its smallest convex superset. The convex hull of a set A is formally defined in the following equivalent ways.*

i) $\operatorname{conv}(A) := \bigcap_{A \subset C \textit{ and } C \textit{ convex}} C$*;*

ii) $\operatorname{conv}(A) := \left\{ \sum_{i=1}^{n} \theta_i \xi^i \,:\, n \geq 1, \quad \theta \in \Delta^n, \quad \xi^i \in A \ \textit{ for } \ i = 1, \ldots, n \right\}$*.*

Remark 2.1.16. *Let $m \geq 1$. If $A = \{\xi^1, \ldots, \xi^m\} \subset \mathcal{H}$ consists out of finitely many vectors, then*

$$\operatorname{conv}(A) = \left\{ \sum_{i=1}^{m} \theta_i \xi^i \,:\, \theta \in \Delta^m, \quad \xi^i \in A \ \textit{ for } \ i = 1, \ldots, m \right\}.$$

Next, we define the projection onto a closed and convex set.

Theorem 2.1.17. *Let $C \subset \mathcal{H}$ be a convex and closed set and $x \in \mathcal{H}$ an arbitrary vector. Then, there exists a unique minimizer of the problem*

$$\min_{y \in C} \|y - x\|,$$

which is denoted by

$$\operatorname*{proj}_C(x) := \arg\min_{y \in C} \|y - x\|.$$

It holds that $z = \operatorname{proj}_C(x)$, if and only if

$$\langle y - z, z - x \rangle \geq 0, \quad \textit{for all} \quad y \in C. \tag{2.1}$$

We refer to (2.1) *as the variational characterization of the projection.*

The following lemma can be found in [18, Lemma 3.7].

Lemma 2.1.18. *Let $\mathcal{H}$ be a real Hilbert space, $C \subset \mathcal{H}$ a convex and compact set and $\xi \in \mathcal{H}$ a fixed vector. Then*

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

We originally introduced the following two lemmas in [217, Lemmas A.1, A.2].

Lemma 2.1.19. *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*

$$\textit{Find} \quad \xi \in \mathcal{H} \quad \textit{such that} \quad \eta = \operatorname*{proj}_{C+\xi}(0), \tag{2.2}$$

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 (2.2). 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 states 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 (2.2) 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 2.1.20. *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*

$$\textit{Find} \quad \xi \in \mathcal{H} \quad \textit{such that} \quad -a(\xi + \nu) = \operatorname*{proj}_{C+\xi}(0), \tag{2.3}$$

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 (2.3). It is easy to check that $-a(\xi + \nu) \in C + \xi$. Define the projection $p \coloneqq \operatorname{proj}_C(\nu)$. 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 (2.3). 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. □

In the following, we collect some results on convex functions starting with a formal definition of convexity.

Definition 2.1.21. *A function $f : \mathcal{H} \to \mathbb{R}$ is called convex, if for all $x, y \in \mathcal{H}$ and all $\lambda \in [0,1]$, it holds that*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Convex functions have nice topological properties especially in the presence of lower semicontinuity. We define sequential lower semicontinuity with respect to the strong and the weak topology in $\mathcal{H}$.

Definition 2.1.22. *Let $f : \mathcal{H} \to \mathbb{R}$ be a function.*

The function f is called sequentially lower semicontinuous in $x \in \mathcal{H}$, if for all $(x^k)_{k\geq 0} \subset \mathcal{H}$ with $x^k \to x$ as $k \to +\infty$, it holds that

$$f(x) \leq \liminf_{k\to+\infty} f(x^k).$$

Moreover, we say that f is sequentially lower semicontinuous if it is so in every $x \in \mathcal{H}$.

The function f is called sequentially weakly lower semicontinuous in $x \in \mathcal{H}$, if for all $(x^k)_{k\geq 0} \subset \mathcal{H}$ with $x^k \rightharpoonup x$ as $k \to +\infty$, it holds that

$$f(x) \leq \liminf_{k\to+\infty} f(x^k).$$

Moreover, we say that f is sequentially weakly lower semicontinuous, if it is so in every $x \in \mathcal{H}$.

We do not want to introduce the concept of lower semicontinuity and weak lower semicontinuity. The proper definition of these notions requires the introduction of nets which we omit in this thesis. The following theorem justifies this decision [32, Theorem 9.1].

Theorem 2.1.23. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex. Then, the following are equivalent:*

i) *f is sequentially weakly lower semicontinuous;*

ii) *f is sequentially lower semicontinuous;*

iii) *f is lower semicontinuous;*

iv) *f is weakly lower semicontinuous.*

The following proposition gives a characterization of smooth convex functions.

Proposition 2.1.24. *Let $f : \mathcal{H} \to \mathbb{R}$ be continuously differentiable. Then, the following are equivalent:*

i) *f is convex;*

ii) *For all $x, y \in \mathcal{H}$ it holds that*

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle.$$

The following lemma is an adaption of the Descent Lemma (Lemma 2.1.12) for convex smooth functions.

Lemma 2.1.25. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f. Then, for all $x, y, z \in \mathcal{H}$ it holds*

$$f(z) - f(x) \leq \langle \nabla f(y), z - x \rangle + \frac{L}{2}\|z - y\|^2.$$

Proof. The proof combines the Descent Lemma (Lemma 2.1.12) and Proposition 2.1.24. Let $x, y, z \in \mathcal{H}$. By the Descent Lemma it holds that

$$f(z) - f(y) \leq \langle \nabla f(y), z - y \rangle + \frac{L}{2}\|z - y\|^2. \tag{2.4}$$

From Proposition 2.1.24, we follow

$$f(y) - f(x) \leq \langle \nabla f(y), y - x \rangle. \tag{2.5}$$

Summing (2.4) and (2.5) gives

$$f(z) - f(x) \leq \langle \nabla f(y), z - x \rangle + \frac{L}{2}\|z - y\|^2.$$

□

We introduce two more restrictive variants of convexity.

Definition 2.1.26. *Let $f : \mathcal{H} \to \mathbb{R}$, $x \mapsto f(x)$ be a function. Then:*

i) *f is called strictly convex, if for all $x, y \in \mathcal{H}$ and all $\lambda \in (0, 1)$, it holds that*

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y);$$

ii) f is called μ-strongly convex for $\mu > 0$, if the function $f(\cdot) - \frac{\mu}{2}\|\cdot\|^2$ is convex.

Proposition 2.1.27. *Let $f : \mathcal{H} \to \mathbb{R}$ be μ-strongly convex and lower semicontinuous. Then f has a unique minimizer $x^* \in \mathcal{H}$ and for all $x \in \mathcal{H}$ it holds that*

$$f(x) \geq f(x^*) + \frac{\mu}{2}\|x - x^*\|^2.$$

We introduce a generalization of the gradient for nonsmooth convex functions.

Definition 2.1.28. *Let $f : \mathcal{H} \to \mathbb{R}$ be a function. The convex subdifferential of f in $x \in \mathcal{H}$ is defined as*

$$\partial f(x) := \{\xi \in \mathcal{H} \,:\, f(y) - f(x) \geq \langle \xi, y - x\rangle \quad \text{for all} \quad y \in \mathcal{H}\}.$$

The elements $\xi \in \partial f(x)$ are called subgradients.

For the convex subdifferential a generalization of Fermat's rule holds.

Proposition 2.1.29. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and lower semicontinuous. Then, $x^* \in \arg\min_{x\in\mathcal{H}} f(x)$ if and only if $0 \in \partial f(x^*)$.*

2.1.4 Clarke subdifferential

In this subsection, we introduce an extension of the derivative from smooth analysis to the class of locally Lipschitz continuous functions $f : \mathcal{H} \to \mathbb{R}$. The introduction we present in the following is part of our publication [215] and is mostly based on [75]. For a comprehensive treatment of generalized derivatives see also [174, 193]. Recall that a function $f : \mathcal{H} \to \mathbb{R}$ is called *locally Lipschitz continuous* in $x \in \mathcal{H}$, if there exist $\varepsilon > 0$ and a constant $L = L(x, \varepsilon) > 0$ with

$$|f(y) - f(z)| \leq L\,\|y - z\| \quad \text{for all } y, z \in B_\varepsilon(x).$$

Similarly, we call f globally Lipschitz continuous on $\mathcal{U} \subseteq \mathcal{H}$, if there exists a constant $L = L(\mathcal{U}) > 0$ with

$$|f(y) - f(z)| \leq L\,\|y - z\| \quad \text{for all } y, z \in \mathcal{U}.$$

We say that f is *locally (or globally) L-Lipschitz continuous*, if we want to point out the specific Lipschitz constant.

Definition 2.1.30. *Let $f : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous. Define the* generalized directional derivative *at x in direction $v \in \mathcal{H}$ as*

$$f^\circ(x; v) := \limsup_{y\to x, t\searrow 0} \frac{f(y + tv) - f(y)}{t}.$$

In the following, we refer to Propositions 2.1.1, 2.1.2 and 2.1.5 in [75] which state the most important facts on the generalized directional derivative.

Proposition 2.1.31. *Let $f : \mathcal{H} \to \mathbb{R}$ be locally L-Lipschitz continuous in $x \in \mathcal{H}$. Then:*

i) The function $v \mapsto f^\circ(x;v)$ *is finite, positively homogeneous, and subadditive on* $\mathcal{H}$ *(i.e.,* $f^\circ(x;tv) = tf^\circ(x;v)$ *and* $f^\circ(x;v+w) \leq f^\circ(x;v) + f^\circ(x,w)$ *for every* $t > 0$ *and* $v, w \in \mathcal{H}$*), and satisfies*

$$|f^\circ(x;v)| \leq L\,\|v\|;$$

ii) $f^\circ(x;v)$ *is upper semicontinuous as a function of* (x,v) *and, as a function of* v *alone, is* L*-Lipschitz continuous on* $\mathcal{H}$*;*

iii) $f^\circ(x;-v) = (-f)^\circ(x;v)$*.*

Using the generalized directional derivative we are able to define the so-called *(Clarke) subdifferential.*

Definition 2.1.32. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be locally Lipschitz continuous. Define the* (Clarke) subdifferential *in* x *as*

$$\partial_C f(x) \coloneqq \big\{\xi \in \mathcal{H}^* : f^\circ(x;v) \geq \xi(v) \text{ for all } v \in \mathcal{H}\big\}.$$

A functional ξ *in the set* $\partial_C f(x)$ *is called a* subderivative *of* f *in* x*.*

Proposition 2.1.33. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be locally* L*-Lipschitz continuous in* $x \in \mathcal{H}$*. Then:*

i) $\partial_C f(x)$ *is a nonempty, convex, weak*-compact subset of* $\mathcal{H}^*$ *and* $\|\xi\|_* \leq L$ *for every* ξ *in* $\partial_C f(x)$*;*

ii) For every v *in* $\mathcal{H}$*, it holds that*

$$f^\circ(x;v) = \max\big\{\xi(v) : \xi \in \partial_C f(x)\big\}.$$

The following result states that in the case of a smooth function, the Clarke subdifferential coincides with the derivative. We can derive this proposition from [75, Proposition 2.2.1].

Proposition 2.1.34. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be continuously differentiable and let* $x \in \mathcal{H}$*. Then,*

$$\partial_C f(x) = \{\mathrm{D}f(x)\}\,,$$

i.e., the Clarke subdifferential is a singleton only containing the Fréchet derivative of f *in* x*.*

Proposition 2.1.35. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be locally Lipschitz continuous in* x*. Then:*

i) We have $\xi \in \partial_C f(x)$ *if and only if* $f^\circ(x;v) \geq \xi(v)$ *for all* $v \in \mathcal{H}$*;*

ii) Let $(x^k)_{k\geq 0}$ *and* $(\xi^k)_{k\geq 0}$ *be sequences in* $\mathcal{H}$ *and* $\mathcal{H}^*$*, respectively, with* $\xi^k \in \partial_C f(x^k)$ *for all* $k \geq 0$*. Suppose that* x^k *converges to* x *as* $k \to +\infty$ *and that* ξ *is a weak*-accumulation point of* $(\xi^k)_{k\geq 0}$*. Then* $\xi \in \partial_C f(x)$*;*

iii) $\partial_C f(x) = \bigcap_{\varepsilon>0} \bigcup_{y \in \overline{B_\varepsilon(x)}} \partial_C f(y)$*.*

Recall that the Clarke subdifferential in infinite dimensions satisfies the well-known *Mean Value Theorem* (cf., [75, Theorem 2.3.7]).

Theorem 2.1.36. *Let $x, y \in \mathcal{H}$ and let $f : \mathcal{H} \to \mathbb{R}$ be Lipschitz continuous on an open set containing the line segment $[x, y]$. Then, there exists a point z on the open line segment (x, y) such that*

$$f(y) - f(x) \in \partial_C f(z)(y - x).$$

Note that, if f is locally Lipschitz continuous on $\mathcal{H}$, then any line segment $[x, y]$ has a neighborhood on which f is globally Lipschitz continuous since $[x, y]$ is compact in $\mathcal{H}$. We conclude this section with a necessary optimality condition based on the Clarke subdifferential [75, Proposition 2.3.2].

Proposition 2.1.37. *Let $f : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous. If $x^* \in \arg\min_{x \in \mathcal{H}} f(x)$, then $0 \in \partial_C f(x^*)$.*

2.2 Differential equations and inclusions

In this section, we present basic results on differential equations and inclusions. These are particularly important in Chapter 4 where we define various differential equations and inclusions with favorable properties in the context of multiobjective optimization.

In Subsection 2.2.1, we present two fundamental existence results in the theory of differential equations, namely Peano's Theorem and the Cauchy–Lipschitz Theorem and highlight the differences between these results. After discussing general differential equations, in Subsection 2.2.2, we move to the less classical subject of differential inclusions. To present this topic in an appropriate manner, we introduce definitions from set-valued analysis. This section concludes with Subsection 2.2.3 which contains a selection of essential differential and integral inequalities.

2.2.1 Differential equations

In this subsection, we present some classical results on differential equations taking values in a Hilbert space. Introductory literature on this topic can be found in [50, 143, 240]. Central to this subsection is the equation

$$\dot{x}(t) = \phi(x(t)),$$

which describes the evolution of a point $x(t)$ in the space $\mathcal{H}$ over time $t \in \mathbb{R}$, governed by a function $\phi : \mathcal{H} \to \mathcal{H}$. Here $\dot{x}(t) = \frac{d}{dt}x(t) \in \mathcal{H}$ denotes the first derivative of the position with respect to time. Differential equations play a fundamental role in all parts of natural sciences with prominent applications in physics, engineering, chemistry and economics.

We present the two most important existence results for differential equations: Peano's Theorem and the Cauchy–Lipschitz Theorem. These results differ significantly. Peano's Theorem is more general in the sense that it only requires the function $\phi(\cdot)$ to be continuous. However, it is limited to cases where the space $\mathcal{H}$ is finite-dimensional, i.e., $\dim(\mathcal{H}) < +\infty$ and guarantees only the local existence of solutions. In contrast, the Cauchy–Lipschitz Theorem also applies in infinite-dimensional spaces and yields existence of global solutions. On the down side, it requires Lipschitz continuity of the function $\phi(\cdot)$, i.e., $\|\phi(x) - \phi(y)\| \leq L\|x - y\|$ for all $x, y \in \mathcal{H}$. An important advantage of the Cauchy–Lipschitz Theorem is that it not only guarantees existence of a solution but also its uniqueness.

We formulate Peano's Theorem and the Cauchy–Lipschitz Theorem in the form of initial value problems which only evolve forward in time given some initial time $t_0 \in \mathbb{R}$ and some initial point $x_0 \in \mathcal{H}$, i.e., we are looking for solutions $x : [t_0, t_0 + \delta) \to \mathcal{H}$ for some positive $\delta > 0$.

Theorem 2.2.1. (Peano's Theorem) *Let $\mathcal{H}$ be a finite dimensional Hilbert space and let $\phi : \mathcal{H} \to \mathcal{H}$ be continuous. Given initial data $t_0 \in \mathbb{R}$ and $x_0 \in \mathcal{H}$, the Cauchy problem*

$$\left| \begin{array}{l} \dot{x}(t) = \phi(x(t)), \quad \text{for all} \quad t > t_0, \\ x(t_0) = x_0, \end{array} \right.$$

has a local solution, i.e., there exists $\delta > 0$ and a function $x : [t_0, t_0 + \delta) \to \mathcal{H}$ such that:

i) $x(\cdot)$ is continuous on $[t_0, t_0 + \delta)$;

ii) $x(\cdot)$ is continuously differentiable on $(t_0, t_0 + \delta)$;

iii) $x(t_0) = x_0$;

iv) For all $t \in (t_0, t_0 + \delta)$ the equation $\dot{x}(t) = \phi(x(t))$ holds.

Proof. The original proof dates back to the works of Peano [190, 191]. A modern proof can be found in [230]. □

Theorem 2.2.2. (Cauchy–Lipschitz Theorem) *Let $\mathcal{H}$ be a Hilbert space and let $\phi : \mathcal{H} \to \mathcal{H}$ be Lipschitz continuous. Given initial data $t_0 \in \mathbb{R}$ and $x_0 \in \mathcal{H}$, the Cauchy problem*

$$\left| \begin{array}{l} \dot{x}(t) = \phi(x(t)), \quad \text{for all} \quad t > t_0, \\ x(t_0) = x_0, \end{array} \right.$$

has a unique solution, i.e., there exists a unique function $x : [t_0, +\infty) \to \mathcal{H}$ such that:

i) $x(\cdot)$ is continuous on $[t_0, +\infty)$;

ii) $x(\cdot)$ is continuously differentiable on $(t_0, +\infty)$;

iii) $x(t_0) = x_0$;

iv) For all $t \in (t_0, +\infty)$ the equation $\dot{x}(t) = \phi(x(t))$ holds.

Proof. A proof of this theorem is contained in [50, Theorem 7.3]. □

2.2.2 Set-valued analysis and differential inclusions

In this subsection, we introduce several basic concepts from set-valued analysis and differential inclusions. The notation is consistent with [26] which includes a comprehensive overview of the topic. The introduction given in this subsection is closely aligned with the one provided in our paper [217, Appendix A].

In this subsection, let $\mathcal{X}$ and $\mathcal{Y}$ be real Hilbert spaces and let

$$G : \mathcal{X} \rightrightarrows \mathcal{Y}, \quad x \mapsto G(x) \subset \mathcal{Y},$$

be a set-valued map.

Definition 2.2.3. *We call $G : \mathcal{X} \rightrightarrows \mathcal{Y}$ upper semicontinuous in $x \in \mathcal{X}$, if for all open set $N \subset \mathcal{Y}$ with $G(x) \subset N$ there exists an open set $M \subset \mathcal{X}$ with $x \in M$ such that $G(M) \subset N$.*

Moreover, we call $G(\cdot)$ upper semicontinuous, if it is upper semicontinuous in every $x \in \mathcal{X}$.

Definition 2.2.4. *We call $G : \mathcal{X} \rightrightarrows \mathcal{Y}$ upper semicontinuous in x in the ε sense, if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $G(B_\delta(x)) \subset G(x) + B_\varepsilon(0)$.*
Moreover, we call $G(\cdot)$ upper semicontinuous in the ε sense, if it is upper semicontinuous in the ε sense in every $x \in \mathcal{X}$.

Proposition 2.2.5. *Let $G : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set valued map. The following statements hold:*

i) If $G(\cdot)$ is upper semicontinuous it is also upper semicontinuous in the ε sense;

ii) If $G(\cdot)$ is upper semicontinuous in the ε sense and takes compact values $G(x) \subset \mathcal{Y}$ for all $x \in \mathcal{X}$, then it is upper semicontinuous.

Definition 2.2.6. *We say that a map $\phi : \mathcal{X} \to \mathcal{Y}$ is locally compact if for every point $x \in \mathcal{X}$ there exists an open set $\mathcal{U} \subset \mathcal{X}$ with $x \in \mathcal{U}$ which is mapped into a compact set, i.e., there exists $\mathcal{K} \subset \mathcal{Y}$ compact with $\phi(\mathcal{U}) \subset \mathcal{K}$.*

We recite the following existence result from [26, p. 98, Theorem 3].

Theorem 2.2.7. *Let $\Omega \subset \mathbb{R} \times \mathcal{X}$ be an open set containing (t_0, x_0). Let $G : \Omega \rightrightarrows \mathcal{X}$ be an upper semicontinuous set-valued map which takes as values nonempty, closed and convex subsets of $\mathcal{X}$. Assume that the map $(t, x) \mapsto \mathrm{proj}_{G(t,x)}(0)$ is locally compact. Then, there exists $T > t_0$ and an absolutely continuous function $x(\cdot)$ defined on $[t_0, T]$ which is a solution of the differential inclusion*

$$\dot{x}(t) \in G(t, x(t)), \quad \textit{for almost all} \quad t \in (t_0, T), \tag{2.6}$$

with $x(t_0) = x_0$.

Remark 2.2.8. *Consider the general differential inclusion (2.6). A solutions $x : [t_0, T] \to \mathcal{X}$ given by Theorem 2.2.7 is not differentiable but merely absolutely continuous. Therefore, the notion $\dot{x}(t) \in G(t, x(t))$ may not hold on the entire domain $[t_0, T]$. An absolutely continuous function $x : [t_0, T] \to \mathcal{X}$ is differentiable almost everywhere in $[t_0, T]$. A solution $x(\cdot)$ to (2.6) satisfies the inclusion $\dot{x}(t) \in G(t, x(t))$ in almost every t, where the derivative $\dot{x}(t)$ is defined. In general $\dot{x}(\cdot)$ will not be continuous. But since $x(\cdot)$ is absolutely continuous with values in a Hilbert space (which satisfies the Radon-Nikodym property), $\dot{x}(\cdot)$ is Bochner integrable and $x(t) = x(t_0) + \int_{t_0}^{t} \dot{x}(s)\, ds$ for all $t \geq t_0$ (see [76, 87]).*

2.2.3 Differential and integral inequalities

A general introduction to differential and integral inequalities can be found in [187]. This subsection primarily presents a collection of technical lemmas used in the analysis of certain differential equations in this thesis. The first result we present is the well-known Gronwall–Bellman Lemma.

Lemma 2.2.9. *Let $t_0 \in \mathbb{R}$ and $T > t_0$. Let $\phi \in L^1([t_0, T], \mathbb{R})$ with $\phi(t) \geq 0$ for almost all $t \in [t_0, T]$ and let $a \geq 0$. Let $u : [t_0, T] \to \mathbb{R}$ be a continuous function satisfying*

$$u(t) \leq a + \int_{t_0}^{t} \phi(s)u(s)\, ds,$$

for all $t \in [t_0, T]$. Then for all $t \in [t_0, T]$ it holds that

$$u(t) \leq a \exp\left(\int_{t_0}^{t} \phi(s)\, ds\right).$$

Proof. A proof of this lemma is contained in [51, Lemma A.4]. □

Lemma 2.2.10. *Let $t_0 \in \mathbb{R}$ and $T > t_0$. Let $\phi \in L^1([t_0, T], \mathbb{R})$ with $\phi(t) \geq 0$ for almost all $t \in [t_0, T]$ and let $a \geq 0$. Let $u : [t_0, T] \to \mathbb{R}$ be a continuous function satisfying*

$$\frac{1}{2}u(t)^2 \leq \frac{1}{2}a^2 + \int_{t_0}^{t} \phi(s)u(s)\, ds,$$

for all $t \in [t_0, T]$. Then for all $t \in [t_0, T]$ it holds that

$$|u(t)| \leq a + \int_{t_0}^{t} \phi(s)\, ds.$$

Proof. A proof of this lemma is contained in [51, Lemma A.5]. □

The following two lemmas are standard results used to establish the convergence of a real-valued function that satisfies a differential inequality.

Lemma 2.2.11. *Let $t_0 \in \mathbb{R}$ and let $h : [t_0, +\infty) \to \mathbb{R}$ be a continuously differentiable function which is bounded from below. Assume*

$$\ddot{h}(t) + \alpha\dot{h}(t) \leq g(t),$$

for almost all $t \geq t_0$, with $\alpha > 0$ and $g \in L^1([t_0, +\infty), \mathbb{R})$. Then, $\lim_{t\to+\infty} h(t)$ exists.

Proof. A proof of this lemma is contained in [19, Lemma 4.2]. □

Lemma 2.2.12. *Let $t_0 \in \mathbb{R}$ and let $h : [t_0, +\infty) \to \mathbb{R}$ be a continuously differentiable function which is bounded from below. Assume*

$$t\ddot{h}(t) + \alpha\dot{h}(t) \leq g(t),$$

for almost all $t \geq t_0$, with $\alpha > 1$ and $g \in L^1([t_0, +\infty), \mathbb{R})$ a nonnegative function. Then, $\lim_{t\to+\infty} h(t)$ exists.

Proof. A proof can be found in [23, Lemma A.6.]. □

We also include the following discrete version of Lemma 2.2.12. Although, this result is not a differential inequality per se it fits best here next to its continuous counterpart.

Lemma 2.2.13. *Let $\alpha \geq 3$ and let $(\theta_k)_{k\geq 0}$, $(\delta_k)_{k\geq 0} \subset [0, +\infty)$ be positive sequences such that for all $k \geq 0$*

$$\theta_{k+1} \leq \frac{k-1}{k+\alpha-1}\theta_k + \delta_k.$$

If $\sum_{k=0}^{\infty} k\delta_k < +\infty$, then $\sum_{k=0}^{\infty} \theta_k < +\infty$.

Proof. A proof can be found in [13, Lemma 5.10.]. □

The following result is a technical lemma on the derivative of the pointwise minimum of a finite collection of absolutely continuous functions. This lemma can be found in our publication [49].

Lemma 2.2.14. *For $i = 1, \dots, m$, let $h_i : [t_0, +\infty) \to \mathbb{R}$ be absolutely continuous functions on every interval $[t_0, T]$ for $T \geq t_0$. Define $h : [t_0, +\infty) \to \mathbb{R}$, $t \mapsto h(t) := \min_{i=1,\dots,m} h_i(t)$. Then, the following statements are true:*

i) The function h is absolutely continuous on every interval $[t_0, T]$ for $T \geq t_0$, and therefore differentiable in almost all $t \geq t_0$;

ii) For almost all $t > t_0$ there exists $i \in \{1, \dots, m\}$ such $h(t) = h_i(t)$ and $\frac{d}{dt}h(t) = \frac{d}{dt}h_i(t)$.

Proof.

i) The minimum of a finite family of absolutely continuous functions is absolutely continuous.

ii) Let $t > t_0$ such that $h(\cdot)$ and $h_i(\cdot)$ are differentiable in t for all $i = 1, \dots, m$. Take an arbitrary sequence $(\tau_k)_{k\geq 0}$ with $\lim_{k\to+\infty} \tau_k = 0$. Then there exists $i \in \{1, \dots, m\}$ and a subsequence $(k_l)_{l\geq 0} \subset \mathbb{N}$ with $h(t + \tau_{k_l}) = h_i(t + \tau_{k_l})$ for all $l \geq 0$. From the continuity of $h(\cdot)$ and $h_i(\cdot)$, it holds $h(t) = h_i(t)$. By the definition of the derivative, we get

$$\frac{d}{dt}h(t) = \lim_{l\to+\infty} \frac{h(t+\tau_{k_l}) - h(t)}{\tau_{k_l}} = \lim_{l\to+\infty} \frac{h_i(t+\tau_{k_l}) - h_i(t)}{\tau_{k_l}} = \frac{d}{dt}h_i(t).$$

□

We recite the following lemma from our publication [49].

Lemma 2.2.15. *Let $\alpha, \beta, a, b > 0$ be given constants, and let $t_0 > 0$. Then,*

$$\int_{t_0}^{t} \alpha s^{-a} \exp(\beta s^b) ds = \mathcal{O}\left(t^{1-(a+b)} \exp(\beta t^b)\right) \quad \text{as } t \to +\infty.$$

Proof. For $t \geq t_0$, we use integration by parts to get

$$\begin{aligned}\int_{t_0}^{t} \alpha s^{-a} \exp\left(\beta s^b\right) ds &= \frac{\alpha}{\beta b}\int_{t_0}^{t} s^{1-(a+b)} \frac{d}{ds}\exp\left(\beta s^b\right) ds \\ &= \frac{\alpha}{\beta b}\left[s^{1-(a+b)}\exp\left(\beta s^b\right)\right]_{t_0}^{t} - \frac{1-(a+b)}{\beta b}\int_{t_0}^{t} \alpha s^{-(a+b)}\exp\left(\beta s^b\right) ds. \end{aligned} \tag{2.7}$$

Since $b > 0$, there exists $t_1 \geq t_0$ such that for all $t \geq t_1$

$$\left|\frac{1-(a+b)}{\beta b}\right| t^{-b} \leq \frac{1}{2}. \tag{2.8}$$

Define $C_1 := \left|\frac{1-(a+b)}{\beta b}\right| \int_{t_0}^{t_1} \alpha s^{-(a+b)} \exp\left(\beta s^b\right) ds$. Then, (2.7) and (2.8) yield for all $t \geq t_0$

$$\begin{aligned}\int_{t_0}^{t} \alpha s^{-a} \exp\left(\beta s^b\right) ds &\leq \frac{\alpha}{\beta b}\left[s^{1-(a+b)}\exp\left(\beta s^b\right)\right]_{t_0}^{t} + C_1 + \left|\frac{1-(a+b)}{\beta b}\right| \int_{t_1}^{t} \alpha s^{-(a+b)}\exp\left(\beta s^b\right) ds \\ &\leq \frac{\alpha}{\beta b}\left[s^{1-(a+b)}\exp\left(\beta s^b\right)\right]_{t_0}^{t} + C_1 + \frac{1}{2}\int_{t_1}^{t} \alpha s^{-a}\exp\left(\beta s^b\right) ds \\ &\leq \frac{\alpha}{\beta b}\left[s^{1-(a+b)}\exp\left(\beta s^b\right)\right]_{t_0}^{t} + C_1 + \frac{1}{2}\int_{t_0}^{t} \alpha s^{-a}\exp\left(\beta s^b\right) ds,\end{aligned}$$

hence

$$\int_{t_0}^{t} \alpha s^{-a} \exp\left(\beta s^b\right) ds \leq \frac{2\alpha}{\beta b}\left[s^{1-(a+b)}\exp\left(\beta s^b\right)\right]_{t_0}^{t} + 2C_1.$$

Defining $C_2 := -\frac{2\alpha}{\beta b}(t_0)^{1-(a+b)} \exp\left(\beta (t_0)^b\right) + 2C_1$, we obtain for all $t \geq t_0$

$$\int_{t_0}^{t} \alpha s^{-a} \exp\left(\beta s^b\right) ds \leq \frac{2\alpha}{\beta b} t^{1-(a+b)} \exp\left(\beta t^b\right) + C_2,$$

and the asymptotic bound holds. □

The following lemma can be seen as a generalization of Lemma 2.2.12. It is a modification of a lemma presented in [146, Lemma 16] which we originally introduced in [49].

Lemma 2.2.16. *Let $t_0 > 0$, $\alpha > 0$, $q \in (0,1)$, and let $k : [t_0, +\infty) \to \mathbb{R}$ be a nonnegative function such that*

$$\left(t \mapsto t^q k(t)\right) \in L^1\left([t_0, +\infty), \mathbb{R}\right). \tag{2.9}$$

Let $h : [t_0, +\infty) \to \mathbb{R}$ be a continuously differentiable function that is bounded from below and possesses an absolutely continuous derivative $\dot{h}(\cdot)$. Further, assume $h(\cdot)$ satisfies

$$\ddot{h}(t) + \frac{\alpha}{t^q}\dot{h}(t) \leq k(t) \quad \textit{for almost all } t \geq t_0. \tag{2.10}$$

Then, $\left(t \mapsto \left[\dot{h}(t)\right]_+\right) \in L^1\left([t_0, +\infty), \mathbb{R}\right)$, where $\left[\dot{h}(t)\right]_+$ denotes the positive part of $\dot{h}(t)$, and further $\lim_{t\to+\infty} h(t)$ exists.

Proof. Define the function

$$\mathfrak{M}:[t_0,+\infty)\to\mathbb{R},\quad t\mapsto\mathfrak{M}(t):=\exp\left(\int_{t_0}^{t}\frac{\alpha}{s^q}ds\right)=C_{\mathfrak{M}}\exp\left(\frac{\alpha}{1-q}t^{1-q}\right),$$

with $C_{\mathfrak{M}}:=\exp\left(-\frac{\alpha}{1-q}t_0^{1-q}\right)$, and $b:=\frac{\alpha}{1-q}>0$. For $t\geq t_0$, using integration by parts, we have

$$\begin{aligned}C_{\mathfrak{M}}\int_t^{+\infty}\frac{ds}{\mathfrak{M}(s)}&=\int_t^{+\infty}\exp\left(-bs^{1-q}\right)ds=-\frac{1}{\alpha}\int_t^{+\infty}s^q\frac{d}{ds}\exp\left(-bs^{1-q}\right)ds\\&=-\frac{1}{\alpha}\left(\left[s^q\exp\left(-bs^{1-q}\right)\right]_t^{+\infty}-\int_t^{+\infty}qs^{q-1}\exp\left(-bs^{1-q}\right)ds\right)\\&=\frac{t^q}{\alpha}\exp\left(-bt^{1-q}\right)+\frac{q}{\alpha}\int_t^{+\infty}s^{q-1}\exp(-bs^{1-q})ds.\end{aligned}\tag{2.11}$$

As $q-1<0$, there exists $t_1\geq t_0$ such that for all $t\geq t_1$ the inequality $\frac{q}{\alpha}t^{q-1}\leq\frac{1}{2}$ holds and hence

$$\frac{q}{\alpha}\int_t^{+\infty}s^{q-1}\exp(-bs^{1-q})ds\leq\frac{1}{2}\int_t^{+\infty}\exp(-bs^{1-q})ds.\tag{2.12}$$

Combining (2.11) and (2.12), we conclude that for all $t\geq t_1$

$$C_{\mathfrak{M}}\int_t^{+\infty}\frac{ds}{\mathfrak{M}(s)}=\int_t^{+\infty}\exp\left(-bs^{1-q}\right)ds\leq\frac{2t^q}{\alpha}\exp\left(-bt^{1-q}\right).\tag{2.13}$$

Using the definition of $\mathfrak{M}(\cdot)$, equality (2.13) yields for all $t\geq t_1$

$$\left(\int_t^{+\infty}\frac{ds}{\mathfrak{M}(s)}\right)\mathfrak{M}(t)=\left(\int_t^{+\infty}\exp\left(-bs^{1-q}\right)\right)\exp\left(bt^{1-q}\right)\leq\frac{2t^q}{\alpha}.\tag{2.14}$$

We multiply (2.14) by $k(\cdot)$, integrate from t_0 to $+\infty$, and apply relation (2.9) to follow

$$\int_{t_0}^{+\infty}\left(\int_t^{+\infty}\frac{ds}{\mathfrak{M}(s)}\right)\mathfrak{M}(t)k(t)dt<+\infty.\tag{2.15}$$

By the definition of $\mathfrak{M}(\cdot)$, we have $\frac{d}{dt}\mathfrak{M}(t)=\mathfrak{M}(t)\frac{\alpha}{t^q}$ and then, by (2.10),

$$\frac{d}{dt}\left(\mathfrak{M}(t)\dot{h}(t)\right)=\mathfrak{M}(t)\ddot{h}(t)+\mathfrak{M}(t)\frac{\alpha}{t^q}\dot{h}(t)\leq\mathfrak{M}(t)k(t)\quad\text{for almost all } t\geq t_0.\tag{2.16}$$

We integrate (2.16) from t_0 to $t\geq t_0$ and observe

$$\mathfrak{M}(t)\dot{h}(t)-\mathfrak{M}(t_0)\dot{h}(t_0)\leq\int_{t_0}^{t}\mathfrak{M}(s)k(s)ds.$$

The function $k(\cdot)$ takes nonnegative values only and we derive for all $t\geq t_0$

$$\left[\dot{h}(t)\right]_+\leq\frac{|\mathfrak{M}(t_0)\dot{h}(t)|}{\mathfrak{M}(t)}+\frac{1}{\mathfrak{M}(t)}\int_{t_0}^{t}\mathfrak{M}(s)k(s)ds.$$

We integrate this inequality from t_0 to $+\infty$ and write

$$\int_{t_0}^{+\infty} \left[\dot{h}(t)\right]_+ dt \leq \int_{t_0}^{t} \frac{|\mathfrak{M}(t_0)\dot{h}(t)|}{\mathfrak{M}(t)} dt + \int_{t_0}^{+\infty} \frac{1}{\mathfrak{M}(t)} \left(\int_{t_0}^{t} \mathfrak{M}(s)k(s)ds \right) dt. \tag{2.17}$$

Since $\mathfrak{M}(\cdot)$ grows at an exponential rate, we have $\int_{t_0}^{+\infty} \frac{|\mathfrak{M}(t_0)\dot{h}(t)|}{\mathfrak{M}(t)} dt < +\infty$. We apply Fubini's Theorem to the second integral in (2.17) and combine it with (2.15) to conclude

$$\int_{t_0}^{+\infty} \frac{1}{\mathfrak{M}(t)} \left(\int_{t_0}^{t} \mathfrak{M}(s)k(s)ds \right) dt = \int_{t_0}^{+\infty} \left(\int_{t}^{+\infty} \frac{ds}{\mathfrak{M}(s)} \right) \mathfrak{M}(t)k(t)dt < +\infty. \tag{2.18}$$

Equation (2.17) and (2.18) imply

$$\int_{t_0}^{+\infty} \left[\dot{h}(t)\right]_+ dt < +\infty,$$

and by the lower boundedness of $h(\cdot)$ we follow that $\lim_{t\to+\infty} h(t)$ exists. $\square$

2.3 Multiobjective optimization

In this section, we describe the *multiobjective optimization problem* that forms the main focus of this thesis. In multiobjective optimization the goal is to simultaneously minimize multiple objective functions. For an introduction to this area and the related field of vector optimization we refer the reader to [88, 90, 133, 169]. The multiobjective optimization problem reads as

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

where the functions $f_i : \mathcal{H} \to \mathbb{R}$ for $i = 1, \dots, m$ are called the *objective functions*. To summarize the objective functions in a single vector valued function, we use the notation

$$F : \mathcal{H} \to \mathbb{R}^m, \quad x \mapsto F(x) := (f_1(x), \dots, f_m(x))^\top . \tag{2.19}$$

The space $\mathcal{H}$ is called the *decision space*, and the space $\mathbb{R}^m$ is referred to as the *image space*. The set $F(\mathcal{H}) = \{F(x) : x \in \mathcal{H}\} \subset \mathbb{R}^m$ is called the *attainable set*. Note, that we use the notation F instead of f in (2.19), as the latter is reserved for scalar functions, i.e., $f : \mathcal{H} \to \mathbb{R}$.

This section is organized as follows. In Subsection 2.3.1, we define solutions to problem (MOP). We introduce different notions of Pareto optimality and provide an illustrative example. Subsection 2.3.2 contains a discussion of necessary first-order optimality conditions. We cover the smooth case which uses gradient information of the objective functions, and the nonsmooth case which is based on the Clarke subdifferential. Afterwards, in Subsection 2.3.3, we define a merit function which forms a key ingredient of our analysis. Merit functions define suitable scalarizations of the objective functions, characterize the quality of a solution, and are important for comparing the convergence speed of different methods. In Subsection 2.3.4, we introduce first-order methods for multiobjective optimization. We define common descent directions, the multiobjective steepest descent direction, and conclude with an illustrative analysis of the multiobjective steepest descent method.

2.3.1 Pareto optimality

The optimization problem (MOP) involves multiple objective functions, and therefore the classical concept of optimality from scalar optimization cannot be applied directly. For every vector in the decision space $x \in \mathcal{H}$ the objective function value in the image space $F(x)$ is an element of the attainable set $F(\mathcal{H})$ which is a subset of $\mathbb{R}^m$. Since there is no total order on $\mathbb{R}^m$, there generally does not exist a unique optimal function value as in scalar optimization, unless the objectives are non-conflicting and share a common minimizer. A suitable generalization of optimality when dealing with multiple objectives is the notion of Pareto optimality which is attributed to Pareto [189]. This notion shifts the focus from optimal solutions to optimal compromises, i.e., we are interested in points where we cannot improve one objective function without worsening at least one other objective. This idea is formalized in the following Definition [169].

Definition 2.3.1. *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 $i = 1, \dots, m$, and $f_j(x) < f_j(x^*)$ for at least one index j;*

ii) A point $x^* \in \mathcal{H}$ is locally Pareto optimal *if there exists $\varepsilon > 0$ such that x^* is Pareto optimal in $B_\varepsilon(x^*)$;*

iii) A point $x^* \in \mathcal{H}$ is weakly Pareto optimal *if there does not exist another point $x \in \mathcal{H}$ such that $f_i(x) < f_i(x^*)$ for all $i = 1, \dots, m$;*

iv) A point $x^* \in \mathcal{H}$ is locally weakly Pareto optimal *if there exists $\varepsilon > 0$ such that x^* is weakly Pareto optimal in $B_\varepsilon(x^*)$.*

From Definition 2.3.1, we immediately derive the following proposition which relates the different concepts of Pareto optimality.

Proposition 2.3.2. *Consider the multiobjective optimization problem* (MOP) *and let $x \in \mathcal{H}$. Then, the following statements hold:*

i) If x is Pareto optimal, then x is weakly Pareto optimal;

ii) If x is Pareto optimal, then x is locally Pareto optimal;

iii) If x is weakly Pareto optimal, then x is locally weakly Pareto optimal;

iv) If x is locally Pareto optimal, then x is locally weakly Pareto optimal.

Remark 2.3.3. *Another concept which we use in various parts of this thesis is the notion of dominance. Let $x, y \in \mathcal{H}$. We say that x dominates y, if $f_i(x) \leq f_i(y)$ for all $i = 1, \dots, m$ and there exists $j \in \{1, \dots, m\}$ with $f_j(x) < f_j(y)$. Furthermore, we say that x strictly dominates y, if $f_i(x) < f_i(y)$ holds for all $i = 1, \dots, m$.*

Besides the notation in Definition 2.3.1, the literature on multiobjective optimization and vector optimization includes a variety of terms to describe Pareto optimal points, such as (Pareto) efficient points, nondominated points or noninferior points [65, 90].

Using the notions of optimality introduced in Definition 2.3.1, we define the so-called *Pareto set* and *Pareto front.*

Definition 2.3.4. *Consider the multiobjective optimization problem* (MOP)*. We define the following sets:*

i) The Pareto set *is denoted by $\mathcal{P} := \{x \in \mathcal{H} : x \text{ is Pareto optimal}\}$;*

ii) The weak Pareto set *is denoted by $\mathcal{P}_w := \{x \in \mathcal{H} : x \text{ is weakly Pareto optimal}\}$;*

iii) The Pareto front *is denoted by $F(\mathcal{P}) = \{F(x) \in \mathbb{R}^m : x \text{ is Pareto optimal}\}$;*

iv) The weak Pareto front *is denoted by $F(\mathcal{P}_w) = \{F(x) \in \mathbb{R}^m : x \text{ is weakly Pareto optimal}\}$.*

Additionally, we often need the lower level set of a vector-valued function which we introduce in the following definition.

Definition 2.3.5. *Consider the multiobjective optimization problem* (MOP).

i) Let $a \in \mathbb{R}^m$. *We define*

$$\mathcal{L}(F,a) \coloneqq \{x \in \mathcal{H} \,:\, F(x) \leqq a\} = \bigcap_{i=1}^{m} \{x \in \mathcal{H} \,:\, f_i(x) \leq a_i\},$$

where "$\leqq$*" denotes the partial order on* $\mathbb{R}^m$ *induced by* $\mathbb{R}^m_+$. *For* $a, b \in \mathbb{R}^m$, *we write* $a \leqq b$ *if and only if* $a_i \leq b_i$ *for all* $i = 1, \dots, m$.

ii) We denote the intersection of a lower level set and the weak Pareto set by

$$\mathcal{LP}_w(F,a) \coloneqq \mathcal{L}(F,a) \cap \mathcal{P}_w.$$

In the following elementary example, we illustrate the Pareto set and the Pareto front.

Example 2.3.6. *Given matrices and vectors*

$$Q_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \textit{and} \quad c_1 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 0 \\ 5 \end{bmatrix},$$

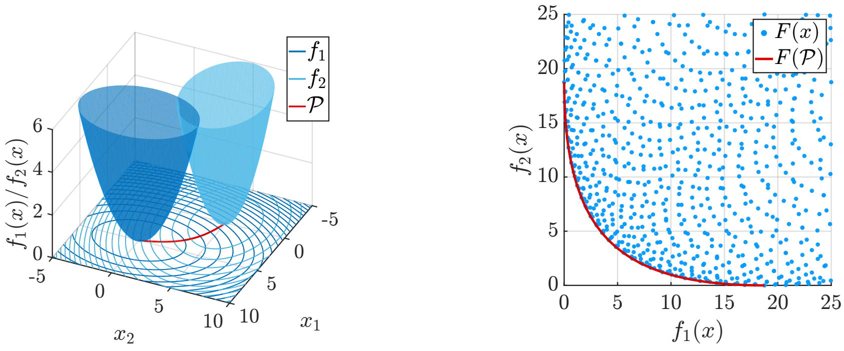
define the objective functions

$$f_i : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto f_i(x) \coloneqq \frac{1}{2}(x - c_i)^\top Q_i (x - c_i), \quad \textit{for} \quad i = 1, 2. \tag{2.20}$$

The objective functions f_1 *and* f_2 *are strongly convex and continuously differentiable with respective unique minimizers* c_1 *and* c_2. *In this example, we define the multiobjective optimization problem*

(MOP-Ex)
$$\min_{x \in \mathbb{R}^2} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix},$$

with the objective functions f_i *for* $i = 1, 2$ *defined in* (2.20).



(a) (b)

Figure 2.1: Subfigure 2.1a shows objective function f_1 and f_2 and the Pareto set $\mathcal{P}$ of (MOP-Ex) in the decision space. Subfigure 2.1a visualizes the attainable set $F(\mathcal{H})$ and the Pareto front $F(\mathcal{P})$ in the image space.

Figure 2.1 illustrates the Pareto set and the Pareto front of (MOP) *given the objective functions* f_1 *and* f_2 *defined in* (2.20). *For this problem the Pareto set* $\mathcal{P}$ *can be computed explicitly. It is shown in Subfigure 2.1a as a red line connecting* c_1 *and* c_2*, the respective minimizers of* f_1 *and* f_2*. The Pareto front* $F(\mathcal{P})$ *of this problem is visualized in Subfigure 2.1b. The attainable set* $F(\mathcal{H})$ *is approximated by an evaluation of the objective functions* f_1 *and* f_2 *on a equidistant grid, and plotted as blue dots. As Definition 2.3.1 suggests, for each point in the Pareto set, there does not exist another point which is strictly better with respect to all objective function values. Already in this simple example, we see that there does not exist a single optimal solution to* (MOP)*, but a continuum of optimal compromises.*

2.3.2 Necessary optimality conditions

The definition of optimality given in Definition 2.3.1 is difficult to apply in practice. Given a point $x \in \mathcal{H}$, to check whether it is Pareto optimal using Definition 2.3.1 directly, we have to compare its function value against the function values of all other points. This is generally not feasible, and therefore we need more sophisticated ways to identify optimal points. For this reason, it is more practical to work with necessary and sufficient optimality conditions. These usually rely on first- and higher-order derivative information of the objective functions and are expressed by systems of equations, which can be verified more efficiently. In addition, these conditions give further insight into the optimization problem and can be used to define numerical methods to solve the optimization problem at hand algorithmically.

For scalar optimization problems $\min_{x \in \mathcal{H}} f(x)$, with a continuously differentiable objective function $f : \mathcal{H} \to \mathbb{R}$ the so-called Karush-Kuhn-Tucker conditions (KKT conditions [135, 141]), which are also known in the smooth and unconstrained case as Fermat's rule $\nabla f(x) = 0$ [32], provide a necessary optimality condition. In Subsection 2.1.4, we have already seen that for locally Lipschitz functions $f : \mathcal{H} \to \mathbb{R}$, this condition can be generalized to $0 \in \partial_C f(x)$, where $\partial_C f(x)$ is the Clarke subdifferential of f in x. In this subsection, we discuss how these optimality conditions can be generalized to the setting of multiobjective optimization and the concept of Pareto optimality introduced in Definition 2.3.1. Additionally, we describe under which further assumptions on the objective functions these conditions are not only necessary but also sufficient.

The case of continuously differentiable objective functions

If the objective functions of (MOP) are continuously differentiable, we can define the following necessary condition for Pareto optimality [141, 169].

Theorem 2.3.7. *Let* $f_i : \mathcal{H} \to \mathbb{R}$ *be continuously differentiable for* $i = 1, \dots, m$ *and let* x^* *be (locally weakly) Pareto optimal. Then, there exists* $\theta \in \Delta^m$ *such that*

$$\sum_{i=1}^{m} \theta_i \nabla f_i(x^*) = 0,$$

which is equivalent to

$$0 \in \operatorname{conv}\left(\{\nabla f_i(x^*) \,:\, i = 1, \dots, m\}\right). \tag{2.21}$$

We call a point $x^* \in \mathcal{H}$ *satisfying* (2.21) Pareto critical.

The following Theorem investigates Pareto critical points under additional assumptions on the objective functions.

Theorem 2.3.8. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable for $i = 1, \dots, m$ and let $x^* \in \mathcal{H}$ be Pareto critical. Then, the following statements hold:*

i) If f_i is convex for all $i = 1, \dots, m$, then x^* is weakly Pareto optimal;

ii) If f_i is strictly convex for all $i = 1, \dots, m$, then x^* is Pareto optimal.

Proof. A proof is contained in [220]. □

In the following example we discuss Pareto critical points.

Example 2.3.9. *Reconsider the multiobjective optimization problem from Example 2.3.6.*

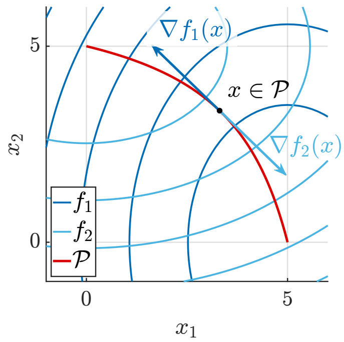


Figure 2.2: Contour plots of f_1 and f_2 and Pareto set $\mathcal{P}$ of (MOP-Ex) with a Pareto optimal point $x \in \mathcal{P}$ and corresponding gradients $\nabla f_1(x)$ and $\nabla f_2(x)$.

Figure 2.2 shows a Pareto optimal point $x \in \mathcal{P}$ and the gradients $\nabla f_i(x)$ for $i = 1, 2$ evaluated in x. As the gradients point in opposing directions, there exists $\theta \in \Delta^2$ such that $\theta_1 \nabla f_1(x) + \theta_2 \nabla f_2(x) = 0$ and therefore x is Pareto critical. The objective function f_1 and f_2 defined in (2.20) *are strictly convex and hence Theorems 2.3.7 and 2.3.8 state that x is Pareto optimal if and only if x is Pareto critical. Therefore, in this example we can compute $\mathcal{P}$ solving the nonlinear, constrained system of equations*

$$\theta_1 \nabla f_1(x) + \theta_2 \nabla f_2(x) = 0, \quad \text{with} \quad x \in \mathcal{H} \quad \text{and} \quad \theta \in \Delta^2,$$

from which we obtain

$$\mathcal{P} = \left\{ \begin{bmatrix} \frac{10\lambda}{\lambda+1} \\ \frac{10-10\lambda}{2-\lambda} \end{bmatrix} : \lambda \in [0,1] \right\}.$$

The case of locally Lipschitz continuous objective functions

In Subsection 2.1.4, the Clarke subdifferential is introduced for a locally Lipschitz continuous function $f : \mathcal{H} \to \mathbb{R}$. Proposition 2.1.37 states that $0 \in \partial_C f(x)$ is a necessary condition for $x \in \arg\min_{x \in \mathcal{H}} f(x)$. This condition can be extended to the multiobjective setting similar to Theorem 2.3.7 as the following theorem shows.

Theorem 2.3.10. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous for all $i = 1, \dots, m$ and let $x^* \in \mathcal{H}$ be (locally weakly) Pareto optimal. Then, there exists $\theta \in \Delta^m$ and $\xi_i \in \partial_C f_i(x^*)$ for $i = 1, \dots, m$ such that*

$$0 = \sum_{i=1}^{m} \theta_i \xi_i.$$

which is equivalent to

$$0 \in \operatorname{conv}\left(\bigcup_{i=1}^{m} \partial_C f_i(x^*) \right). \tag{2.22}$$

If a vector $x^* \in \mathcal{H}$ satisfies (2.22) *we call it* Pareto critical.

Proof. A proof can be found in [159, Theorem 12]. Note that in [159] the finite-dimensional case ($\mathcal{H} = \mathbb{R}^n$) is considered. However, the proof can also be applied in general Hilbert spaces, since the used arguments only require properties of the generalized directional derivative and the Clarke subdifferential that we state in Propositions 2.1.31, 2.1.35 and Theorem 2.1.36. □

Remark 2.3.11. *Notice that the definitions of Pareto critical points in Theorems 2.3.7 and 2.3.10 are equivalent in the smooth case. If the objective functions f_i are continuously differentiable, Proposition 2.1.34 states that $\partial_C f_i(x) = \{\mathrm{D} f_i(x)\}$ for all $i = 1, \dots, m$. Then, it holds that* $\operatorname{conv}(\bigcup_{i=1}^{m} \partial_C f_i(x)) = \operatorname{conv}(\{\mathrm{D} f_i(x) : i = 1, \dots, m\})$ *and hence* $0 \in \operatorname{conv}(\bigcup_{i=1}^{m} \partial_C f_i(x))$ *if and only if* $0 \in \operatorname{conv}(\{\nabla f_i(x) : i = 1, \dots, m\})$.

Remark 2.3.12. *In the literature, optimality conditions for constrained multiobjective optimization problems and sufficient optimality conditions that rely on second-order derivatives, have been established [131]. In this thesis, we focus on unconstrained optimization problems and we are concerned solely with first-order methods and gradient dynamical systems. Therefore, we do not present more sophisticated results in these directions.*

2.3.3 Merit functions

In scalar optimization, all optimal points yield the same function value. In contrast, in multiobjective optimization the function values of the different objectives vary along the Pareto set. As a result, it is not straightforward to express the optimality of a point solely in terms of its function values. One intuitive approach is to consider the distance of the objective function value to the Pareto front, i.e., $\operatorname{dist}(F(x), F(\mathcal{P})) = \inf_{F^* \in F(\mathcal{P})} \|F(x) - F^*\|$ with a suitable norm $\|\cdot\|$ on $\mathbb{R}^m$, to define a measure of optimality. However, this is impractical since the Pareto front $F(\mathcal{P}) \subset \mathbb{R}^m$ is not convex, and the mapping $x \mapsto \operatorname{dist}(F(x), F(\mathcal{P}))$ suffers in general from bad analytical properties. Despite these challenges, a measure of optimality based on function values is valuable as it provides further insight into the multiobjective optimization problem and serves as a basis to compare the complexity of numerical methods. In this subsection, we introduce a merit function for multiobjective optimization. Generally, a merit (or gap) function for a problem is a scalar function that attains nonnegative values everywhere and vanishes only at solutions. Merit functions for multiobjective optimization were first applied for convex and linearly constrained problems [66], and later extended to more general settings [89, 153, 220].

In this thesis, we use the following merit function related to the multiobjective optimization problem (MOP),

$$\varphi : \mathcal{H} \to \mathbb{R}, \quad x \mapsto \varphi(x) := \sup_{z\in\mathcal{H}} \min_{i=1,\dots,m} \left(f_i(x) - f_i(z)\right). \tag{2.23}$$

The following theorem states that $\varphi(\cdot)$ is a merit function with respect to weak Pareto optimality.

Theorem 2.3.13. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be an arbitrary function for $i = 1, \dots, m$ and let $\varphi(\cdot)$ be defined as in* (2.23)*. For all $x \in \mathcal{H}$ it holds that $\varphi(x) \geq 0$. Moreover, $x \in \mathcal{H}$ is weakly Pareto optimal for* (MOP)*, if and only if $\varphi(x) = 0$.*

Proof. A proof can be found in [224, Theorem 3.1]. □

In the single objective case, i.e., for $m = 1$ and $f_1 := f$, it holds $\varphi(x) = f(x) - \inf_{z\in\mathcal{H}} f(z)$ for all $x \in \mathcal{H}$, which is a merit function for the scalar optimization problem $\min_{x\in\mathcal{H}} f(x)$. This provides another justification for using $\varphi(\cdot)$ as a measure of optimality in multiobjective optimization. One should note that, even if all objective functions f_i are smooth, the function $\varphi(\cdot)$ is not smooth in general.

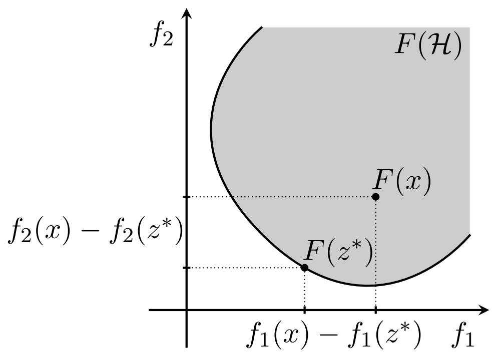


Figure 2.3: Attainable set of a convex multiobjective optimization problem with two objective functions. Given $x \in \mathcal{H}$ the point $z^* \in \mathcal{H}$ is the solution to $\sup_{z\in\mathcal{H}} \min_{i=1,\dots,m} f_i(x) - f_i(z)$ which satisfies $f_i(x) - f_i(z^*) = \varphi(x)$ for all $i = 1, \dots, m$.

Figure 2.3 visualizes how the merit function value $\varphi(x)$ can be computed for a convex multiobjective optimization problem as a solution to the problem $\sup_{z\in\mathcal{H}} \min_{i=1,\dots,m} f_i(x) - f_i(z)$. In the well-behaved case there exists $z^* \in \mathcal{LP}_w(F, F(x))$ such that $f_i(x) - f_i(z^*) = \varphi(x)$ for all $i = 1, \dots, m$. In this case, $\varphi(x) = \operatorname{dist}_{\|\cdot\|_\infty}(F(x), F(\mathcal{P}_w)) = \inf_{F^*\in F(\mathcal{P}_w)} \|F(x) - F^*\|_\infty$, where $\|\cdot\|_\infty$ is the infinity norm on $\mathbb{R}^m$.

Theorem 2.3.14. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be sequentially weakly lower semicontinuous for $i = 1, \dots, m$. Then the function $\varphi(\cdot)$ defined in* (2.23) *is sequentially weakly lower semicontinuous, i.e., for $(x^k)_{k\geq 0}$ with $x^k \rightharpoonup x^\infty \in \mathcal{H}$ it holds that*

$$\varphi(x^\infty) \leq \liminf_{k\to+\infty} \varphi(x^k).$$

Proof. Let $(x^k)_{k\geq 0} \subset \mathcal{H}$ be a sequence with $x^k \rightharpoonup x^\infty$ as $k \to +\infty$. The minimum of a finite family of weakly lower semicontinuous functions is weakly lower semicontinuous, i.e., for all $z \in \mathcal{H}$

$$\min_{i=1,\ldots,m} f_i(x^\infty) - f_i(z) \leq \liminf_{k\to+\infty} \min_{i=1,\ldots,m} f_i(x^k) - f_i(z). \tag{2.24}$$

Since (2.24) holds for all $z \in \mathcal{H}$, we conclude

$$\begin{aligned}\varphi(x^\infty) &= \sup_{z\in\mathcal{H}} \min_{i=1,\ldots,m} f_i(x^\infty) - f_i(z) \leq \sup_{z\in\mathcal{H}} \liminf_{k\to+\infty} \min_{i=1,\ldots,m} f_i(x^k) - f_i(z) \\ &= \sup_{z\in\mathcal{H}} \sup_{l\geq 0} \inf_{k\geq l} \min_{i=1,\ldots,m} f_i(x^k) - f_i(z) = \sup_{l\geq 0} \sup_{z\in\mathcal{H}} \inf_{k\geq l} \min_{i=1,\ldots,m} f_i(x^k) - f_i(z) \\ &\leq \sup_{l\geq 0} \inf_{k\geq l} \sup_{z\in\mathcal{H}} \min_{i=1,\ldots,m} f_i(x^k) - f_i(z) = \sup_{l\geq 0} \inf_{k\geq l} \varphi(x^k) = \liminf_{k\to+\infty} \varphi(x^k).\end{aligned}$$

□

By Theorem 2.3.14, we conclude that every weak accumulation point of a sequence $(x^k)_{k\geq 0}$ with $\lim_{k\to+\infty} \varphi(x^k) = 0$ is weakly Pareto optimal. The following lemma provides a way to obtain the value of $\varphi(x)$ without taking the supremum with respect to the whole space $\mathcal{H}$. This lemma is important in the analysis of first-order methods and gradient dynamical systems for multiobjective optimization. The version of the lemma we use can be found in [49].

Lemma 2.3.15. *For $x_0 \in \mathcal{H}$ and $a \in \mathbb{R}^m_+$, assume that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ holds for all $x \in \mathcal{L}(F, F(x_0) + a)$. Then, for all $x \in \mathcal{L}(F, F(x_0) + a)$ it holds that*

$$\begin{aligned}\varphi(x) &= \sup_{z\in\mathcal{LP}_w(F,F(x_0)+a)} \min_{i=1,\ldots,m} f_i(x) - f_i(z) \\ &= \sup_{F^*\in F(\mathcal{LP}_w(F,F(x_0)+a))} \inf_{z\in F^{-1}(\{F^*\})} \min_{i=1,\ldots,m} f_i(x) - f_i(z).\end{aligned}$$

Proof. We start by proving the first equality. Let $x \in \mathcal{L}(F, F(x_0) + a)$ be fixed. Obviously,

$$\sup_{z\in\mathcal{LP}_w(F,F(x_0)+a)} \min_{i=1,\ldots,m} f_i(x) - f_i(z) \leq \sup_{z\in\mathcal{H}} \min_{i=1,\ldots,m} f_i(x) - f_i(z) = \varphi(x). \tag{2.25}$$

Next, we show that $\min_{i=1,\ldots,m} f_i(x) - f_i(z) \leq \sup_{z'\in\mathcal{L}(F,F(x))} \min_{i=1,\ldots,m} f_i(x) - f_i(z')$ holds for all $z \in \mathcal{H}$. We assume that there exists $z \notin \mathcal{L}(F, F(x))$ with $\min_{i=1,\ldots,m} f_i(x) - f_i(z) > \sup_{z'\in\mathcal{L}(F,F(x))} \min_{i=1,\ldots,m} f_i(x) - f_i(z')$. Since $z \notin \mathcal{L}(F, F(x))$, there exists $j \in \{1, \ldots, m\}$ with $f_j(z) > f_j(x)$. Therefore

$$0 > \min_{i=1,\ldots,m} f_i(x) - f_i(z) \geq \sup_{z'\in\mathcal{L}(F,F(x))} \min_{i=1,\ldots,m} f_i(x) - f_i(z') \geq 0,$$

which leads to a contradiction. Hence,

$$\sup_{z\in\mathcal{H}} \min_{i=1,\ldots,m} f_i(x) - f_i(z) \leq \sup_{z\in\mathcal{L}(F,F(x))} \min_{i=1,\ldots,m} f_i(x) - f_i(z). \tag{2.26}$$

Next, we show that $\sup_{z\in\mathcal{L}(F,F(x))}\min_{i=1,\ldots,m} f_i(x)-f_i(z) \leq \sup_{z\in\mathcal{LP}_w(F,F(x))}\min_{i=1,\ldots,m} f_i(x)-f_i(z)$. By assumption, for all $z\in\mathcal{L}(F,F(x))$ there exists $z'\in\mathcal{LP}_w(F,F(z))\subseteq\mathcal{LP}_w(F,F(x))$. Since $z'\in\mathcal{LP}_w(F,F(z)))$, it holds $f_i(z')\leq f_i(z)$ for all $i=1,\ldots,m$, hence

$$\min_{i=1,\ldots,m} f_i(x)-f_i(z) \leq \min_{i=1,\ldots,m} f_i(x)-f_i(z'). \tag{2.27}$$

From (2.27), we conclude

$$\sup_{z\in\mathcal{L}(F,F(x))}\min_{i=1,\ldots,m} f_i(x)-f_i(z) \leq \sup_{z\in\mathcal{LP}_w(F,F(x))}\min_{i=1,\ldots,m} f_i(x)-f_i(z). \tag{2.28}$$

Since $x\in\mathcal{L}(F,F(x_0)+a)$, we have $\mathcal{LP}_w(F,F(x))\subseteq\mathcal{LP}_w(F,F(x_0)+a)$, hence

$$\sup_{z\in\mathcal{LP}_w(F,F(x))}\min_{i=1,\ldots,m} f_i(x)-f_i(z) \leq \sup_{z\in\mathcal{LP}_w(F,F(x_0)+a)}\min_{i=1,\ldots,m} f_i(x)-f_i(z). \tag{2.29}$$

Combining (2.26), (2.28) and (2.29), it yields

$$\varphi(x) \leq \sup_{z\in\mathcal{LP}_w(F,F(x_0)+a)}\min_{i=1,\ldots,m} f_i(x)-f_i(z), \tag{2.30}$$

which proves the first equality. The second inequality follows since for all $z\in F^{-1}(\{F^*\})$ it holds that $F(z)=F^*$. □

2.3.4 A representative first-order method for multiobjective optimization

In this subsection, we introduce the multiobjective steepest descent method [104, 177]. This method serves as a natural starting point for first-order methods in multiobjective optimization, and the more sophisticated algorithms presented later in this thesis can be seen as advancements of this foundational approach.

The multiobjective steepest descent direction

Before introducing the steepest descent direction, we first describe the general concept of a descent direction. Since we are dealing with multiple objectives, it is necessary to identify a direction that yields local descent with respect to all objectives.

Definition 2.3.16. *Let $f_i:\mathcal{H}\to\mathbb{R}$ be continuously differentiable for $i=1,\ldots,m$ and let $x\in\mathcal{H}$. A vector $v\in\mathcal{H}$ is a called a common descent direction in x if it satisfies*

$$\langle\nabla f_i(x),v\rangle<0 \quad \textit{for all} \quad i=1,\ldots,m.$$

To construct a descent method, a mere descent direction is not sufficient to define an efficient algorithm. A stronger condition is required. In the next definition, we introduce the multiobjective steepest descent direction, as presented in [104, 177].

Definition 2.3.17. *Let $f_i:\mathcal{H}\to\mathbb{R}$ be continuously differentiable for $i=1,\ldots,m$ and let $x\in\mathcal{H}$. Then, the multiobjective steepest descent direction in x is defined as the unique solution to the strongly convex optimization problem*

$$\min_{d\in\mathcal{H}}\max_{i=1,\ldots,m}\langle\nabla f_i(x),d\rangle+\frac{1}{2}\|d\|^2. \tag{2.31}$$

The objective function in (2.31) is a piecewise linear approximation of $d \mapsto \max_{i=1,\dots,m} f_i(x+d) - f_i(x)$ with a quadratic regularization term. The problem (2.31) is nonsmooth, but it can be transformed to a quadratic optimization problem with linear constraints. In the following proposition, we present a duality result to obtain a different formulation of the multiobjective steepest descent direction.

Proposition 2.3.18. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable for $i = 1, \dots, m$ and let $x \in \mathcal{H}$. Let d^* be the the multiobjective steepest descent direction at x. Then it holds that*

$$d^* = -\sum_{i=1}^{m} \theta_i^* \nabla f_i(x),$$

where $\theta^* \in \mathbb{R}^m$ is the solution to

$$\begin{aligned} \min_{\theta \in \mathbb{R}^m} \ & \frac{1}{2} \left\| \sum_{i=1}^{m} \theta_i \nabla f_i(x) \right\|^2, \\ \text{s.t.} \quad & \sum_{i=1}^{m} \theta_i = 1, \\ & \theta_i \geq 0, \quad \textit{for} \quad i = 1, \dots, m. \end{aligned} \tag{2.32}$$

Furthermore, d^* can be written as

$$d^* = -\operatorname*{proj}_{C(x)}(0), \tag{2.33}$$

where

$$C(x) \coloneqq \operatorname{conv}\left(\{\nabla f_i(x) \, : \, i = 1, \dots, m\}\right).$$

Proof. Rewrite the problem (2.31) into

$$\text{(P)} \qquad \begin{aligned} \min_{(d,\beta) \in \mathcal{H} \times \mathbb{R}} \ & \Phi(d, \beta) \coloneqq \beta + \frac{1}{2} \|d\|^2, \\ \text{s.t.} \quad & g_i(d, \beta) \coloneqq \langle \nabla f_i(x), d \rangle - \beta \leq 0, \quad \text{for} \quad i = 1, \dots, m. \end{aligned}$$

Define the Lagrangian function

$$\begin{aligned} \mathcal{L} : (\mathcal{H} \times \mathbb{R}) \times \mathbb{R}^m \to \mathbb{R}, \quad ((d, \beta), \theta) \mapsto \ & \Phi(d, \beta) + \sum_{i=1}^{m} \theta_i g_i(d, \beta) \\ = \ & \beta + \frac{1}{2}\|d\|^2 + \sum_{i=1}^{m} \theta_i \left(\langle \nabla f_i(x), d \rangle - \beta \right) \\ = \ & \left(1 - \sum_{i=1}^{m} \theta_i \right) \beta + \frac{1}{2}\|d\|^2 + \left\langle \sum_{i=1}^{m} \theta_i \nabla f_i(x), d \right\rangle. \end{aligned}$$

The dual problem

$$\inf_{\substack{\theta \in \mathbb{R}^m \\ \text{s.t. } \theta_i \geq 0, \text{ for } i = 1, \dots, m}} \ \sup_{(d, \beta) \in \mathcal{H} \times \mathbb{R}} \mathcal{L}((d, \beta), \theta),$$

reduces to

$$
\text{(D)} \qquad \begin{aligned} \min_{\theta\in\mathbb{R}^m} \ & \frac{1}{2}\left\|\sum_{i=1}^{m}\theta_i\nabla f_i(x)\right\|^2, \\ \text{s.t.} \quad & \sum_{i=1}^{m}\theta_i = 1, \\ & \theta_i \geq 0, \quad \text{for} \quad i=1,\dots,m. \end{aligned}
$$

Since (P) is a convex optimization problem with linear constraints and an objective function which is bounded from below on the feasible set, strong duality holds [155]. Hence, for an optimal solution $d^* \in \mathcal{H}$ to (2.31) which induces the optimal solution (d^*, β^*), with $\beta^* = \max_{i=1,\dots,m}\langle\nabla f_i(x), d^*\rangle$ to (P), there exists a Lagrange multiplier $\theta^* \in \mathbb{R}^m$, such that $((d^*,\beta^*),\theta^*)$ is a KKT point of (P). A KKT point is a saddle point of the Lagrangian and hence $\nabla_d \mathcal{L}((d^*,\theta^*),\beta^*) = 0$, and we conclude $d^* = -\sum_{i=1}^{m}\theta_i^*\nabla f_i(x)$. By strong duality, $\theta^* \in \mathbb{R}^m$ is a solution to (D). By Definition 2.1.15 and Remark 2.1.16, it follows that

$$
d^* = -\operatorname*{proj}_{C(x)}(0).
$$

□

Remark 2.3.19. *Motivated by Proposition 2.3.18, we usually denote the steepest descent direction by* $-\operatorname{proj}_{C(x)}(0)$, *where* $C(x) := \operatorname{conv}(\{\nabla f_i(x) : i=1,\dots,m\})$ *is the convex hull of the gradients.*

Example 2.3.20. *Reconsider the multiobjective optimization problem from Example 2.3.6.*

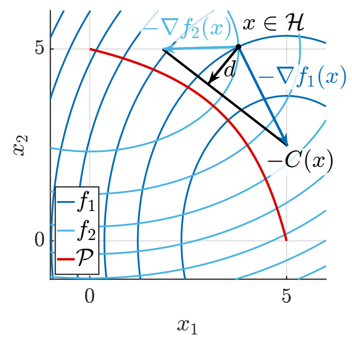


Figure 2.4: Contour plots of f_1 and f_2 and Pareto set $\mathcal{P}$ of (MOP-Ex) with a point $x \in \mathcal{H}$ and corresponding negative gradients $-\nabla f_1(x)$ and $-\nabla f_2(x)$ and the multiobjective steepest descent direction $d = -\operatorname{proj}_{C(x)}(0)$.

Figure 2.4 visualizes the multiobjective steepest descent direction computed in a point $x \in \mathcal{H}$. *The figure shows the the contour plots of the functions* f_1 *and* f_2 *in the decision space and the Pareto set* $\mathcal{P}$. *For* $x \in \mathcal{H}$ *the set* $-C(x)$ *is the convex hull of the negative gradients* $-\nabla f_1(x)$ *and* $-\nabla f_2(x)$. *The steepest descent direction* $d = -\operatorname{proj}_{C(x)}(0)$ *is perpendicular to the set* $-C(x)$.

The following proposition describes the regularity of the multiobjective steepest descent direction $-\operatorname{proj}_{C(x)}(0)$ with respect to changes $x \in \mathcal{H}$.

Proposition 2.3.21. *Let $f_i : \mathcal{H} \to \mathbb{R}^m$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i. Then the function*

$$\phi : \mathcal{H} \to \mathcal{H}, \quad x \mapsto -\operatorname*{proj}_{C(x)}(0),$$

is locally Hölder continuous with exponent $\frac{1}{2}$, i.e., for all $x \in \mathcal{H}$ there exists $\varepsilon > 0$ and a constant $C > 0$, such that

$$\|\phi(y) - \phi(z)\| \leq C\|y - z\|^{\frac{1}{2}}, \quad \text{for all} \quad y, z \in B_\varepsilon(x).$$

Proof. A proof of this proposition is contained in [219]. □

Remark 2.3.22. *The paper [219] not only establishes the Hölder continuity of the multiobjective steepest descent direction with exponent $\frac{1}{2}$, but also shows that this is the optimal exponent in the general case. Consequently the multiobjective steepest descent direction is not Lipschitz continuous in general. A counterexamples to the Lipschitz continuity of the multiobjective steepest descent direction is provided in [219, Proposition 3.2].*

The multiobjective steepest descent direction is directly connected to the concept of Pareto criticality as the following Proposition describes.

Proposition 2.3.23. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable. Then x^* is Pareto critical, if and only if $0 = \operatorname{proj}_{C(x^*)}(0)$.*

Proof. The proof follows immediately by the definition of Pareto critical points given in Theorem 2.3.10. □

Proposition 2.3.23 motivates the investigation of the closedness of the operator $x \mapsto \operatorname{proj}_{C(x)}(0)$. The following lemma states a demiclosedness property of the set-valued mapping $C : \mathcal{H} \rightrightarrows \mathcal{H}$, $x \mapsto C(x)$ (see [24, Lemma 2.4] and [16, Lemma 4.10]).

Lemma 2.3.24. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable for $i = 1, \ldots, m$. 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.*

The multiobjective steepest descent method

Using the multiobjective steepest descent direction introduced previously, we define the so-called multiobjective steepest descent method. Given an initial iterate $x^0 \in \mathcal{H}$ and step size $h > 0$ define the multiobjective steepest descent method by the scheme

$$\text{(MGD)} \qquad \left.\begin{array}{rcl} \theta^k & \in & \arg\min_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i \nabla f_i(x^k) \right\|^2, \\ x^{k+1} & = & x^k - h \sum_{i=1}^m \theta_i^k \nabla f_i(x^k), \end{array}\right\} \quad \text{for} \quad k \geq 0.$$

This method can more concisely be written as

$$\text{(MGD')} \qquad x^{k+1} = x^k - h \operatorname*{proj}_{C(x^k)}(0), \quad \text{for} \quad k \geq 0.$$

While (MGD) is more explicit from an algorithmic point of view, the formulation (MGD') provides valuable insight from an analytical perspective, as the following proposition shows.

Proposition 2.3.25. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for all $i = 1, \ldots, m$, let $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ be the sequence defined by* (MGD)*. Then, for all $i = 1, \ldots, m$ and all $k \geq 0$*

$$f_i(x^{k+1}) \leq f_i(x^k) - \frac{1}{2h}\|x^{k+1} - x^k\|^2.$$

Proof. We use the variational characterization of (MGD'). Since $\nabla f_i(x^k) \in C(x^k)$ and $\frac{1}{h}(x^k - x^{k+1}) = \operatorname{proj}_{C(x^k)}(0)$, we follow from the variational characterization of the projection (see Theorem 2.1.17) for all $i = 1, \ldots, m$ and all $k \geq 0$

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

By Lemma 2.1.25, we have for all $i = 1, \ldots, m$ and all $k \geq 0$

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

Since $0 < h \leq \frac{1}{L}$, we can bound $\frac{L}{2} - \frac{1}{h} \leq -\frac{1}{2h}$ and conclude the proof. □

From Proposition 2.3.25, we derive immediate consequences which we collect in the following corollary.

Corollary 2.3.26. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for all $i = 1, \ldots, m$, let $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ be the sequence defined by* (MGD)*. Then, the following statements hold:*

i) For all $i = 1, \ldots, m$, the limit $\lim_{k\to+\infty} f_i(x^k) \in \mathbb{R} \cup \{-\infty\}$ exists;

ii) The sequence $(\varphi(x^k))_{k\geq 0}$ is monotonically decreasing and the limit $\lim_{k\to+\infty} \varphi(x^k)$ exists;

iii) Assume there exists $i \in \{1, \ldots, m\}$ such that the level set $\left\{x \in \mathcal{H} : f_i(x) \leq f_i(x^0)\right\}$ is bounded. Then, the sequence $(x^k)_{k\geq 0}$ remains bounded;

iv) Assume there exists $i \in \{1, \ldots, m\}$ such that $\lim_{k\to+\infty} f_i(x^k) > -\infty$. Then,

$$\sum_{k=0}^{\infty}\|x^{k+1} - x^k\|^2 < +\infty.$$

Proposition 2.3.25 shows that (MGD) is a common descent method, i.e., for all $i = 1, \dots, m$ it holds that $(f_i(x^k))_{k \geq 0}$ is a monotonically decreasing sequence. In the following theorem, we describe more concrete asymptotical results.

Theorem 2.3.27. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for all $i = 1, \dots, m$, let $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k \geq 0}$ be the sequence defined by* (MGD)*. Then, for all $z \in \mathcal{H}$ it holds that*

$$\lim_{k \to +\infty} \min_{i=1,\dots,m} f_i(x^k) - f_i(z) \leq 0.$$

Proof. Let $z \in \mathcal{H}$. By Lemma 2.1.25, for all $i = 1, \dots, m$ and all $k \geq 0$

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

We apply the minimum and derive for all $k \geq 0$

$$\begin{aligned} &\min_{i=1,\dots,m} f_i(x^{k+1}) - f_i(z) \leq \sum_{i=1}^{m} \theta_i^k (f_i(x^{k+1}) - f_i(z)) \\ \leq &\left\langle \sum_{i=1}^{m} \theta_i^k \nabla f_i(x^k), x^{k+1} - z \right\rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ = &\frac{1}{h} \left\langle x^k - x^{k+1}, x^{k+1} - z \right\rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ = &\frac{1}{2h} \left[\|x^k - z\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - z\|^2 \right] + \frac{L}{2} \|x^{k+1} - x^k\|^2. \end{aligned} \tag{2.34}$$

We use $0 < h \leq \frac{1}{L}$ and obtain from (2.34)

$$\min_{i=1,\dots,m} f_i(x^{k+1}) - f_i(z) \leq \frac{1}{2h} \left[\|x^k - z\|^2 - \|x^{k+1} - z\|^2 \right]. \tag{2.35}$$

Since $f_i(x^k)$ is monotonically decreasing for all $i = 1, \dots, m$, we follow that for all $z \in \mathcal{H}$ the sequence $\min_{i=1,\dots,m} f_i(x^{k+1}) - f_i(z)$ is monotonically decreasing and hence for all $k \geq 0$

$$\min_{i=1,\dots,m} f_i(x^{k+1}) - f_i(z) \leq \min_{i=1,\dots,m} f_i(x^0) - f_i(z). \tag{2.36}$$

We combine (2.35) and (2.36) to get for all $z \in \mathcal{H}$ and all $k \geq 0$

$$\begin{aligned} (k+1) \min_{i=1,\dots,m} f_i(x^{k+1}) - f_i(z) &\leq \sum_{l=0}^{k} f_i(x^{l+1}) - f_i(z) \leq \frac{1}{2h} \sum_{l=0}^{k} \left[\|x^l - z\|^2 - \|x^{l+1} - z\|^2 \right] \\ &\leq \frac{1}{2h} \left[\|x^0 - z\|^2 - \|x^{k+1} - z\|^2 \right] \leq \frac{1}{2h} \|x^0 - z\|^2. \end{aligned}$$

Therefore, we have for all $k \geq 1$

$$\min_{i=1,\dots,m} f_i(x^k) - f_i(z) \leq \frac{\|x^0 - z\|^2}{2hk}. \tag{2.37}$$

By taking the limit for $k \to +\infty$, the statement holds. □

Theorem 2.3.27 describes that there does not exist a $z \in \mathcal{H}$ which strictly dominates x^k in the limit. Under an additional assumption on the objective functions we can derive the following refined version of this theorem.

Theorem 2.3.28. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for all $i = 1, \ldots, m$, let $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ be the sequence defined by* (MGD)*. Assume the functions f_i are bounded from below for $i = 1, \ldots, m$. Then,*

$$\lim_{k\to+\infty} \varphi(x^k) = 0.$$

Proof. By Corollary 2.3.26, $\varphi^\infty := \lim_{k\to+\infty} \varphi(x^k) \in \mathbb{R} \cup \{+\infty\}$ exists. Since the functions f_i are bounded from below for all $i = 1, \ldots, m$, it holds that $\varphi^\infty < +\infty$. We give a proof by contradiction to show $\varphi^\infty = 0$. Assume $\varphi^\infty > 0$. By the definition of $\varphi(\cdot)$ for all $k \geq 0$ there exists $z^k \in \mathcal{H}$ with

$$\min_{i=1,\ldots,m} f_i(x^k) - f_i(z^k) > \frac{\varphi^\infty}{2}.$$

Since $(f_i(x^k))_{k\geq 0}$ is monotonically decreasing for all $i = 1, \ldots, m$ by Proposition 2.3.25 and since f_i is bounded from below by assumption, it holds that $f_i^\infty := \lim_{k\to+\infty} f_i(x^k) \in \mathbb{R}$ exists for all $i = 1, \ldots, m$. For all $a, b \in \mathbb{R}^m$ it holds that $\min_{i=1,\ldots,m} a_i \leq \max_{i=1,\ldots,m}(a_i - b_i) + \min_{i=1,\ldots,m} b_i$. By this, we obtain

$$\frac{\varphi^\infty}{2} < \min_{i=1,\ldots,m} f_i(x^k) - f_i(z^k) \leq \max_{i=1,\ldots,m} f_i(x^k) - f_i^\infty + \min_{i=1,\ldots,m} f_i^\infty - f_i(z^k). \tag{2.38}$$

Since $f_i(x^k) - f_i^\infty \to 0$ as $k \to +\infty$ for all $i = 1, \ldots, m$ there exists $K \geq 0$ such that

$$\max_{i=1,\ldots,m} f_i(x^K) - f_i^\infty \leq \frac{\varphi^\infty}{4}. \tag{2.39}$$

Combining (2.38) and (2.39), we get

$$\frac{\varphi^\infty}{4} < \min_{i=1,\ldots,m} f_i^\infty - f_i(z^K) = \lim_{k\to+\infty} f_i(x^k) - f_i(z^K),$$

which contradicts Theorem 2.3.27. □

Building on Theorems 2.3.27 and 2.3.28, we derive asymptotic convergence rates for the method (MGD) with respect to the merit function values $\varphi(x^k)$. To obtain these rates, we require the following technical assumption. We postpone a detailed discussion of this assumption to later sections, where it is applied to more involved first-order methods and gradient dynamical systems in multiobjective optimization, as our goal for this subsection is merely to summarize the most important results.

($\mathcal{A}$) For all $x_0 \in \mathcal{H}$ and for all $x \in \mathcal{L}(F, F(x_0))$ it holds that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ and further

$$R := \sup_{F^* \in F(\mathcal{LP}_w(F,F(x_0)))} \inf_{z\in F^{-1}(\{F^*\})} \frac{1}{2}\|z - x^0\|^2 < +\infty. \tag{2.40}$$

Theorem 2.3.29. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for all $i = 1, \dots, m$, let $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ be the sequence defined by* (MGD)*. Assume that Assumption $(\mathcal{A})$ holds. Then for all $k \geq 1$, it holds that*

$$\varphi(x^k) \leq \frac{R}{hk},$$

where $R > 0$ is defined in Assumption $(\mathcal{A})$.

Proof. The result follows immediately from the proof of Theorem 2.3.27. Inequality 2.34 states for all $z \in \mathcal{H}$ and $k \geq 1$

$$\min_{i=1,\dots,m} f_i(x^k) - f_i(z) \leq \frac{\|x^0 - z\|^2}{2hk}.$$

If we apply the supremum and infimum as in $(\mathcal{A})$ and use Lemma 2.3.15 with $a = 0$, we obtain

$$\begin{aligned}\varphi(x^k) &= \sup_{F^* \in F(\mathcal{LP}_w(F, F(x_0)))} \inf_{z \in F^{-1}(\{F^*\})} \min_{i=1,\dots,m} f_i(x^k) - f_i(z) \\ &\leq \sup_{F^* \in F(\mathcal{LP}_w(F, F(x_0)))} \inf_{z \in F^{-1}(\{F^*\})} \frac{\|x^0 - z\|^2}{2hk} = \frac{R}{hk}.\end{aligned}$$

□

We conclude the analysis of the multiobjective steepest descent method with the following theorem which states that the sequence $(x^k)_{k\geq 0}$ given by (MGD) converges weakly to a weakly Pareto optimal point of (MOP) under a boundedness assumption.

Theorem 2.3.30. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for all $i = 1, \dots, m$, let $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ be the sequence defined by* (MGD)*. Assume the sequence $(x^k)_{k\geq 0}$ remains bounded. Then x^k converges weakly to a weakly Pareto optimal point of* (MOP)*.*

Proof. Define the set

$$S := \left\{ z \in \mathcal{H} \,:\, f_i(z) \leq \lim_{k\to+\infty} f_i(x^k) \quad \text{for all} \quad i = 1, \dots, m \right\}.$$

Note that the limit $\lim_{k\to+\infty} f_i(x^k) \in \mathbb{R}$ exists for all $i = 1, \dots, m$ by the boundedness of $(x^k)_{k\geq 0}$ and the fact that $f_i(x^k)$ is monotonically decreasing by Proposition 2.3.25. We will apply Opial's Lemma (Lemma 2.1.5) to prove that x^k converges weakly to an element $x^\infty \in \mathcal{H}$. In a subsequent step, we show that $x^\infty \in \mathcal{P}_w$.

We begin by verifying the assumptions required for Opial's Lemma. Since $(x^k)_{k\geq 0}$ is bounded it posses at least one sequential cluster point $x^\infty \in \mathcal{H}$, i.e., there exists a monotonically increasing subsequence $(k_l)_{l\geq 0}$ with $x^{k_l} \rightharpoonup x^\infty$ as $l \to +\infty$. By the weak lower semicontinuity of the objective functions we follow for all $i = 1, \dots, m$

$$f_i(x^\infty) \leq \liminf_{l\to+\infty} f_i(x^{k_l}) = \lim_{k\to+\infty} f_i(x^k).$$

Therefore, $x^\infty \in S$ and by the same argument each weak sequential cluster point of $(x^k)_{k\geq 0}$ belongs to S. Let $z \in S$. By (2.35), we follow for all $k \geq 0$

$$\frac{1}{2h}\left[\|x^{k+1} - z\|^2 - \|x^k - z\|^2\right] \leq - \min_{i=1,\ldots,m} f_i(x^{k+1}) - f_i(z).$$

Since $z \in S$, it holds that $\min_{i=1,\ldots,m} f_i(x^{k+1}) - f_i(z) \geq 0$ for all $i = 1, \ldots, m$ and therefore for all $k \geq 0$

$$\|x^{k+1} - z\| \leq \|x^k - z\|.$$

Hence, for all $z \in S$, the limit $\lim_{k\to+\infty}\|x^k - z\|$ exists. Then by Opial's Lemma (Lemma 2.1.5) it follows that x^k converges weakly to an element in S, i.e., $x^k \rightharpoonup x^\infty \in S$ as $k \to +\infty$. Since x^k is bounded, it holds that $f_i(x^k)$ is uniformly bounded from below for all $k \geq 0$ and $i = 1, \ldots, m$. Then, by Corollary 2.3.26, we have

$$\sum_{k=0}^{\infty}\left\|\operatorname*{proj}_{C(x^k)}(0)\right\|^2 = \frac{1}{h^2}\sum_{k=0}^{\infty}\|x^{k+1} - x^k\|^2 < +\infty.$$

Therefore, $\operatorname{proj}_{C(x^k)}(0) \to 0$ as $k \to +\infty$. By Lemma 2.3.24, it follows that x^∞ is Pareto critical and hence weakly Pareto optimal due to the convexity of the objective functions. $\square$

Chapter 3

A descent method for nonconvex locally Lipschitz continuous multiobjective optimization

In this chapter, we consider the multiobjective optimization problem

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

with objective functions $f_i : \mathcal{H} \to \mathbb{R}$ for $i = 1, \ldots, m$. Naturally, there are applications, where the objectives f_i feature nonsmoothness. For example, in [168], an obstacle problem with an elastic string is considered, where one objective is maximization of the contact area between the string and a given obstacle and another objective is minimization of the total force applied to the string.

There is a vast amount of methods available for solving various types of *finite*-dimensional optimization problems, but while most of them are designed to deal with *either* nonsmoothness [28] *or* multiple objectives [169, 92], algorithms for nonsmooth multiobjective optimization problems are scarce. Two possible methods designed for nonsmooth multiobjective optimization problems are the *proximal bundle method* [158, 160] and the *gradient sampling method* [107].

Combining nonsmoothness, multiple objectives *and* an infinite-dimensional Hilbert space setting becomes additionally challenging. When presented with such a nonsmooth multiobjective optimization problem in infinite dimensions, there are several options to proceed, among them:

1. Discretize the infinite-dimensional nonsmooth multiobjective optimization problem and then use a solver for finite-dimensional problems, e.g., one of those presented in [158, 160, 107].
2. Scalarize the problem and then use a solver for infinite-dimensional nonsmooth scalar optimization, e.g., [169, 37].
3. Design a method that is capable of treating infinite dimensions, nonsmooth objective functions and multiple objectives at the same time.

Option 1 does not incorporate the underlying infinite-dimensional problem's topology and can therefore suffer from mesh-dependent behavior such as inconsistent termination criteria between different meshes; see, e.g., the discussion in [132, Sections 3.2.2-3.2.4]. Option 2, as in the smooth case, struggles in the presence of nonconvexity or when the number of objectives exceeds two. Option 3 suffers from neither of these drawbacks but is technically challenging to realize. Previously, infinite-dimensional nonsmooth multiobjective optimization problems have been mostly addressed under additional assumptions on the structure, such as convexity or composite structure (e.g., [41, 43, 115]). To the best of our knowledge, the method introduced in [215], which we discuss in this chapter, presents the first nonscalarizing method for solving general, unstructured nonsmooth infinite-dimensional multiobjective optimization problems.

In this chapter, we generalize the common descent method based on subderivative sampling presented in [107] from finite-dimensional to infinite-dimensional (Hilbert space) settings. The main idea in [107] is to replace the Clarke subdifferential [75] in the design of the descent direction in the dynamic gradient approach of [17] with the Goldstein ε-subdifferential [113], and to approximate the latter via an adaptive gradient sampling scheme. This way, a descent direction for nonsmooth multiobjective optimization problems can be computed. Combining this descent direction with an Armijo-backtracking-type step size control yields a descent method, for which convergence to points satisfying a necessary optimality condition has been shown. This algorithmic approach can be extended to a general Hilbert space setting in a relatively straightforward manner, but the convergence analysis of the algorithm requires modifications to account for the loss of compactness of bounded and closed sets. Additionally, the notions of optimality employed in [107] will be adapted. While the Clarke subdifferential and the Goldstein ε-subdifferential have already been defined on Hilbert spaces [75, 164, 163], their multiobjective counterparts require additional attention. We generalize these objects and prove that they satisfy a generalized demiclosedness property, and employ them in the derivation of necessary conditions for Pareto optimality.

This chapter is organized as follows. In Section 3.1, we extend the Goldstein ε-subdifferential to the multiobjective, infinite-dimensional setting and investigate its properties. Theorem 3.1.10 describes a demiclosedness property of the multiobjective ε-subdifferential, which is important for the convergence proof of the introduced method. The main results of this chapter are presented in Section 3.2. First, in Subsection 3.2.1 we describe how descent directions satisfying a sufficient descent property for all objective functions can be obtained theoretically using the generalized subdifferential from the previous section. In Subsection 3.2.2, we present an algorithm to efficiently compute such descent directions (under the assumption that at least one subderivative can be computed at every point) and prove its feasibility. Using this algorithm, we introduce a descent method for locally Lipschitz continuous multiobjective optimization problems in general Hilbert spaces (Algorithm 3) in Subsection 3.2.3. We prove that this method generates sequences of iterates with Pareto critical accumulation points in Theorem 3.2.10. In Section 3.3, we demonstrate and analyze the behavior of our method in application to a multiobjective obstacle problem on a two-dimensional domain.

The content of this chapter was already published in the following paper:

[215] Sonntag, K., Gebken, B., Müller, G., Peitz, S., and Volkwein, S. *A descent method for nonsmooth multiobjective optimization in Hilbert spaces.* In: *Journal of Optimization Theory and Applications* 203 (1) (2024), pp. 455–487. doi: 10.1007/s10957-024-02520-4.

3.1 Generalized derivatives

In this section, we introduce a generalization of the Clarke subdifferential for the class of locally Lipschitz continuous functions $f : \mathcal{H} \to \mathbb{R}$. In Subsection 3.1.1 we introduce the so-called Goldstein ε-subdifferential, and then extend it to the multiobjective setting in Subsection 3.1.2.

3.1.1 Goldstein ε-subdifferential

In finite dimensions, $\partial_C f(x)$ is the convex hull of the limits of the derivatives of f in all sequences (where the derivatives are defined) that converge to x. Thus, if we evaluate the Fréchet derivative $\mathrm{D}f$ in a number of points close to x (where it is defined) and take the convex hull, we expect the resulting set to be an approximation of $\partial_C f(x)$. To formalize this, we introduce the following definition (see [113, 138]).

Definition 3.1.1. *Let $f : \mathcal{H} \to \mathbb{R}$, $\varepsilon \geq 0$ and $x \in \mathcal{H}$. Then, we define the* (Goldstein) ε-subdifferential of f in x *by*

$$\partial_C^{\varepsilon} f(x) := \overline{\operatorname{conv}}^* \left(\bigcup_{y \in \overline{B_\varepsilon(x)}} \partial_C f(y) \right),$$

which is the weak-closure of the convex hull of the union of the Clarke subdifferentials of the objective function f evaluated in the closed ball centered at x with radius ε. We call $\xi \in \partial_C^{\varepsilon} f(x)$ an* ε-subderivative.

Note that $\partial_C^0 f(x) = \partial_C f(x)$ and $\partial_C f(x) \subseteq \partial_C^{\varepsilon} f(x)$ for all $\varepsilon > 0$.

Proposition 3.1.2. *Let $x \in \mathcal{H}$ and let $f : \mathcal{H} \to \mathbb{R}$ be globally Lipschitz continuous on the ball $B_{\bar{\varepsilon}}(x)$ for some $\bar{\varepsilon} > 0$. Moreover, suppose that $\varepsilon \in [0, \bar{\varepsilon})$. Then, $\partial_C^{\varepsilon} f(x)$ is nonempty, convex and weak*-compact.*

Proof. For $\partial_C^{\varepsilon} f(x)$, the claim was shown in [164, Proposition 2.3]. To apply the proof we need a neighbourhood of $\overline{B_\varepsilon(x)}$, where f is globally Lipschitz continuous. For that reason we introduce the open ball $B_{\bar{\varepsilon}}(x) \supsetneq \overline{B_\varepsilon(x)}$ in the formulation of this proposition. □

In the following, we present a theorem that is a stronger version of parts *ii)* and *iii)* of Proposition 2.1.35. This result relates the ε-subdifferential to the Clarke subdifferential. Before we state the theorem we prove a preparatory lemma.

Lemma 3.1.3. *Let $f : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous in $x \in \mathcal{H}$, $v \in \mathcal{H} \setminus \{0\}$ and $\alpha \in \mathbb{R}$. If*

$$\alpha > \xi(v) \quad \textit{for all} \quad \xi \in \partial_C f(x), \tag{3.1}$$

then there exists an $\bar{\varepsilon} > 0$, such that for all $0 \le \varepsilon \le \bar{\varepsilon}$

$$\alpha > \xi(v) \quad \textit{for all} \quad \xi \in \partial_C^{\varepsilon} f(x). \tag{3.2}$$

Proof of Lemma 3.1.3. We do not show (3.2) directly but first conclude that the separation holds in the weaker form of

$$\alpha > \xi(v) \quad \text{for all} \quad \xi \in \bigcup_{y \in \overline{B_\varepsilon(x)}} \partial_C f(y) \subset \partial_C^{\varepsilon} f(x), \tag{3.3}$$

which is a consequence of Proposition 2.1.35 as we prove in the following.

Let $v \in \mathcal{H}\setminus\{0\}$ and $\alpha \in \mathbb{R}$. Assume for all $\bar{\varepsilon} > 0$ there exists an $\varepsilon \in (0, \bar{\varepsilon}]$ and $\xi \in \bigcup_{y \in \overline{B_\varepsilon(x)}} \partial_C f(x)$ with $\xi(v) \ge \alpha$. Then, there exist a sequence $(\varepsilon_k)_{k \ge 0}$ of positive real numbers and sequences $(y^k)_{k\ge 0}$ and $(\xi^k)_{k \ge 0}$ of elements in $\mathcal{H}$ and $\mathcal{H}^*$, respectively, such that ε_k converges to zero as $k \to +\infty$, $\xi^k \in \partial_C f(y^k)$, $\xi^k(v) \ge \alpha$ for all $k \ge 0$ and $\|y^k - x\| < \varepsilon_k$ converges to zero. Since f is locally Lipschitz continuous in x, there exists an $K \ge 0$ such that for all $k \ge K$ the mapping f is locally L-Lipschitz continuous in y^k. Then, Proposition 2.1.33 *i)* states that for all $k \ge K$ the elements of the sequence $(\xi^k)_{k\ge 0}$ are contained in the weak*-compact set $\overline{B_L(0)} \subset \mathcal{H}^*$. Therefore, the sequence $(\xi^k)_{k \ge 0}$ has a sequential weak*-accumulation point ξ^*. By Proposition 2.1.35 *ii)* the point ξ^* is an element of $\partial_C f(x)$. Since $\xi^k(v) \ge \alpha$ for all $k \ge 0$ we get by the weak*-convergence of a subsequence of $(\xi^k)_{k\ge 0}$ to ξ^*, that $\xi^*(v) \ge \alpha$ which is a contradiction to (3.1). Therefore, (3.3) holds.

The remainder of the proof follows by the definition of the ε-subdifferential (see Definition 3.1.1). If a set lies on one side of a hyperplane, then also its convex hull lies on that side and also its closure. □

Remark 3.1.4. *Lemma 3.1.3 states that the ε-subdifferential contracts in a well-behaved manner to the Clarke subdifferential as $\varepsilon \to 0$. In view of Proposition 2.1.35 iii) this lemma states that we do not have to take the full set-valued limit to contract the ε-subdifferential to one side of the hyperplane defined by v and α.*

Theorem 3.1.5. *Let $f : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous in $x \in \mathcal{H}$. Then, the following statements hold:*

i) *Let $(x^k)_{k\ge 0}$ be a sequence in $\mathcal{H}$ converging to x and $(\varepsilon_k)_{k\ge 0}$ a sequence in $\mathbb{R}_{>0}$ tending to 0. Suppose that the sequence $(\xi^k)_{k\ge 0}$ satisfies $\xi^k \in \partial_C^{\varepsilon_k} f(x^k)$ for all $k \ge 0$. Let ξ be a sequential weak*-accumulation point of $(\xi^k)_{k \ge 0}$. Then, $\xi \in \partial_C f(x)$;*

ii) $\partial_C f(x) = \bigcap_{\varepsilon > 0} \partial_C^{\varepsilon} f(x)$.

Proof. *i)* Since $\xi^k \in \partial_C^{\varepsilon_k} f(x^k)$ it follows that $\xi^k \in \partial_C^{\kappa_k} f(x)$, with $\kappa_k := \varepsilon_k + \|x^k - x\|$. Assume $\xi \notin \partial_C f(x)$. Then, since $\partial_C f(x)$ is convex and weak*-compact, it is closed and the strict separation theorem states that there exists $v \in \mathcal{H} \setminus \{0\}$ and $\alpha \in \mathbb{R}$ satisfying

$$\xi(v) > \alpha > \eta(v) \quad \text{for all} \quad \eta \in \partial_C f(x).$$

Since κ_k converges to 0, Lemma 3.1.3 states that there exists an $K \geq 0$ such that

$$\xi(v) > \alpha > \eta(v) \quad \text{for all} \quad \eta \in \partial_C^{\kappa_k} f(x), \quad k \geq K,$$

and hence

$$\xi(v) > \alpha > \xi^k(v) \quad \text{for all} \quad k \geq K.$$

This is a contradiction to the fact that ξ is a sequential weak*-accumulation point of $(\xi^k)_{k\geq 0}$.

ii) From Proposition 2.1.35 *iii)*, we immediately get the inclusion

$$\partial_C f(x) = \bigcap_{\varepsilon>0} \bigcup_{y\in\overline{B_\varepsilon(x)}} \partial_C f(y) \subseteq \bigcap_{\varepsilon>0} \overline{\operatorname{conv}}\Big(\bigcup_{y\in\overline{B_\varepsilon(x)}} \partial_C f(y)\Big) = \bigcap_{\varepsilon>0} \partial_C^\varepsilon f(x).$$

The other inclusion is a consequence of Lemma 3.1.3 and we prove it analogously to part *i)*. Assume that $\xi \in \cap_{\varepsilon>0}\partial_C^\varepsilon f(x)$, but $\xi \notin \partial_C f(x)$. Then, since $\partial_C f(x)$ is convex and weak*-compact and therefore closed, the strict separation theorem states that there exist $v \in \mathcal{H} \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that

$$\xi(v) > \alpha > \eta(v) \quad \text{for all} \quad \eta \in \partial_C f(x).$$

Lemma 3.1.3 states that there exists an $\varepsilon > 0$ such that

$$\xi(v) > \alpha > \eta(v) \quad \text{for all} \quad \eta \in \partial_C^\varepsilon f(x)$$

and hence $\xi \notin \partial_C^\varepsilon f(x)$. Therefore, it follows that $\xi \notin \bigcap_{\varepsilon>0} \partial_C^\varepsilon f(x)$, which is a contradiction. In total, we derive $\cap_{\varepsilon>0}\partial_C^\varepsilon f(x) \subseteq \partial_C f(x)$ which completes the proof.

□

3.1.2 Multiobjective ε-subdifferential

In this subsection, we extend the Goldstein ε-subdifferential to the multiobjective setting and investigate its main properties.

Definition 3.1.6. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous for $i = 1, \dots, m$ and let $\varepsilon \geq 0$ and $x \in \mathcal{H}$. Then, we define the multiobjective ε-subdifferential by*

$$C^\varepsilon : \mathcal{H} \rightrightarrows \mathcal{H}, \quad x \mapsto C^\varepsilon(x) := \overline{\operatorname{conv}}^* \left(\bigcup_{i=1}^m \partial_C^\varepsilon f_i(x) \right),$$

which is the weak-closure of the convex hull of the union of the Goldstein ε-subdifferentials of the objective function f_i evaluated in x.*

For a visualization of $C^\varepsilon(x)$ in case of a finite-dimensional decision space, we refer to Example 3.1 in [107]. We use the multiobjective ε-subdifferential to give an approximate notion of criticality with the following definition. Additionally, in the descent method we present in the next section the step directions are defined using convex combinations of ε-subderivatives which are elements in $C^\varepsilon(x)$.

Definition 3.1.7. *We say that $x \in \mathcal{H}$ is* (ε, δ)-critical *for constants $\varepsilon \geq 0$ and $\delta \geq 0$, if there exists a $\xi \in C^\varepsilon(x)$ with $\|\xi\|_* \leq \delta$, or equivalently if $\left\|\mathrm{proj}_{C^\varepsilon(x)}(0)\right\|_* \leq \delta$.*

Lemma 3.1.8. *The convex hull of a finite union of convex, weak*-compact sets is weak*-compact.*

Proof. Although the proof utilizes standard arguments, we state it here for the sake of completeness. Let $A^i \subseteq \mathcal{H}^*$ be nonempty, convex and weak*-compact for all $i = 1, \ldots, m$ and set $A := \mathrm{conv}\left(\cup_{i=1}^m A^i\right)$. Let $(\xi^k)_{k\geq 0}$ be an arbitrary sequence in A. The sets A^i are convex and therefore we can write

$$\xi^k = \sum_{i=1}^m \lambda_i^k \xi_i^k \quad \text{for all} \quad k \geq 0,$$

with $\lambda^k = (\lambda_1^k, \ldots, \lambda_m^k)^\top \in \Delta^m$ and $\xi_i^k \in A^i$. Since Δ^m is compact and the sets A^i are sequentially weak*-compact, there exists a subsequence $(k_l)_{l\geq 0}$ such that $\lambda^{k_l} \to \lambda^* \in \Delta^m$ and $\xi_i^{k_l} \rightharpoonup^* \xi_i^* \in A^i$ as $l \to +\infty$ for all $i = 1, \ldots, m$. Then, ξ^{k_l} converges to $\xi^* = \sum_{i=1}^m \lambda_i^* \xi_i^* \in \mathrm{conv}\left(\cup_{i=1}^m A_i\right)$ in the weak*-topology, which completes the proof. □

Now, we formulate the following result analogously to Proposition 3.1.2.

Proposition 3.1.9. *Let $x \in \mathcal{H}$, $0 \leq \varepsilon < \bar{\varepsilon}$ and let $f_i : \mathcal{H} \to \mathbb{R}$ be globally Lipschitz continuous on $B_{\bar{\varepsilon}}(x)$ for $i = 1, \ldots, m$. Then, $C^\varepsilon(x)$ is nonempty, convex and weak*-compact. Furthermore,*

$$C^\varepsilon(x) = \mathrm{conv}\left(\bigcup_{i=1}^m \partial_C^\varepsilon f_i(x)\right),$$

in other words, the closure in Definition 3.1.6 *is superfluous in this case.*

Proof. The proof follows from Proposition 3.1.2 and Lemma 3.1.8. □

The following theorem extends Theorem 3.1.5 to the multiobjective setting.

Theorem 3.1.10. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous in $x \in \mathcal{H}$ for $i = 1, \ldots, m$. Let $(\varepsilon_k)_{k\geq 0}$ be a sequence of positive numbers that converges to 0. Let $(x^k)_{k\geq 0}$ and $(\xi^k)_{k\geq 0}$ be sequences in $\mathcal{H}$ and $\mathcal{H}^*$, respectively, and assume that x^k converges to x and that ξ^k converges to ξ in the weak*-topology as $k \to +\infty$. Further, assume that $\xi^k \in C^{\varepsilon_k}(x_k)$ for all $k \geq 0$. Then,*

$$\xi \in C^0(x) = \mathrm{conv}\left(\bigcup_{i=1}^m \partial_C f_i(x)\right).$$

Proof. Since the functions f_i are locally Lipschitz continuous for $i = 1, \dots, m$, there exists $\varepsilon > 0$ such that f_i is L-Lipschitz continuous on $B_\varepsilon(x)$ for all $i = 1, \dots, m$. Similar to the proof of Theorem 3.1.5 we define $\kappa_k := \varepsilon_k + \|x^k - x\|$ for all $k \geq 0$ and fix $K \geq 0$ such that for all $k \geq K$ it holds that $\kappa_k \leq \varepsilon$. From $\partial_C^{\varepsilon_k} f(x^k) \subseteq \partial_C^{\kappa_k} f(x)$ it follows that $C^{\varepsilon_k}(x^k) \subseteq C^{\kappa_k}(x)$. Proposition 3.1.9 implies that $C^{\kappa_k}(x)$ is nonempty, convex and weak*-compact and

$$C^{\kappa_k}(x) = \operatorname{conv}\left(\bigcup_{i=1}^m \partial_C^{\kappa_k} f_i(x)\right).$$

The remainder of the proof can be seen as a combination of the proofs of Theorem 3.1.5 and Proposition 3.1.9. Since ξ^k is an element of $C^{\kappa_k}(x)$ for all $k \geq K$ it can be written as

$$\xi^k = \sum_{i=1}^m \lambda_i^k \xi_i^k,$$

with $\lambda^k = (\lambda_1^k, \dots \lambda_m^k) \in \Delta^m$ and $\xi_i^k \in \partial_C^{\kappa_k} f_i(x)$. Since $\kappa_k \leq \varepsilon$ it follows that ξ_i^k is contained in the weak*-compact set $\overline{B_L(0)} \subset \mathcal{H}^*$. Hence, there exists a subsequence $(k_l)_{l \geq 0}$ such that

$$\lambda^{k_l} \to \lambda^* \in \Delta^m \quad \text{and} \quad \xi_i^{k_l} \rightharpoonup^* \xi_i^* \in \overline{B_L(0)} \quad \text{for all} \quad i = 1, \dots, m \quad \text{as} \quad l \to +\infty.$$

From Theorem 3.1.5, it follows that $\xi_i^* \in \partial_C f_i(x)$. Then, $\xi^{k_l} = \sum_{i=1}^m \lambda_i^{k_l} \xi_i^{k_l}$ converges to $\xi^* = \sum_{i=1}^m \lambda_i^* \xi_i^* \in \operatorname{conv}(\cup_{i=1}^m \partial_C f_i(x))$ in the weak*-topology as $l \to +\infty$. Since the this limit is unique and ξ^k converges to ξ in the weak*-topology, the proof is complete. □

The next corollary follows directly from Theorem 3.1.10 and gives a sufficient condition for a point to be Pareto critical.

Corollary 3.1.11. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous in $x \in \mathcal{H}$ for $i = 1, \dots, m$. Assume that*

$$0 \in C^\varepsilon(x) \quad \textit{for all} \quad \varepsilon > 0.$$

Then x is Pareto critical, i.e.,

$$0 \in \operatorname{conv}\left(\bigcup_{i=1}^m \partial_C f_i(x)\right).$$

After describing the optimality conditions for (MOP), we now move towards the algorithms from [107] that we extend to the infinite-dimensional setting.

3.2 Derivation of the descent method

In this section, we present a line-search based *common descent method*, meaning that, starting from a point $x^0 \in \mathcal{H}$, we generate a sequence $(x^k)_{k \geq 0}$ in $\mathcal{H}$ in which each point is an improvement over the previous point with respect to all objective functions, that is

$$f_i(x^{k+1}) < f_i(x^k) \quad \text{for all} \quad k \geq 0 \quad \text{and} \quad i = 1, \dots, m,$$

and where $x^{k+1} = x^k + t_k v^k$ for a search direction $v^k := R^{-1}(\xi^k)$ generated from a dual element $\xi^k \in \mathcal{H}^*$ and corresponding step lengths $t_k > 0$. The critical computation of the search direction generalizes the method from [107] to the infinite-dimensional setting.

3.2.1 Descent directions obtained from the multiobjective ε-subdifferential

The foundation of our approach is the following result from convex analysis.

Theorem 3.2.1. *Let $\Xi \subseteq \mathcal{H}^*$ be convex and closed. Then,*

$$\bar{\xi} \coloneqq \underset{\xi \in -\Xi}{\arg\min} \|\xi\|_*^2 \tag{3.4}$$

is well-defined and unique. Further, it holds that either $\bar{\xi} \neq 0$ and

$$\langle \bar{\xi}, \xi \rangle_* \leq -\|\bar{\xi}\|_*^2 < 0 \quad \text{for all } \xi \in \Xi,$$

or $\bar{\xi} = 0$ and there is no $\tilde{\xi} \in \mathcal{H}$ with $\langle \tilde{\xi}, \xi \rangle_* < 0$ for all $\xi \in \Xi$.

Proof. This theorem is stated in [31, Theorem 3.14]. □

When considering $\Xi = C^\varepsilon(x)$ (which is convex and closed by definition), then this immediately yields the following corollary.

Corollary 3.2.2. *Let $f_i \colon \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous in $x \in \mathcal{H}$ for $i = 1, \ldots, m$ and let $\varepsilon \geq 0$.*

i) If x is locally weakly Pareto optimal, then

$$0 \in C^\varepsilon(x);$$

ii) Let $x \in \mathcal{H}$ and

$$\bar{\xi} \coloneqq \underset{\xi \in -C^\varepsilon(x)}{\arg\min} \|\xi\|_*^2. \tag{3.5}$$

Then either $\bar{\xi} \neq 0$ and

$$\langle \bar{\xi}, \xi \rangle_* \leq -\|\bar{\xi}\|_*^2 < 0 \quad \text{for all} \quad \xi \in C^\varepsilon(x),$$

or $\bar{\xi} = 0$ and there is no $\tilde{\xi} \in \mathcal{H}$ with $\langle \tilde{\xi}, \xi \rangle_* < 0$ for all $\xi \in C^\varepsilon(x)$.

This means that, when working with the ε-subdifferential instead of the Clarke subdifferential, we still have a necessary optimality condition and a way to compute descent directions, although the optimality conditions are weaker and descent can be expected to be weaker than when using the unrelaxed subdifferential.

For the direction from (3.5), we can find a lower bound for a step size up to which we have guaranteed descent in each objective function f_i.

Lemma 3.2.3. *Let $f_i \colon \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous in $x \in \mathcal{H}$ for $i = 1, \ldots, m$. Moreover, we assume that $\varepsilon \geq 0$ holds and we define $\bar{v} \coloneqq R^{-1}(\bar{\xi})$ for the solution $\bar{\xi} \in -C^\varepsilon(x)$ of (3.5). Then*

$$f_i(x + t\bar{v}) \leq f_i(x) - t\,\|\bar{v}\|^2 \quad \text{for all} \quad 0 \leq t \leq \frac{\varepsilon}{\|\bar{v}\|} \quad \text{and} \quad i = 1, \ldots, m.$$

Proof. The proof of [107, Lemma 3.2] can be adapted to the infinite-dimensional case using the fact that the Mean Value Theorem (Theorem 2.1.36) holds for the Clarke subdifferential also in infinite dimensions and because $\|\bar{v}\| = \|\bar{\xi}\|_*$. □

However, solving (3.5) generally requires the knowledge of the entire ε-subdifferential, which is impractical. Instead, we will use Theorem 3.2.1 to compute a finitely generated approximation Ξ of $\operatorname{conv}(\cup_{i=1}^m \partial_C^\varepsilon f_i(x))$, where the resulting direction is guaranteed to have sufficient descent.

3.2.2 Computation of descent directions by adaptive subderivative sampling

In practice, it is generally not possible to compute the entire Clarke subdifferential $\partial_C f_i(x)$, unless additional structure of f_i is known. In this subsection, we describe how the solution of (3.5) can be replaced by a suboptimal one when for every $i \in \{1, \dots, m\}$, only a single subderivative from $\partial_C f_i(x)$ is available at every $x \in \mathcal{H}$. Similar to the gradient sampling approach, the idea behind this method is to use instead of $C^\varepsilon(x)$ in (3.5) the convex hull of a finite number of ε-subderivatives $\xi^0, \dots, \xi^j$ from $C^\varepsilon(x)$ for $j \geq 0$. Since it is impossible to know a priori how many and which ε-subderivatives are required to obtain a good descent direction, we solve (3.5) multiple times in an iterative manner while enriching our approximation until a satisfying direction has been found. To this end, in the following, we will specify how to enrich our current approximation $\operatorname{conv}(\{\xi^0, \dots, \xi^j\})$ and how to characterize an acceptable descent direction.

Suppose that $\Xi = \{\xi^0, \dots, \xi^j\} \subseteq C^\varepsilon(x)$ and define

$$\tilde{\xi} := \operatorname*{arg\,min}_{\xi \in -\operatorname{conv}(\Xi)} \|\xi\|_*^2. \tag{3.6}$$

Let $c \in (0, 1)$. Motivated by Lemma 3.2.3, we regard $\tilde{v} := R^{-1}(\tilde{\xi})$ as an *acceptable* descent direction, if

$$f_i\left(x + \frac{\varepsilon}{\|\tilde{v}\|}\tilde{v}\right) \leq f_i(x) - c\varepsilon\|\tilde{v}\| \quad \text{for all} \quad i = 1, \dots, m. \tag{3.7}$$

If the set $I \subseteq \{1, \dots, m\}$ for which (3.7) is violated is nonempty, then we have to find a new ε-subderivative $\xi' \in C^\varepsilon(x)$ such that $\Xi \cup \{\xi'\}$ yields a better descent direction. Intuitively, (3.7) being violated means that the local behavior of f_i, for $i \in I$, in x in the direction $\tilde{v}$ is not sufficiently captured in Ξ. Thus, for each $i \in I$, we expect that there exists some $t' \in (0, \varepsilon/\|\tilde{v}\|]$ such that $\xi' \in \partial_C f_i(x + t'\tilde{v})$ improves the approximation of $C^\varepsilon(x)$. This is stated in the following lemma. For a proof, we refer to [107, Lemma 3.3].

Lemma 3.2.4. *Let $c \in (0, 1)$, $\Xi = \{\xi^0, \dots, \xi^j\} \subseteq C^\varepsilon(x)$ and $\tilde{v} := R^{-1}(\tilde{\xi})$ for the solution $\tilde{\xi}$ of (3.6) and assume $\tilde{v} \neq 0$. If*

$$f_i\left(x + \frac{\varepsilon}{\|\tilde{v}\|}\tilde{v}\right) > f_i(x) - c\,\varepsilon\|\tilde{v}\| \quad \textit{for some } i \in \{1, \dots, m\},$$

then there exists a $t' \in (0, \varepsilon/\|\tilde{v}\|]$ and $\xi' \in \partial_C f_i(x + t'\tilde{v})$ such that

$$\langle \tilde{\xi}, \xi' \rangle_* > -c\,\|\tilde{\xi}\|_*^2. \tag{3.8}$$

In particular, $\xi' \in C^\varepsilon(x) \setminus \operatorname{conv}(\Xi)$.

Note that Lemma 3.2.4 only shows the existence of t' and ξ' without stating a way how to actually compute them. To solve this problem, let $i \in \{1, \dots, m\}$ be the index of an objective function for which (3.7) is not satisfied, define

$$h_i : \mathbb{R} \to \mathbb{R}, \quad t \mapsto f_i(x + t\tilde{v}) - f_i(x) + ct \|\tilde{v}\|^2,$$

and consider Algorithm 1. If f_i is continuously differentiable around x, then (3.8) is equivalent

---

Algorithm 1 (Computing of a new subderivative)

Require: Current point $x \in \mathcal{H}$, direction $\tilde{v} = R^{-1}(\tilde{\xi}) \in \mathcal{H}$, tolerance $\varepsilon > 0$, Armijo parameter $c \in (0, 1)$.

1: Set $a = 0$, $b = \varepsilon / \|\tilde{v}\|$ and $t = (a + b)/2$.
2: **for** $j = 1, 2, \dots$ **do**
3:    Compute a $\xi' \in \partial_C f_i(x + t\tilde{v})$.
4:    **if** $\langle \tilde{\xi}, \xi' \rangle_* > -c \|\tilde{\xi}\|_*^2$ **then**
5:       stop.
6:    **end if**
7:    **if** $h_i(b) > h_i(t)$ **then**
8:       set $a = t$.
9:    **else**
10:       set $b = t$.
11:    **end if**
12:    Set $t = (a + b)/2$.
13: **end for**
14: **return** Current $\xi' \in \partial_C f_i(x + t\tilde{v})$.

---

to $h_i'(t') > 0$, in other words, h_i being monotonically increasing around t'. Thus, the idea of Algorithm 1 is to find some t such that h_i is monotonically increasing around t, while checking if (3.8) is satisfied for a subderivative $\xi \in \partial_C f_i(x + t\tilde{v})$. For a more thorough discussion of the behavior and termination of Algorithm 1, we refer to [106, 107]. Note that the computation of a subderivative in Step 3 of the algorithm is a problem specific task that may be challenging on its own, see also Section 3.3.3 for details on how this is solved in the numerical example.

We use this method of finding new subderivatives to construct an algorithm that computes descent directions of nonsmooth multiobjective optimization problems, namely Algorithm 2. In Theorem 3.2.8, we will show that Algorithm 2 stops after a finite number of iterations and produces an acceptable descent direction, i.e., a direction that satisfies (3.7). In the infinite-dimensional setting, the proof of [107, Theorem 3.1] cannot be applied directly. The proof uses the fact that the closed ball $\overline{B_\varepsilon(x)}$ is a compact subset of $\mathbb{R}^n$ to conclude that there exists a common Lipschitz constant L on $B_\varepsilon(x)$ for the locally Lipschitz continuous objective functions f_i . This premise does not hold for infinite-dimensional Hilbert spaces. In fact one can construct a function f that is locally Lipschitz continuous on $\mathcal{H}$ but not Lipschitz continuous on $B_2(0)$, as demonstrated in the following example.

Example 3.2.5. *Let $\mathcal{H}$ be a separable Hilbert space with orthonormal basis $\left(e^i\right)_{i \in \mathbb{N}}$. For $i \geq 2$ we define by $\mathcal{B}_i := \overline{B_{1/i}(e_i)}$ a family of closed balls. Obviously, we have $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ for $i \neq j$,*

Algorithm 2 (Computing a descent direction)

Require: Current point $x \in \mathcal{H}$, tolerances $\varepsilon, \delta > 0$, Armijo parameter $c \in (0,1)$.

Compute $\xi_i^0 \in \partial_C^\varepsilon f_i(x)$ for all $i = 1, \dots, m$. Set $\Xi_0 = \{\xi_1^0, \dots, \xi_m^0\}$ and $l = 0$.
for $l = 0, 1, 2, \dots$ **do**
    Compute $\xi^l = \arg\min_{\xi \in -\operatorname{conv}(\Xi_l)} \|\xi\|_*^2$ and set $v^l = R^{-1}\left(\xi^l\right)$.
    if $\left\|\xi^l\right\|_* \leq \delta$ **then**
        return v^l.
    else
        Find all objective functions for which there is insufficient descent:

$$I_l = \left\{ i \in \{1, \dots, m\} \, : \, f_i\left(x + \varepsilon v^l / \|v^l\|\right) > f_i(x) - c\,\varepsilon \,\|v^l\| \right\}.$$

        if $I_l = \emptyset$ **then**
            stop.
        else
            For each $i \in I_l$ compute $t_i \in (0, \varepsilon/\|v^l\|]$ and $\xi_i^l \in \partial_C f_i(x + t_i v^l)$ with $\langle \xi^l, \xi_i^l \rangle_* > -c\,\|\xi^l\|_*^2$ by applying Algorithm 1.
            Set $\Xi_{l+1} = \Xi_l \cup \{\xi_i^l \, : \, i \in I_l\}$.
        end if
    end if
end for

---

since $\|e_i - e_j\| = \sqrt{2} > 1/i + 1/j$. *Using the sets* $\mathcal{B}_i$ *define the function*

$$f : \mathcal{H} \to \mathbb{R}, \quad x \mapsto \begin{cases} i\,\|x - e_i\| & \text{if } x \in \mathcal{B}_i, \\ 1 & \text{otherwise.} \end{cases}$$

The local Lipschitz continuity can be derived from the definition of f. *In fact, the set* $\mathcal{H} \setminus \bigcup_{i \geq 2} \mathcal{B}_i$ *is open and hence for every* $x \in \mathcal{H} \setminus \bigcup_{i \geq 2} \mathcal{B}_i$ *there exists a neighborhood of* x *on which* f *is constant and therefore Lipschitz continuous. If* $x \in \mathcal{B}_i$ *for some* $i \geq 2$ *there exists an open neighborhood* $\mathcal{U}$ *of* x *such that* $\mathcal{U} \cap \mathcal{B}_j = \emptyset$ *for* $j \neq i$. *Then, for all* $y, z \in \mathcal{U}$ *it holds that* $|f(y) - f(z)| \leq i\|y - z\|$, *which can be verified by a simple case separation considering all the case where* y *and* z *belong to* $\mathcal{H} \setminus \bigcup_{i \geq 2} \mathcal{B}_i$ *or* $\mathcal{B}_i$.
If f *would be Lipschitz continuous on* $B_2(0)$ *with some Lipschitz constant* $L > 0$ *we arrive at a contradiction because then* $B_i(0) = \partial_C f(e_i) \subseteq B_L(0)$ *has to hold since* $e_i \in B_2(0)$ *for all* $i \geq 2$.

Nevertheless, we can show that Algorithm 2 still converges for an infinite-dimensional Hilbert space. We can recover the main argument of the proof of [107, Theorem 3.1] but need some preparatory results to bypass the fact that we cannot use a common Lipschitz constant for the functions f_i on $B_\varepsilon(x)$. To this end, we introduce the two following lemmas.

Lemma 3.2.6. *Let* $C_1 \subseteq C_2 \subseteq \mathcal{H}^*$ *be two convex and closed subsets. Define*

$$\xi^1 := \arg\min_{\xi \in C_1} \|\xi\|_*^2 \quad \text{and} \quad \xi^2 := \arg\min_{\xi \in C_2} \|\xi\|_*^2.$$

Note that ξ^1 and ξ^2 are well-defined and unique. Let $\delta \geq 0$ such that $\|\xi^2\|_* \geq \delta$. Then

$$\|\xi^1 - \xi^2\|_*^2 \leq \|\xi^1\|_*^2 - \delta^2.$$

Proof. Simply rewriting the squared norm yields

$$\|\xi_1 - \xi_2\|_*^2 = \|\xi_1\|_*^2 - \|\xi_2\|_*^2 + 2\,\langle \xi_2, \xi_2 - \xi_1 \rangle_*.$$

From $\xi_1 \in C_2$ we infer the projection property $\langle \xi_2, \xi_2 - \xi_1 \rangle_* \leq 0$. In addition with the relation $-\|\xi_2\|_*^2 \leq -\delta^2$ we get the desired result. □

In the proof of the following lemma we directly incorporate Lemma 3.2.6.

Lemma 3.2.7. *Let $(\xi^l)_{l\geq 0}$ be an arbitrary sequence in $\mathcal{H}^*$. Define $\Xi_l := \{\xi_0, \ldots, \xi_l\}$ for $l \geq 0$. Let the sequence $(\psi^l)_{l\geq 0} \subset \mathcal{H}^*$ be given by*

$$\psi^l = \underset{\psi \in -\operatorname{conv}(\Xi_l)}{\arg\min} \ \|\psi\|_*^2.$$

Then ψ^l converges strongly in $\mathcal{H}^*$.

Proof. From the definition of the elements ψ^l we obtain that $(\|\psi^l\|_*)_{l\geq 0}$ is monotonically decreasing. Hence we can conclude that there exists a $\delta \geq 0$ such that

$$\lim_{l\to+\infty} \|\psi^l\|_*^2 =: \delta^2 \geq 0.$$

Using the limit $\delta^2 \geq 0$ and Lemma 3.2.6, we will show that $(\psi^l)_{l\geq 0}$ is a Cauchy sequence in $\mathcal{H}^*$. Let $l, m \geq 1$ and consider $\|\psi^l - \psi^{l+m}\|_*$. Choosing $C_1 = -\operatorname{conv}(\Xi_l)$, $C_2 = -\operatorname{conv}(\Xi_{l+m})$, $\xi^1 = \psi^l$ and $\xi^2 = \psi^{l+m}$ with $\|\psi^{l+m}\|_* \geq \delta$ we infer from Lemma 3.2.6 that

$$\|\psi^l - \psi^{l+m}\|_*^2 \leq \|\psi^l\|_*^2 - \delta^2.$$

Since $\lim_{l\to+\infty}\|\psi^l\|_*^2 = \delta^2$ it follows that $(\psi^l)_{l\geq 0}$ is a Cauchy sequence in $\mathcal{H}^*$. Consequently, ψ^l converges strongly in $\mathcal{H}^*$. □

Using Lemmas 3.2.6 and 3.2.7, we can adapt the proof of [107, Theorem 3.1] to show that Algorithm 2 terminates in the Hilbert space setting.

Theorem 3.2.8. *Let $f_i\colon \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous for $i = 1, \ldots, m$. Then, Algorithm 2 terminates and hence the sequence $(v^l)_{l\geq 0}$ is finite. If $\tilde{v}$ is the last element of $(v_l)_{l\geq 0}$ and $\tilde{\xi} = R(\tilde{v})$, then either $\|\tilde{\xi}\|_* \leq \delta$ or $\tilde{v}$ is an acceptable descent direction, that is,*

$$f_i\left(x + \frac{\varepsilon}{\|\tilde{v}\|}\tilde{v}\right) \leq f_i(x) - c\varepsilon\|\tilde{v}\| \quad \textit{for all} \quad i = 1, \ldots, m.$$

Proof. Assume Algorithm 2 does not terminate, that is, the sequences $(\xi^l)_{l\geq 0}$ and $(v^l)_{l\geq 0} = \left(R^{-1}(\xi^l)\right)_{l\geq 0}$ are infinite sequences. Independently from Steps 7 and 11, Lemma 3.2.7 guarantees that ξ^l converges to an element $\tilde{\xi} \in \mathcal{H}^*$ as $l \to +\infty$, and, accordingly, v^l converges to $\tilde{v} = R^{-1}(\tilde{\xi})$ as $l \to +\infty$. Hence, the scalars $t_i^l \in (0, \varepsilon/\|v^l\|]$ chosen in Step 11 are bounded for all

$l \geq 0$ and $i \in I_l$. Using this, we choose a subsequence $(l_k)_{k\geq 0}$ such that $I_{l_k} = \tilde{I}$ remains constant and $t_i^{l_k} \to \tilde{t}_i \in [0, \varepsilon/\|\tilde{v}\|]$ for $k \to +\infty$ for all $i \in \tilde{I}$. Accordingly, $x + t_i^{l_k} v_{l_k}$ converges to $x + \tilde{t}_i \tilde{v}$ as $k \to +\infty$.

Since the functions f_i are locally Lipschitz continuous, there exists a common local Lipschitz constant $L \geq 0$ such that all objective functions f_i are L-Lipschitz continuous in a neighborhood of $x + \tilde{t}_i \tilde{v}$, respectively. Due to the convergence of the sequences, we can find an index $K \geq 0$ and $\kappa \geq 0$ such that

$$\left\|\xi_i^{l_k}\right\|_* \leq L + \kappa \quad \text{for all} \quad k \geq K \quad \text{and} \quad i \in \tilde{I}. \tag{3.9}$$

On the other hand, we can bound $\|\xi^l\|_* \leq \|\xi^0\|_* \leq \max\left(\|\xi^0\|_*, L + \kappa\right)$ for all $l \geq 0$. For convenience, we update $L \to \max\left(\|\xi^0\|_*, L + \kappa\right)$ for the remainder of the proof to get a uniform bound for $\|\xi_i^{l_k}\|_*$ and $\|\xi^l\|_*$ for all $k \geq K$, $i \in \tilde{I}$ and $l \geq 0$.
Now, let $k \geq K$ and $i \in \tilde{I}$. Since $\xi_i^{l_{k-1}} \in \Xi_{l_k}$ and $-\xi^{l_{k-1}} \in \operatorname{conv}\left(\Xi_{l_{k-1}}\right) \subseteq \operatorname{conv}\left(\Xi_{l_k}\right)$, we have $(1-s)(-\xi^{l_{k-1}}) + s\xi_i^{l_{k-1}} \in \operatorname{conv}\left(\Xi_{l_k}\right)$ for all $s \in [0,1]$. Therefore, the minimization property of ξ^{l_k} yields that

$$\begin{aligned}
\left\|\xi^{l_k}\right\|_*^2 &\leq \left\| -\xi^{l_{k-1}} + s(\xi_i^{l_{k-1}} + \xi^{l_{k-1}})\right\|_*^2 \\
&= \left\|\xi^{l_{k-1}}\right\|_*^2 - 2s\left\langle \xi^{l_{k-1}}, \xi_i^{l_{k-1}} + \xi^{l_{k-1}}\right\rangle_* + s^2\left\|\xi_i^{l_{k-1}} + \xi^{l_{k-1}}\right\|_*^2 \\
&= \left\|\xi^{l_{k-1}}\right\|_*^2 - 2s\left\langle \xi^{l_{k-1}}, \xi_i^{l_{k-1}}\right\rangle_* - 2s\left\|\xi^{l_{k-1}}\right\|_*^2 + s^2\left\|\xi_i^{l_{k-1}} + \xi^{l_{k-1}}\right\|_*^2,
\end{aligned} \tag{3.10}$$

for all $s \in [0,1]$. Since $i \in \tilde{I}$ we must have

$$\left\langle \xi^{l_{k-1}}, \xi_i^{l_{k-1}}\right\rangle_* > -c\left\|\xi^{l_{k-1}}\right\|_*^2, \tag{3.11}$$

by Step 11. From inequality (3.9) and the choice of the Lipschitz constant L, we can conclude that

$$\left\|\xi_i^{l_{k-1}} + \xi^{l_{k-1}}\right\|_* \leq 2L. \tag{3.12}$$

Combining (3.10) with (3.11) and (3.12) yields

$$\begin{aligned}
\left\|\xi^{l_k}\right\|_*^2 &< \left\|\xi^{l_{k-1}}\right\|_*^2 + 2sc\left\|\xi^{l_{k-1}}\right\|_*^2 - 2s\left\|\xi^{l_{k-1}}\right\|_*^2 + 4s^2L^2 \\
&= \left\|\xi^{l_{k-1}}\right\|_*^2 - 2s(1-c)\left\|\xi^{l_{k-1}}\right\|_*^2 + 4s^2L^2.
\end{aligned}$$

Now, we choose $s := (1-c)\|\xi^{l_{k-1}}\|_*^2/(4L^2)$. Since $1 - c \in (0,1)$ and $\|v^{l_{k-1}}\|_* \leq L$ we have $s \in (0,1)$. Thus, we obtain

$$\begin{aligned}
\left\|\xi^{l_k}\right\|_*^2 &< \left\|\xi^{l_{k-1}}\right\|_*^2 - \frac{2(1-c)^2}{4L^2}\left\|\xi^{l_{k-1}}\right\|_*^4 + \frac{(1-c)^2}{4L^2}\left\|\xi^{l_{k-1}}\right\|_*^4 \\
&= \left(1 - \frac{(1-c)^2}{4L^2}\left\|\xi^{l_{k-1}}\right\|_*^2\right)\left\|\xi^{l_{k-1}}\right\|_*^2.
\end{aligned}$$

We have assumed that Algorithm 2 does not terminate. Therefore, we must have $\|\xi^{l_{k-1}}\|_* > \delta$, which implies

$$\left\|\xi^{l_k}\right\|_*^2 < r\left\|\xi^{l_{k-1}}\right\|_*^2,$$

with $r := \left(1 - \left(\frac{1-c}{2L}\,\delta\right)^2\right)$. Note that we have $\delta < \|\xi^{l_k}\|_* \leq L$ for all $k \geq 0$, so $r \in (0,1)$. Additionally, r does not depend on l_k, so we have

$$\left\|\xi^{l_k}\right\|_*^2 < r\left\|\xi^{l_{k-1}}\right\|_*^2 < r^2\left\|\xi^{l_{k-2}}\right\|_*^2 < \ldots < r^k\left\|\xi^{l_0}\right\|_*^2 \leq r^k L^2.$$

In particular, there exists some k such that $\|\xi^{l_k}\|_* \leq \delta$, which is a contradiction. □

Remark 3.2.9. *The proof of Theorem 3.2.8 shows that for convergence of Algorithm 2, it would be sufficient to consider only a single $i \in I_l$ in Step 11. Similarly, for the initial approximation Ξ_0, a single element of $\partial^\varepsilon_C f_i(x)$ for any $i \in \{1,\ldots,m\}$ would be enough. A modification of either step can potentially reduce the number of executions of Step 11 (i.e., Algorithm 1) in Algorithm 2 in case the ε-subdifferentials of multiple objective functions are similar. However, we will forgo these modifications and leave Algorithm 2 as it is, since both modifications also introduce a bias towards certain objective functions, which we want to avoid.*

3.2.3 The final descent method

Building on Algorithm 2, it is now straightforward to construct the descent method for locally Lipschitz continuous multiobjective optimization problems given in Algorithm 3.

Algorithm 3 (Nonsmooth descent method)

Require: Initial point $x^0 \in \mathcal{H}$, parameters for stopping criterion $\overline{\delta}, \overline{\varepsilon} \geq 0$, tolerance sequences $(\delta_k)_{k\geq 0}, (\varepsilon_k)_{k\geq 0} \subseteq \mathbb{R}_{>0}$, Armijo parameters $c \in (0,1)$, $t_0 > 0$.
1: **for** $k = 0,1,2,\ldots$ **do**
2: Compute a descent direction v^k via Algorithm 2 with inputs $(x^k, \varepsilon_k, \delta_k, c)$.
3: Use backtracking line search to determine

$$\bar{s} = \inf\left\{s \in \mathbb{N} \cup \{0\} \,:\, f_i(x^k + 2^{-s}t_0 v^k) \leq f_i(x^k) - 2^{-s}ct_0\left\|v^k\right\|^2 \text{ for all } i \in \{1,\ldots,m\}\right\}$$

and set $\bar{t} = \max\left(2^{-\bar{s}}t_0, \varepsilon_k/\|v^k\|\right)$.
4: **if** $\|v^k\| \leq \overline{\delta}$ and $\varepsilon_k \leq \overline{\varepsilon}$ **then**
5: **return** $(\overline{\varepsilon}, \overline{\delta})$-critical point x^k
6: **else**
7: Set $x^{k+1} = x^k + \bar{t}v^k$.
8: **end if**
9: **end for**

In Step 3, the classical Armijo backtracking line search is used (see [103]) for the sake of simplicity. Note that it is well-defined due to Step 7 in Algorithm 2.

Clearly, the stopping condition matches the Definition 3.1.7 of the current iterate being $(\overline{\varepsilon}, \overline{\delta})$-critical exactly. Thus, when Algorithm 3 terminates, it will in fact return an $(\overline{\varepsilon}, \overline{\delta})$-critical point. We state a convergence as well as a termination result for Algorithm 3. First off, in Theorem 3.2.10, we address the case, where the tolerances $\overline{\varepsilon}$ and $\overline{\delta}$ are both set to 0. The theorem states that Algorithm 3 converges (in the sense of subsequences) to Pareto critical points in the limit. Then, in Theorem 3.2.11 we show that the algorithm is capable of finding $(\overline{\varepsilon}, \overline{\delta})$-critical points, for generalized parameter settings.

Theorem 3.2.10. *For $i = 1, \dots, m$ let $f_i \colon \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous. Let x^0 be an element in $\mathcal{H}$ and $(\delta_k)_{k\geq 0}, (\varepsilon_k)_{k\geq 0} \subseteq \mathbb{R}_{>0}$ be two sequences with*

$$\delta_k \to 0, \quad \varepsilon_k \to 0 \quad \text{as} \quad k \to +\infty \quad \text{and} \quad \sum_{k=1}^{\infty} \varepsilon_k \delta_k = \infty.$$

Let further $\overline{\varepsilon} = \overline{\delta} = 0$, $c \in (0,1)$ and $t_0 > 0$. Assume Algorithm 3 does not converge after finitely many steps. Let $(x^k)_{k\geq 0}$ be the sequence generated by Algorithm 3 with inputs $(x^0, \overline{\delta}, \overline{\varepsilon}, (\delta_k)_{k\geq 0}, (\varepsilon_k)_{k\geq 0}, c, t_0)$. Then, the following statements hold:

i) Every accumulation point of $(x^k)_{k\geq 0}$ is Pareto critical;

ii) If one f_i is bounded from below then $(x^k)_{k\geq 0}$ possesses a subsequence $(x^{k_l})_{l\geq 0}$ such that $\|v^{k_l}\| \to 0$ as $l \to \infty$.

Proof.

i) In the following proof we choose appropriate subsequences of $(x_k)_{k\geq 0}$ multiple times. We do this without relabeling the sequence and only comment when doing so. Let x^* be an accumulation point of $(x^k)_{k\geq 0}$. Then, there exists a subsequence (no relabeling) with $x^k \to x^*$ as $k \to \infty$.

First we show that $\|v^k\| \leq \delta_k$ is true for infinitely many $k \geq 0$. In each iteration of Algorithm 3, we use Algorithm 2. Therefore at least one of the stopping criteria of Algorithm 2 is met infinitely many times. Assume the stopping criteria $\|v^k\| < \delta$ in Step 4 of Algorithm 2 (where $\|v^k\| = \|\xi^k\|_*$) is only met finitely many times. Then, there exists $J \geq 0$ such that for all $k \geq J$ it holds that

$$f_i(x^{k+1}) \leq f_i(x^k) - c\varepsilon_k \|v^k\| \quad \text{for all} \quad i = 1, \dots, m \quad \text{and} \quad \|v^k\| > \delta_k. \tag{3.13}$$

The first inequality follows from the active stopping criterion in Step 7 of Algorithm 2 and the way the backtracking rule in Step 3 of Algorithm 3 is defined. We show that these inequalities lead to a contradiction. Let $i \in \{1, \dots, m\}$ and $K \geq J$. Then, we have

$$\begin{aligned} f_i(x^{K+1}) - f_i(x^J) &= \sum_{k=J}^{K} f_i(x^{k+1}) - f_i(x^k) \leq \sum_{k=J}^{K} -c\varepsilon_k \|v^k\| \\ &< -c \sum_{k=J}^{K} \varepsilon_k \delta_k. \end{aligned} \tag{3.14}$$

We know by the assumptions on $(\delta_k)_{k\geq 0}$ and $(\varepsilon_k)_{k\geq 0}$ that the last series diverges. Accordingly, the sequential continuity of f_i yields that

$$f_i(x^*) - f_i(x^J) = -c \lim_{K\to+\infty} \sum_{k=J}^{K-1} \varepsilon_k \delta_k = -\infty,$$

which is a contradiction as the difference on the left hand side is finite.

Therefore, $\|v^k\| \leq \delta_k$ holds for infinitely many $k \geq 1$. This means, we can choose an appropriate subsequence of $(x^k)_{k\geq 0}$ (no relabeling) such that

$$x^k \to x^* \quad \text{as} \quad k \to +\infty \quad \text{and} \quad \|v^k\| < \delta_k \quad \text{for all} \quad k \geq K.$$

By Theorem 3.1.10 it follows that $0 \in \operatorname{conv}(\cup_{i=1}^m \partial_C f_i(x^*))$. Hence x^* is Pareto critical.

ii) The proof follows from inequalities (3.13) and (3.14) and the fact that $\|v^k\| \leq \delta_k$ has to hold for infinitely many $k \geq 0$ if f_i is bounded from below.

□

In practice, we will rely on Algorithm 3 terminating after a finite number of iterations due to the stopping criterion for tolerances $\overline{\varepsilon}, \overline{\delta} > 0$ instead of generating infinite sequences of iterates. The following theorem states that the algorithm will in fact terminate after a finite number of iterations, for example, if the sequences $(\varepsilon_k)_{k\geq 0}$ and $(\delta_k)_{k\geq 0}$ are chosen as certain constants.

Theorem 3.2.11. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be locally Lipschitz continuous for $i = 1, \ldots, m$. We suppose that x^0 is an element in $\mathcal{H}$ and set $\overline{\varepsilon}, \overline{\delta} > 0$. Let $(\delta_k)_{k\geq 0}, (\varepsilon_k)_{k\geq 0} \subseteq \mathbb{R}_{>0}$ be constant sequences with $\delta_k = \overline{\delta}$, $\varepsilon_k = \overline{\varepsilon}$ for all $k \geq 0$, $c \in (0, 1)$ and $t_0 > 0$. Let $(x^k)_{k\geq 0}$ be the sequence generated by Algorithm 3 with inputs $(x^0, \overline{\delta}, \overline{\varepsilon}, (\delta_k)_{k\geq 0}, (\varepsilon_k)_{k\geq 0}, c, t_0)$ and assume that one objective function f_i is bounded from below. Then, Algorithm 3 returns an $(\overline{\varepsilon}, \overline{\delta})$-critical point after finitely many iterations.*

Proof. Assume Algorithm 3 does not terminate after finitely many steps and produces an infinite sequence $(x^k)_{k\geq 0}$. Since the condition $\varepsilon_k \leq \overline{\varepsilon}$ is fulfilled in every iteration of Algorithm 3, we show that $\|v^k\| \leq \overline{\delta}$ has to hold for one $k \geq 1$. Then, Algorithm 3 stops since the condition $\varepsilon_k \leq \overline{\varepsilon}$ is fulfilled in every step. Again one of the stopping criteria of Algorithm 2 has to be fulfilled infinitely many times. If $\|v^k\| \leq \overline{\delta}$ in Step 4 of Algorithm 2 is fulfilled then also Algorithm 3 stops. If this is not the case then Algorithm 2 only stops due to the stopping condition in Step 8 and we conclude that for all $k \geq 0$ it holds that

$$f_i(x^{k+1}) \leq f_i(x^k) - c\,\overline{\varepsilon}\|v^k\| \quad \text{for all} \quad i = 1, \ldots, m \quad \text{and} \quad \|v^k\| > \overline{\delta}.$$

Combining these inequalities, we have for all $k \geq 0$

$$f_i(x^{k+1}) \leq f_i(x^k) - c\,\overline{\varepsilon}\overline{\delta} \quad \text{for all} \quad i = 1, \ldots, m. \tag{3.15}$$

This leads to a contradiction. Fix $i \in \{1, \ldots, m\}$ such that f_i is bounded from below. Then, by (3.15), we have for all $K \geq 1$

$$f_i(x^K) - f_i(x^0) = \sum_{k=0}^{K-1} f_i(x^{k+1}) - f_i(x^k) \leq \sum_{k=0}^{K-1} -c\,\overline{\varepsilon}\overline{\delta} = -Kc\,\overline{\varepsilon}\overline{\delta}. \tag{3.16}$$

Since the right-hand side of (3.16) diverges to $-\infty$ for $K \to \infty$, we arrive at a contradiction given that f_i is bounded from below. □

Remark 3.2.12. *The choice of the tolerance sequences $(\delta_k)_{k\geq 0}$ and $(\varepsilon_k)_{k\geq 0}$ in Theorem 3.2.11 can be further relaxed. We are not forced to use constant sequences $\delta_k = \overline{\delta}$ and $\varepsilon_k = \overline{\varepsilon}$. Instead, we could choose arbitrary sequences with $\delta_k \in (0, \overline{\delta}]$ and $\varepsilon_k \in (0, \overline{\varepsilon}]$ that satisfy the condition $\sum_{k=0}^{\infty} \delta_k \varepsilon_k = \infty$ similar to the requirements of Theorem 3.2.10. This could be further relaxed to arbitrary positive sequences $\delta_k > 0$ and $\varepsilon_k > 0$ provided that they remain bounded by $\overline{\delta}$ and $\overline{\varepsilon}$ for almost all iterations and that they also satisfy the summability property $\sum_{k=0}^{\infty} \delta_k \varepsilon_k = \infty$. The proof in these settings follows analogously to the proof of Theorem 3.2.11.*

3.3 Application in bicriterial optimal control of an obstacle problem

In this section, we examine the behavior of Algorithm 3 applied to a classic, nonsmooth obstacle-constrained optimal control problem – see, for example, [136, Section 6] – on the two-dimensional domain $\Omega := (-1, 1)^2$ for two objective functions.

The forward problem, that is, the constraint in the optimal control problem, can be interpreted as the problem of finding a displacement $y : \Omega \to \mathbb{R}$ of a clamped membrane under external, distributed vertical forces $u : \Omega \to \mathbb{R}$ (assuming small displacements with linear response) with a rigid obstacle, described by $\psi : \Omega \to \mathbb{R}$, limiting the vertical displacement to $y \leq \psi$.

This constrained problem can be equivalently formulated as a convex energy minimization problem or via the corresponding partial differential variational inequality, and it is well understood. Most importantly, the control-to-state operator is known to be well-defined, Lipschitz continuous and Hadamard- but generally not Fréchet-differentiable everywhere [122, 151, 171]. There is also extensive literature on computational aspects for obstacle constrained dynamics, including efficient solvers [116, 117, 167, 232].

Various aspects of optimal control problems with the obstacle constraint have previously been considered in a broad range of publications (e.g., [72, 171, 231]), but, to the best of our knowledge, obstacle-constrained optimization problems have not been considered in the context of infinite-dimensional multiobjective optimization (though their discretizations have been dealt with in finite-dimensional, nonsmooth multiobjective optimization [168]). Due to the nonlinearity of the control-to-state operator, these problems are generally nonconvex and nonsmooth. However, (varying notions of) subdifferentials of the control-to-state operator have been characterized in [201], and [200, Theorem 5.7] shows how to compute an element of the Clarke subdifferential of control reduced optimal control of the obstacle problem – which is what we require in order to employ our common descent method. Note that this exact technique for computing subderivatives was applied in scalar optimal control of obstacle-constrained problems using an inexact bundle method in function space [128, 129].

3.3.1 Problem description

The domain we consider is the two dimensional square $\Omega = (-1,1)^2 \subseteq \mathbb{R}^2$ with an obstacle described by a function $\psi \in H^1(\Omega)$ (to be specified later), yielding the set of admissible displacements

$$K \coloneqq \left\{ y \in V \coloneqq H_0^1(\Omega) : y \leq \psi \text{ a.e. on } \Omega \right\},$$

which is guaranteed to be nonempty by choosing ψ appropriately. The variational inequality formulation of the constraining obstacle problem for a fixed, distributed external load $f \in V^* \coloneqq H^{-1}(\Omega)$ amounts to finding $y \in K$, such that

$$\langle Ay - f, v - y \rangle_{V^*,V} \geq 0 \text{ for all } v \in K. \tag{3.17}$$

Here, $A\colon V \to V^*$ is a linear, continuous and coercive partial differential operator (we will be using the weak form of $A = -\Delta$ in the following), and $\langle \cdot , \cdot \rangle_{V^*,V}$ denotes the dual pairing. In the optimal control problem, we consider the control space $U \coloneqq L^2(\Omega)$ with the standard $U \hookrightarrow V^*$ Gelfand-type embedding to let $u \in U$ assume the role of f in (3.17).

Given a desired state and reference control $y_d \in H \coloneqq L^2(\Omega)$, $u_d \in U$, we then fix the two cost functionals to obtain the optimal control problem

$$\begin{aligned} \min_{(y,u) \in K \times U} \ & \frac{1}{2} \begin{pmatrix} \|y - y_d\|_H^2 \\ C \|u - u_d\|_U^2 \end{pmatrix}, \\ \text{s.t. } & \langle Ay - u, v - y \rangle_{V^*,V} \geq 0 \text{ for all } v \in K, \end{aligned} \tag{3.18}$$

with a hyperparameter $C > 0$. Note that C is essentially introduced in order to scale the axes in the plots of the Pareto fronts, so that they are easier to interpret. Introducing and tuning the parameter C can be interpreted as preconditioning of the problem.

Problem (3.18) is an optimal control problem and clearly a constrained problem. To make it fit into the realm of unconstrained optimization, which we have formulated the algorithms in this chapter for, we simply make use of the existence of the Hadamard-differentiable solution operator of the obstacle problem $S\colon U \to K \subseteq V$ mapping a control u to the solution $y = S(u)$ of the constraining variational inequality of (3.18) to obtain the equivalent control-reduced multiobjective optimization problem

$$\min_{u \in U} \frac{1}{2} \begin{pmatrix} \|S(u) - y_d\|_H^2 \\ C \|u - u_d\|_U^2 \end{pmatrix}. \tag{3.19}$$

Using the direct method of variational calculus, one can easily show, that the weighted-sum-scalarized problems corresponding to this problem possess solutions, and hence the Pareto set and the Pareto front of this problem are nonempty. What remains to be fixed in the remainder is the choice of the algorithmic parameters, the desired states and controls y_d, u_d and the specific obstacle ψ.

We describe the choice of the free parameters in the following subsection. In all cases, we ensure that our problem configuration in fact captures the nonsmooth behavior of the problem. As mentioned above, the nonsmoothness of the problem is introduced by the solution operator. More specifically, the points of non-Fréchet-differentiability are precisely those of so called *weak contact*, that is, where the control corresponds to a state that is in contact with the obstacle, but where there are no normal forces actively preventing penetration on a sufficiently large area (in the sense of Sobolev capacities). Such configurations of "coincidental" contact are exactly those, where the problem transfers from a free Poisson problem to a full constrained problem.

3.3.2 Computational procedure and joint parameters

The goal of our numerical procedure is to find an approximate representation of the Pareto front and Pareto set of the obstacle-constrained optimal control problem (3.19). To this end, we apply Algorithm 3 starting from a number of varying initial values. As shown in Theorem 3.2.11, for each initial value, Algorithm 3 terminates at an $(\overline{\varepsilon}, \overline{\delta})$-critical point after finitely many steps. As the terminal iterate of the algorithm typically varies with varying initial guesses, we obtain a representation of the Pareto front and the Pareto set of (3.19) by $(\overline{\varepsilon}, \overline{\delta})$-critical points. We chose the different initial controls $u_0 \in U$ constant on the entire domain. Specifically, we apply the algorithm for constant initial controls for all values $u_0 \equiv \hat{u} \in \{1, 2, \ldots, 8\}$ and for all mesh discretizations $h_{\max} \in \{0.2, 0.1, 0.05, 0.02\}$.

For all experiments, we fix the scaling parameter $C = 1.5\text{e}{-2}$ and the hyperparameters $\overline{\varepsilon} = \overline{\delta} = 1\text{e}{-4}$, $c = 1\text{e}{-1}$ and the constant sequences $(\varepsilon_k)_{k\geq 0} \equiv \overline{\varepsilon}, (\delta_k)_{k\geq 0} \equiv \overline{\delta}$. Further, we set $y_d \equiv 2$ and $u_d \equiv 0$. This choice yields a setting where the first cost functional improves when the state is pushed upwards towards the desired state, while the second objective is optimal for vanishing controls, leading to a setting where optimal compromises can be expected to achieve some upwards deformation of the state using controls "efficiently". This suggests that contact should be established in optimal compromises, but no additional forces are to be applied, leading to a nonsmooth weak-contact situation in the optimal compromise. We therefore expect the algorithm to have to deal with increasing nonsmoothness over the course of the run.

Note that at this point, only the obstacle remains to be fixed in each of the examples. We will specify the obstacles we use in the experiment runs in Subsection 3.3.4.

3.3.3 Implementation details

We discretize the optimal control problem using Lagrangian $P1$ finite elements on a triangulation of Ω supplied by MATLAB's PDE-toolbox with a predetermined target maximum element edge length $h_{\max}$ (which is typically only violated by fractions of a percent) and nodally interpolate the obstacle ψ to essentially enforce the nonpenetration constraint nodally. The discretizations of Ω we use are those corresponding to $h_{\max} \in \{0.2, 0.1, 0.05, 0.02\}$. Additionally, we compute a reference solution u^*_{ref} for $h_{\max} = 0.01$ to emulate the exact solution in order to investigate convergence of the solutions for finer meshes. The number of finite elements corresponding to each mesh discretization can be seen in Table 3.1, ranging from 135 to 45 857 elements.

$h_{\max}$	0.2	0.1	0.05	0.02	0.01
# FEM	135	494	1 909	11 682	45 857

Table 3.1: Number of finite elements for different maximum edge lengths $h_{\max}$.

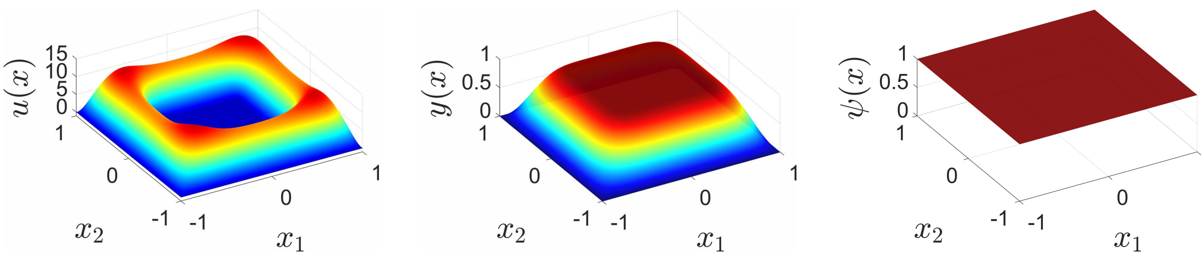
The control-to-state operator is implemented using an active-set strategy applied to the equivalent energy minimization formulation of the obstacle problem and the subderivatives are obtained based on the discretized analogue of the adjoint-based computations in [200, Theorem 5.7], where the discrete approximation to the adjoint state is computed using MATLAB's *mldivide* routine to solve the corresponding linear system. Our implementations of Algorithms 1-3 is also in MATLAB. The preconditioner that maps generalized subderivatives to primal objects, in other words, Riesz's operator (in, e.g., Lemma 3.2.3), is chosen as the canonical $L^2(\Omega)$-Riesz operator.

3.3.4 Numerical results

In this subsection, we present the numerical results obtained by Algorithm 3 for the optimal control problem described in Subsection 3.3.1. The settings of the parameters for Algorithm 3 are specified in Subsection 3.3.2, while the implementation details to handle the PDE-constraints are described in Subsection 3.3.3. To conduct the experiments, we only have to choose the shape of the obstacle ψ, which we do in two example instances below. We consider a constant obstacle and a more involved example. Further, we analyze the size of the approximated Goldstein ε-subdifferential, which is computed in every iteration of Algorithm 3 using Algorithm 2, in order to investigate the behavior of our algorithm.

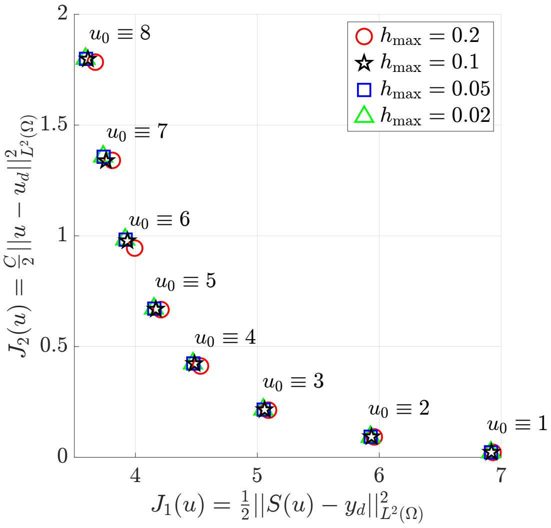
Configuration 1: Constant obstacle

For the first example configuration, we set $\psi \equiv 1$. Since the desired state is $y_d \equiv 2$, the minimization of $J_1(u) = 1/2\, \|S(u) - y_d\|_H^2$ is expected to lead to configurations with contact $y(x) = \psi(x)$ for some points $x \in \Omega$. On the other hand, the second objective function $J_2(u) = C/2\, \|u - u_d\|_U^2$, with $u_d \equiv 0$, penalizes the control cost. We end up in a scenario with conflicting objective functions, with solutions drawn to the obstacle by one objective. An (approximate) optimal

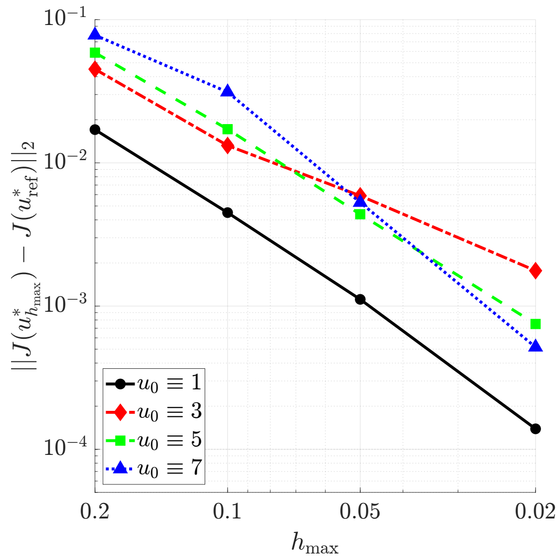


(a) Control (b) State (c) Obstacle

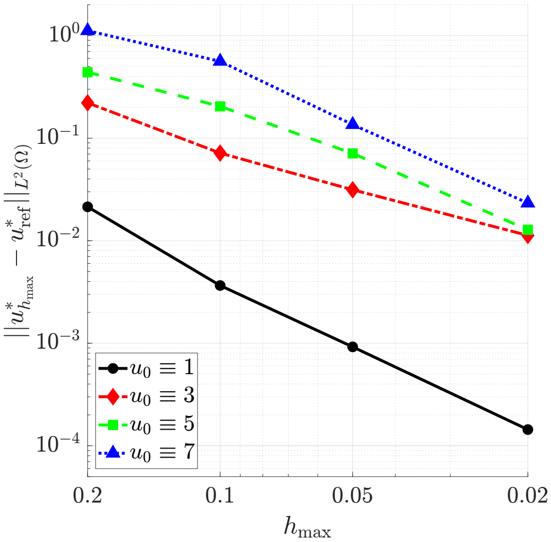
Figure 3.1: A Pareto optimal control computed with Algorithm 3 for mesh size $h_{\max} = 0.02$, initial control $u_0 \equiv 8$ and the constant obstacle $\psi \equiv 1$.



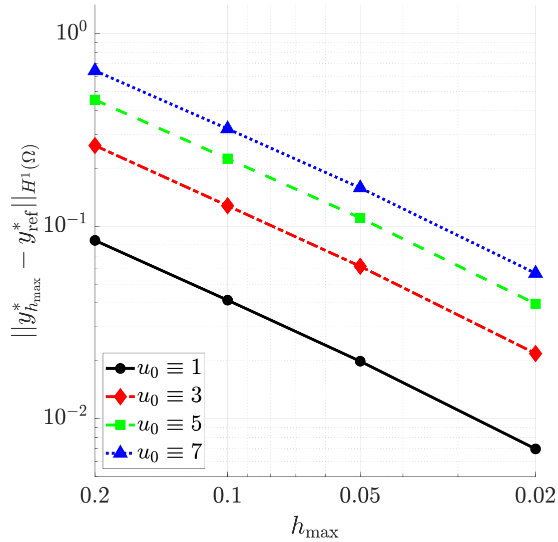
(a) Pareto fronts for different mesh discretizations.



(b) Euclidean distance between optimal values in the image space.



(c) L^2-distance between optimal control and reference control.



(d) H^1-distance between optimal state and reference state.

Figure 3.2: Qualitative analysis of the solutions derived by Algorithm 3 for different discretizations for the constant obstacle. Subfigures (b) - (d) use the reference solution u^*_{ref} corresponding to mesh size $h_{\max} = 0.01$.

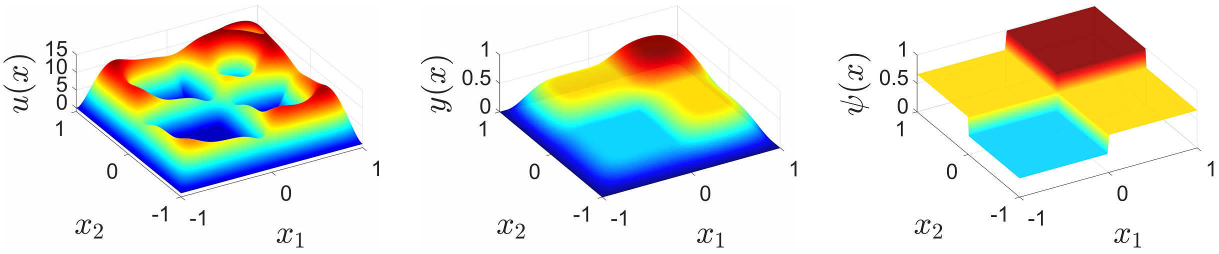
		$h_{\max}$				
		0.2	0.1	0.05	0.02	0.01
u_0	1	4	4	3	3	3
	2	4	5	4	4	4
	3	21	22	11	37	32
	4	106	160	236	264	91
	5	426	2 281	1 116	542	210
	6	3 885	3 894	1 619	1 025	323
	7	4 756	6 190	2 918	1 370	657
	8	2 491	3 576	3 194	2 697	822

Table 3.2: Configuration 1: Number of iterations of Algorithm 3 for different initial values u_0 and mesh sizes $h_{\max}$.

compromise in this conflicting setting can be seen in Figure 3.1.

Subfigure 3.1a shows the optimal control u computed over 2 697 iterations. The corresponding state y is shown in Subfigure 3.1b with the obstacle ψ in Subfigure 3.1c. All solutions obtained by Algorithm 3 for the different meshes and initial states share similar features. In the middle of the domain, there is an area of contact, that is, a region with $y(x) = \psi(x)$. In this area the control $u(x)$ vanishes. This is intuitive, since increasing the control at a point with contact only increases the objective function value of $C/2\,\|u - u_d\|_U^2$ without decreasing the objective function value of $1/2\,\|S(u) - y_d\|_H^2$. The size of the area of contact is influenced by the magnitude of the initial control u_0. For larger control values, we observe a larger area of contact in the solution, while for smaller values, the size of the area of contact is smaller. If we start with small initial control (e.g., $u_0 \equiv 1$), we get solutions with no contact at all, in other words, solutions where the obstacle problem reduces to Poisson's equation and the obstacle ψ can be ignored.

A complete picture of the solutions obtained by Algorithm 3 and the convergence behavior is depicted in Figure 3.2 and Table 3.2. A qualitative analysis of the solutions is included in Figure 3.2. The iteration numbers required for each run are summarized in Table 3.2. For all initial values and mesh sizes the algorithm successfully terminates before reaching the maximum number of 10 000 iterations and computes an $(\bar{\varepsilon}, \bar{\delta})$-critical point. Subfigure 3.2a shows the obtained solutions in the objective space for all initial values ranging from $u_0 \equiv 0$ to $u_0 \equiv 8$ and for all mesh sizes $h_{\max} \in \{0.2, 0.1, 0.05, 0.02\}$ marked with different symbols and colors, respectively. The solutions with the same initial value (but for different mesh discretizations) cluster, while solutions for different initial values are evenly distributed and form a curved front. The clustering behaviour in the objective space will be examined further in Subfigure 3.2b. The figure shows the distance of the objective function values of the obtained solutions to the objective function values of the reference solution which corresponds to a mesh size of $h_{\max} = 0.01$. The plot contains one line for the different initial values $u_0 \equiv \hat{u} \in \{1, 3, 5, 7\}$ and shows how the distance evolves for finer meshes. Linear decay of the distances in double logarithmic scale can be observed, suggesting convergence of the front for $h_{\max} \to 0$. Similar behaviour can be observed



(a) Control (b) State (c) Obstacle

Figure 3.3: A Pareto optimal control computed with Algorithm 3 for mesh size $h_{\max} = 0.02$, initial control $u_0 \equiv 8$ and the piecewise constant obstalce ψ defined in (3.20).

in Subfigures 3.2c and 3.2d. Subfigure 3.2c shows how the distance of the obtained control to the reference control in the L^2-norm evolves for finer meshes. The distance of the corresponding states to the reference state in the H^1-norm can be seen in Subfigure 3.2d. In both subfigures, we can observe linear decay in the double logarithmic scale, indicating convergence of the controls and states computed by Algorithm 3 for finer mesh sizes.

Table 3.2 contains the number of iterations Algorithm 3 performed for the different initial values and mesh sizes. For all mesh sizes the number of iterations increase with the magnitude of the initial control u_0. For $u_0 \equiv 1$ and $u_0 \equiv 2$ there is no contact between the state and the obstacle over the course of the optimization resulting in a small number of iterations. The number of iterations does not increase for finer meshes and we expect to converge to a finite value for $h_{\max} \to 0$ for all initial values.

Configuration 2: Piecewise constant obstacle

In the second example, we choose an obstacle ψ given by a piecewise constant function defined by

$$\psi\colon \Omega \to \mathbb{R}, \quad x \mapsto \begin{cases} 1/3 & \text{if } x_1 \leq 0 \text{ and } x_2 \leq 0, \\ 1 & \text{if } x_1 \geq 0 \text{ and } x_2 \geq 0, \\ 2/3 & \text{otherwise.} \end{cases} \tag{3.20}$$

This obstacle can be interpreted analogously to that in Subsection 3.3.4. An approximate Pareto optimal control obtained by Algorithm 3 for initial value $u_0 \equiv 8$ and $h_{\max} = 0.02$ together with the corresponding state can be seen in Subfigure 3.3a and Subfigure 3.3b. The obstacle ψ, defined in (3.20), is shown in Subfigure 3.3c. Due to the nonconstant obstacle, we see a less structured behaviour in the control and state. Similarly to the first example, we observe vanishing control in areas with contact of the state with the obstacle. Algorithmically, solving this problem configuration is expected to be more challenging compared to the first configuration with the constant obstacle, as the area of contact of the state changes more dynamically over the course of the algorithm's run, in other words, the problems nondifferentiability is more pronounced.

Figure 3.4 contains a qualitative analysis of the solutions obtained by Algorithm 3 for the piecewise constant obstacle. The objective function values obtained from Algorithm 2 for different

		$h_{\max}$				
		0.2	0.1	0.05	0.02	0.01
u_0	1	10	8	37	18	17
	2	8	8	16	22	15
	3	41	85	48	48	28
	4	33	863	528	286	150
	5	324	3 135	2 381	1 254	1 013
	6	4 070	3 701	2 696	1 046	513
	7	9 344	7 907	5 079	1 705	827
	8	9 719	4 539	4 757	2 387	970

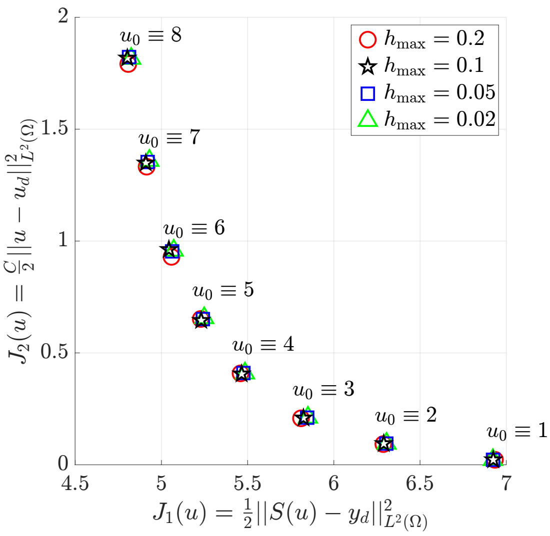
Table 3.3: Configuration 2: Number of iterations of Algorithm 3 for nonconstant obstacle.

initial values $u_0 \equiv \hat{u} \in \{1, 2, \ldots, 8\}$ and different mesh sizes $h_{\max} \in \{0.2, 0.1, 0.05, 0.02\}$ are visualized in Subfigure 3.4a. The objective function values form a front in the image space and solutions for different mesh discretizations but with same initial control cluster. This clustering is further examined in Subfigure 3.4a, where the diminishing mesh size $h_{\max}$ is plotted over the distance between the computed objective function value and the objective function value of the reference solution. We observe linear decay of the distance in double logarithmic scale. Subfigures 3.4c and 3.4d contain the distance of the obtained optimal control to reference control in the L^2-norm and the distance of the corresponding state to the reference stated in the H^1-norm, respectively. Again, we note linear decay for distances for smaller values of $h_{\max}$ in the double logarithmic scale. These plots indicate convergence of the solutions obtained by Algorithm 3 for finer meshes.

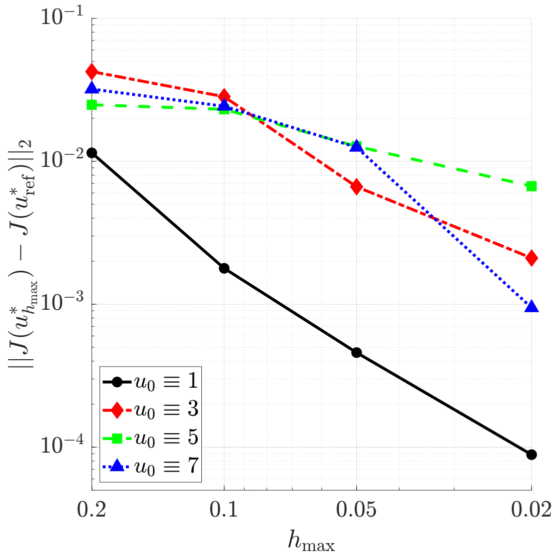
Table 3.3 contains a comparison of the number of iterations performed to reach the stopping criterion in Algorithm 3 for the different initial controls and the different meshes. We see the same trend as in the first example. However, for the piecewise constant obstacle, the iteration numbers are higher for almost all runs compared to the results for the constant obstacle. For all meshes we see an increasing number of iterations with an increasing magnitude of the initial control u_0. This is expected since for a higher magnitude of the initial control, we have more points with contact in the beginning. The number of iterations is bounded for the different mesh sizes and we expect convergence for $h_{\max} \to 0$ for all initial values of u_0.

Size of the approximated multiobjective ε-subdifferential

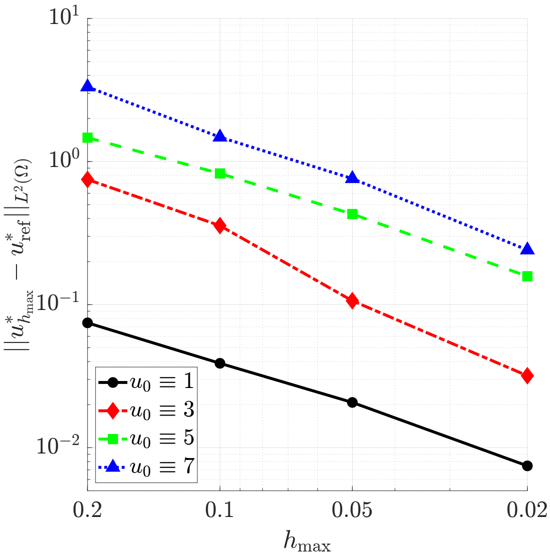
In this part, we take a closer look at Step 2 in Algorithm 3. In this step a common descent direction yielding sufficient descent for all objective functions is computed using Algorithm 2. Algorithm 2 computes a descent direction by iteratively updating an approximation Ξ_l to the multiobjective ε-subdifferential, using subderivatives of the objective functions. The number of subderivatives (i.e., the size of the final Ξ_l) depends on the number of iterations of Algorithm 2 (and the size of the I_l in Step 7). Figure 3.5 shows the number of subderivatives in the final approximated ε-subdifferential in each iteration of a run of Algorithm 3 with initial control



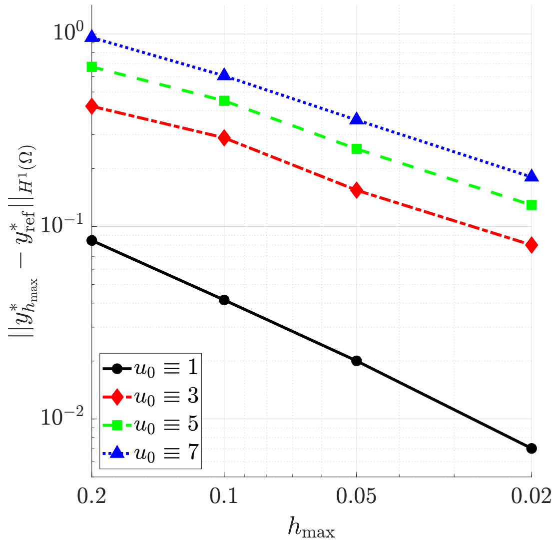
(a) Pareto fronts for different mesh discretizations.



(b) Euclidean distance of optimal values in the image space.



(c) L^2-distance between optimal control and reference control.



(d) H^1-distance between optimal state and reference state.

Figure 3.4: Qualitative analysis of the solutions derived by Algorithm 3 for different discretizations for the nonconstant obstacle. Subfigures (b)-(d) use the reference solution u^*_{ref} corresponding to mesh size $h_{\max} = 0.01$.

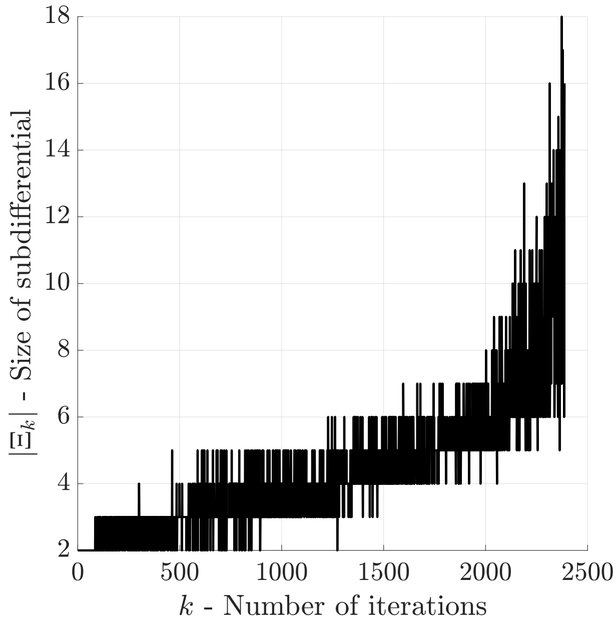


Figure 3.5: Size of the approximated subdifferential for each iteration. Results obtained by Algorithm 3 for the piecewise constant obstacle with mesh size $h_{\max} = 0.02$ and initial control $u_0 \equiv 8$.

$u_0 \equiv 8$ and maximum edge length $h_{\max} = 0.02$. We observe an increasing trend for the size of the subdifferential with the number of iterations. Up to iteration 900 the algorithm regularly only requires two subderivatives. From iteration 1 500 onwards at least four subderivatives get used in every iteration. In the end, the subdifferential consists of up to 18 subderivatives. This behaviour is not surprising: We expect the first objective function to be nonsmooth close to optima of the multiobjective control problem (3.19) (for the chosen initial control u_0), and hence, the algorithm converges to points, where the first objective function is not differentiable. To find a common descent direction in these areas, we need a sufficient number of subderivatives to describe the local behaviour of the objective function. The behaviour in Figure 3.5 can be observed across different mesh sizes and initial values. This indicates that the concept of Algorithm 3 and the approximation of the multiobjective ε-subdifferential in Algorithm 2 behave as expected.

Chapter 4

Gradient dynamical systems for convex multiobjective optimization

In this chapter, we consider the multiobjective optimization problem

$$\text{(MOP)} \qquad \min_{x\in\mathcal{H}} F(x) \coloneqq \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

with convex and continuously differentiable objective functions $f_i : \mathcal{H} \to \mathbb{R}$ for $i = 1, \ldots, m$. The main contributions of this chapter are the introduction and analysis of three novel gradient dynamical systems which are connected to the problem (MOP). Our interest in gradient dynamical systems for multiobjective optimization is motivated by the ongoing research on fast gradient methods and their relationship to accelerated gradient dynamics in scalar optimization. In the following, we provide a brief overview of the foundational ideas behind these developments.

Let $f : \mathcal{H} \to \mathbb{R}$ be a convex and continuously differentiable function with L-Lipschitz continuous gradient ∇f with $L > 0$. Consider the scalar optimization problem

$$\text{(SOP)} \qquad \min_{x\in\mathcal{H}} f(x).$$

One of the simplest iterative methods to solve the problem (SOP) is the *gradient descent method*, which dates back at least to Cauchy [62, 149]. For an initial iterate $x^0 \in \mathcal{H}$ and a fixed step size $h > 0$, define the sequence $(x^k)_{k\geq 0}$ by

$$\text{(GD)} \qquad x^{k+1} = x^k - h\nabla f(x^k), \quad \text{for} \quad k \geq 0.$$

Under the assumption $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$ and for a step size $0 < h \leq \frac{1}{L}$, it holds that $f(x^k) - \inf_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(\frac{1}{k}\right)$ as $k \to +\infty$ [181]. The method (GD) is naturally linked to the *steepest descent dynamical system*

$$\text{(SD)} \qquad \dot{x}(t) + \nabla f(x(t)) = 0.$$

The method (GD) can be derived from the system (SD) using an explicit discretization. The system (SD) shares the same asymptotical features as the method (GD). For convex and smooth functions, it holds that $f(x(t)) \to \inf_{x\in\mathcal{H}} f(x)$ as $t \to +\infty$. Further, the convergence rate $f(x(t)) - \inf_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(\frac{1}{t}\right)$ and weak convergence, i.e., $x(t) \rightharpoonup x^\infty \in \arg\min_{x\in\mathcal{H}} f(x)$, can be obtained under the additional assumption $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$. A discussion of the system (SD) is contained in Section 4.1. While the steepest descent method (GD) is straightforward to implement, in practice it suffers from slow convergence, especially for ill-conditioned problems. A general idea to improve the convergence is proposed in [196], where inertia is introduced into (GD) to obtain

$$\text{(IGD)} \qquad x^{k+1} = x^k + \beta(x^k - x^{k-1}) - h\nabla f(x^k), \quad \text{for} \quad k \geq 0,$$

with a fixed step size $h > 0$ and fixed constant $\beta > 0$. The method (IGD) is still straightforward to implement but behaves in practice better than (GD). Additionally, for strongly convex functions, it can be shown that the method (IGD) converges at an improved linear rate in comparison to (GD) [196]. The continuous version of (IGD) is the so-called *heavy ball with friction dynamical system*

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

with fixed constant $\mu, \gamma > 0$. The system (HBF) can be seen as a model of a ball with mass $\mu > 0$ rolling down the graph of the function f, subject to friction $\gamma > 0$ [19]. In the context of scalar optimization, the system (HBF) has improved properties in comparison to (SD) [4, 114]. Another way to improve the convergence of the method (GD) is presented in the seminal paper [182], where *Nesterov's accelerated gradient method* is proposed, which is given by the scheme

$$\text{(NAG)} \qquad \left.\begin{aligned} y^k &= x^k + \tfrac{k-1}{k+\alpha-1}(x^k - x^{k-1}), \\ x^{k+1} &= y^k - h\nabla f(y^k), \end{aligned}\right\} \quad \text{for} \quad k \geq 0,$$

with fixed step size $h > 0$ and a constant parameter $\alpha > 0$. Compared to (IGD) the method (NAG) does not use a constant momentum factor $\beta > 0$ but a time-dependent parameter $\frac{k-1}{k+\alpha-1}$, which converges to 1 as $k \to +\infty$. Further, in (NAG) the acceleration step and the gradient update step are separated. The system (NAG) has the following improved convergence properties [15]. Given a step size satisfies $h \leq \frac{1}{L}$, $\alpha \geq 3$ and assuming $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$, it holds that $f(x^k) - \inf_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(k^{-2}\right)$ as $k \to +\infty$. Further, if $\alpha > 3$, then $f(x^k) - \inf_{x\in\mathcal{H}} f(x) = o\left(k^{-2}\right)$ as $k \to +\infty$ and x^k converges weakly to a point in $\arg\min_{x\in\mathcal{H}} f(x)$. A key contribution to the understanding of the method (NAG) can be found in the seminal paper [218]. In this paper, it is shown that the system (NAG) can be obtained from a discretization of the *inertial gradient system with asymptotic vanishing damping*

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

For this system fast convergence rates of the function values can be shown. These results match the asymptotic convergence of (NAG) and can be found, e.g., in [13, 166]. If $\alpha \geq 3$, then $f(x(t)) - \inf_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(t^{-2}\right)$ as $t \to +\infty$. For $\alpha > 3$, it holds that $f(x(t)) - \inf_{x\in\mathcal{H}} f(x) = o\left(t^{-2}\right)$ as $t \to +\infty$ and $x(\cdot)$ converges weakly to a point in $\arg\min_{x\in\mathcal{H}} f(x)$.

It remained an open question whether similar results could be obtained in the context of multi-objective optimization [16]. This question was only recently addressed. While there have been first results on generalization of the steepest descent dynamical system in convex multiobjective optimization [17, 18] and first attempts to include inertia in this system [16], a satisfactory generalization of accelerated dynamical systems was still lacking. In this chapter, we provide a positive answer to this question and demonstrate that accelerated gradient dynamics can be extended to the multiobjective setting.

This chapter is outlined as follows. As a motivating example, in Section 4.1, we carry out the asymptotic analysis of the steepest descent dynamical system (SD) in the context of scalar optimization. This serves as the baseline for the analysis of dynamical systems in relation to optimization problems and helps to highlight the differences in the analysis of gradient dynamical systems for scalar versus multiobjective optimization problems. Section 4.2 includes a literature review on existing gradient methods for multiobjective optimization. Additionally, this section includes the analysis of the *multiobjective steepest descent dynamical system*

$$\text{(MSD)} \qquad \dot{x}(t) + \operatorname*{proj}_{C(x(t))}(0) = 0,$$

where $C(x) \coloneqq \operatorname{conv}(\{\nabla f_i(x) : i = 1, \dots, m\})$ is the convex hull of the gradients, and a comparison of (MSD) with the analysis of the system (SD). Before introducing the novel gradient dynamics for multiobjective optimization, in Section 4.3, we give an existence result for a generalized differential equation. The systems introduced in the following chapters are special instances of this generalized equation, and this way we unify the discussion of existence of solutions. In Section 4.4, we define the *inertial multiobjective gradient system*

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

with $\alpha > 0$, which generalizes the system (HBF) to the multiobjective setting. For this system, we prove weak convergence of trajectories to weakly Pareto optimal points. Building on this, in Section 4.5, we present the *multiobjective gradient system with asymptotic vanishing damping*

$$\text{(MAVD)} \qquad \frac{\alpha}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0,$$

with $\alpha > 0$. We show that this system generalizes (AVD) in a satisfactory way to the multiobjective setting, giving fast convergence of the function values and weak convergence of the trajectories to weakly Pareto optimal points. Finally, in Section 4.6, we consider a further modification of the system (MAVD), namely, the *multiobjective Tikhonov regularized inertial gradient system*

$$\text{(MTRIGS)} \qquad \frac{\alpha}{t^q}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t)}(0) = 0,$$

with constant parameters $\alpha, \beta > 0$, $q \in (0, 1]$, $p \in (0, 2]$. The system (MTRIGS) adapts (MAVD) by a vanishing Tikhonov regularization term which improves the convergence properties of the system. We discuss the system (MTRIGS) for different values of the parameters, and prove fast convergence rates of the function values and convergence of the trajectories $x(\cdot)$ to weakly Pareto optimal points. Convergence is achieved in either the weak or strong topology of the underlying Hilbert space, depending on the choice of parameters.

The content of this chapter is based on the following publications. Specific references to these publications are provided in the introductions of the respective sections.

[49] Boţ, R. I. and Sonntag, K. *Inertial dynamics with vanishing Tikhonov regularization for multiobjective optimization.* In: *Journal of Mathematical Analysis and Applications* 554 (2) (2025). doi: 10.1016/j.jmaa.2025.129940.

[216] Sonntag, K. and Peitz, S. *Fast convergence of inertial multiobjective gradient-like systems with asymptotic vanishing damping.* In: *SIAM Journal on Optimization* 34 (3) (2024), pp. 2259–2286. doi: 10.1137/23M1588512.

[217] Sonntag, K. and Peitz, S. *Fast Multiobjective Gradient Methods with Nesterov Acceleration via Inertial Gradient-Like Systems.* In: *Journal of Optimization Theory and Applications* 201 (2024), pp. 539–582. doi: 10.1007/s10957-024-02389-3.

4.1 An introductory example from scalar optimization

In this section, we demonstrate the general procedure for analyzing a gradient dynamical system related to an optimization problem. This serves as a template for the analysis of the novel multiobjective gradient systems introduced in later sections. Furthermore, comparing the multiobjective gradient systems with the system discussed in this section, allows to highlight the challenges of using gradient systems in multiobjective optimization. We outline the general analysis by means of the following problem from scalar optimization. Consider the optimization problem

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

with a convex and continuously differentiable objective function $f : \mathcal{H} \to \mathbb{R}$ with Lipschitz continuous gradient ∇f. To the scalar optimization problem (SOP), we associate the *steepest descent dynamical system*

$$\text{(SD)} \qquad \dot{x}(t) + \nabla f(x(t)) = 0, \quad \text{for} \quad t > t_0,$$

with initial data $t_0 > 0$ and $x(t_0) = x_0 \in \mathcal{H}$. The system (SD), in connection with (SOP), is well-studied in the literature under various assumptions on the objective function f [54, 196, 199, 208]. The results for the convex case discussed in this section can be found in [7, 51, 52, 53, 56], while the behavior of (SD) in the nonconvex setting is more involved, as counterexamples show [1, 80, 188].

In this thesis, the analysis of gradient dynamical systems follows a consistent structure, described as follows:

1. Discussion of existence and uniqueness of solutions;
2. Preparatory results;
3. Asymptotic analysis.

Before analyzing the properties of the solutions, we verify the existence of solutions to the dynamical system under consideration and discuss their uniqueness. Then, we collect preparatory results, which include energy estimates and statements on the boundedness and regularity of solutions. Finally, we present asymptotic results, which are the main focus of the analysis. The asymptotic analysis contains convergence rates of function values $f(x(t))$, as well as statements on weak or strong convergence of trajectories $x(t)$ to an optimal point of the considered optimization problem.

Discussion of existence and uniqueness of solutions

In this part, we provide a formal definition of a solution to (SD), followed by a theorem that states the existence and uniqueness of solutions, given the objective function f is sufficiently smooth.

Definition 4.1.1. *A function $x : [t_0, +\infty) \to \mathcal{H}$ is called a solution to* (SD) *if it satisfies the following properties:*

i) $x \in C^1([t_0, +\infty), \mathcal{H})$, *i.e.,* $x(\cdot)$ *is continuously differentiable;*

ii) $x(t_0) = x_0$;

iii) For all $t > t_0$ *it holds that* $\dot{x}(t) + \nabla f(x(t)) = 0$.

Theorem 4.1.2. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be continuously differentiable with Lipschitz continuous gradient* ∇f*. Then, for all* $t_0 > 0$ *and* $x_0 \in \mathcal{H}$ *there exists a unique solution* $x(\cdot)$ *to* (SD) *in the sense of Definition 4.1.1.*

Proof. The proof follows immediately by the Cauchy–Lipschitz Theorem (Theorem 2.2.2). □

Preparatory results

In this part, we show that the function f is an energy function for the system (SD). Generally speaking, the term energy function refers to any function that depends on a solution $x(\cdot)$ of (SD) and is monotonically decreasing with respect to time.

Proposition 4.1.3. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be continuously differentiable with L-Lipschitz continuous gradient* ∇f *and let* $x(\cdot)$ *be a solution to* (SD) *in the sense of Definition 4.1.1. Then, for all* $t > t_0$

$$\frac{d}{dt} f(x(t)) = -\|\dot{x}(t)\|^2 \leq 0. \tag{4.1}$$

Proof. The proof follows immediately by applying the chain rule and using equation (SD). Let $t > t_0$, then

$$\frac{d}{dt} f(x(t)) = \langle \nabla f(x(t)), \dot{x}(t) \rangle = -\|\dot{x}(t)\|^2 \leq 0.$$

□

Corollary 4.1.4. *Let* $f : \mathcal{H} \to \mathbb{R}$ *be continuously differentiable with L-Lipschitz continuous gradient* ∇f *and let* $x : [t_0, +\infty) \to \mathcal{H}$ *be a solution to* (MSD) *in the sense of Definition 4.1.1. If* f *is bounded from below, then*

$$\int_{t_0}^{+\infty} \|\dot{x}(s)\|^2 \, ds < +\infty.$$

Proof. By integration inequality (4.1) in Proposition 4.1.3, we have for all $t \geq t_0$

$$f(x(t)) \leq f(x(t_0)) - \int_{t_0}^{t} \|\dot{x}(s)\|^2 \, ds. \tag{4.2}$$

From inequality (4.2), we follow

$$\int_{t_0}^{+\infty} \|\dot{x}(s)\|^2 \, ds \leq f(x(t_0)) - \inf_{x \in \mathcal{H}} f(x). \tag{4.3}$$

Since f is bounded from below, the right-hand side of (4.3) is bounded and the statement follows. □

Asymptotic analysis

In this part, we present the asymptotic analysis of solutions $x(\cdot)$ to (SD). The term asymptotic analysis refers to the examination of the properties of the solution $x(t)$ as $t \to +\infty$. First, we prove that for any smooth convex objective function f the function values along the trajectories converge to the minimal value, i.e, $f(x(t)) \to \inf_{x\in\mathcal{H}} f(x)$ as $t \to +\infty$. Afterwards, under the condition $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$, we show that $f(x(t)) - \inf_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(\frac{1}{t}\right)$ as $t \to +\infty$. Additionally, we prove under this condition that solutions $x(\cdot)$ converge weakly to optimal points of (SOP), i.e., $x(t) \rightharpoonup x^\infty \in \arg\min_{x\in\mathcal{H}} f(x)$ as $t \to +\infty$.

Theorem 4.1.5. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f and let $x(\cdot)$ be a solution to* (SD) *in the sense of Definition 4.1.1. Then*

$$\lim_{t\to+\infty} f(x(t)) = \inf_{x\in\mathcal{H}} f(x).$$

Proof. For $z \in \mathcal{H}$, define the anchored energy function

$$\mathcal{E}_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto (t - t_0)\left(f(x(t)) - f(z)\right) + \frac{1}{2}\|x(t) - z\|^2. \tag{4.4}$$

This function is continuously differentiable and we compute the derivative using the chain rule to obtain

$$\begin{aligned} \frac{d}{dt}\mathcal{E}_z(t) =&(t - t_0)\langle \nabla f(x(t)), \dot{x}(t)\rangle + f(x(t)) - f(z) + \langle x(t) - z, \dot{x}(t)\rangle \\ \leq& -(t - t_0)\|\dot{x}(t)\|^2, \end{aligned}$$

where the last inequality follows from the convexity of f (see Proposition 2.1.24). Therefore, the function $\mathcal{E}_z(\cdot)$ is monotonically decreasing and we follow for all $t \geq t_0$

$$(t - t_0)(f(x(t)) - f(z)) \leq \mathcal{E}_z(t) \leq \mathcal{E}_z(t_0) = \frac{1}{2}\|x_0 - z\|^2. \tag{4.5}$$

From inequality (4.5), we obtain for all $t > t_0$

$$f(x(t)) - f(z) \leq \frac{\|x(t_0) - z\|}{2(t - t_0)}. \tag{4.6}$$

By Proposition 4.1.3, $t \mapsto f(x(t))$ is monotonically decreasing and hence $\lim_{t\to+\infty} f(x(t)) \in \mathbb{R} \cup \{-\infty\}$ exists. From inequality (4.6), we conclude $\lim_{t\to+\infty} f(x(t)) - f(z) \leq 0$ for all $z \in \mathcal{H}$ and hence $\lim_{t\to+\infty} f(x(t)) = \inf_{x\in\mathcal{H}} f(x)$. □

Example 4.1.6. *Theorem 4.1.5 states that any solution $x(\cdot)$ of* (SD) *minimizes the objective function f. However, this result is to some extend unsatisfactory as it does not provide further information on how long it will take for $\lim_{t\to+\infty} f(x(t)) = \inf_{x\in\mathcal{H}} f(x)$ to converge. In the following example, we show that even for functions which are bounded from below the convergence can be arbitrary slow. Let $\mathcal{H} = \mathbb{R}$ with the euclidean inner product and norm. For a fixed $p \in (0, 1)$, consider the objective function*

$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} -px + p + 1, & \text{if } x \leq 1, \\ \frac{1}{x^p}, & \text{else.} \end{cases} \tag{4.7}$$

The function f is convex and continuously differentiable with globally Lipschitz continuous gradient. Consider the system

$$\text{(SD)} \qquad \dot{x}(t) + \nabla f(x(t)) = 0, \quad \textit{for} \quad t > t_0,$$

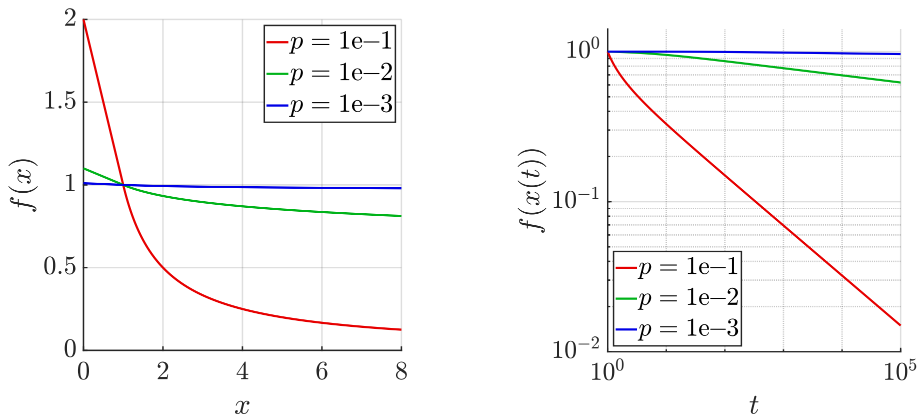
with $t_0 = 1$ and $x(t_0) = 1$. For the objective function f defined in (4.7)*, the unique solution to* (SD) *is given by $x(t) = (1+p(p+2)(t-1))^{\frac{1}{p+2}}$. As Theorem 4.1.5 states, we observe convergence to the optimal value*

$$f(x(t)) = (1 + p(p+2)(t-1))^{-\frac{p}{p+2}} \to 0 = \inf_{x \in \mathcal{H}} f(x), \quad \textit{as} \quad t \to +\infty. \tag{4.8}$$

From (4.8)*, we infer the convergence rate*

$$f(x(t)) - \inf_{x \in \mathcal{H}} f(x) = \mathcal{O}\left(t^{-\frac{p}{p+2}}\right), \quad \textit{as} \quad t \to +\infty. \tag{4.9}$$

In the beginning of this example the parameter $p \in (0,1)$ was chosen arbitrarily. For $p \to 0$ the convergence in (4.9) *gets arbitrary slow.*



(a) (b)

Figure 4.1: Function f defined in (4.7) and function values $f(x(t))$ of solutions $x(\cdot)$ to (SD) for different values of $p \in \{1\text{e}{-1}, 1\text{e}{-2}, 1\text{e}{-3}\}$.

The slow decay of the function values which is formally described in (4.9) *is illustrated in Figure 4.4. Subfigure 4.4a shows the objective function f for different values of $p \in \{1\text{e}{-1}, 1\text{e}{-2}, 1\text{e}{-3}\}$. For all choices of p, we have $\inf_{x\in\mathbb{R}} f(x) = \lim_{x\to+\infty} f(x) = 0$. For smaller values of p the functions decay slower as $x \to +\infty$. Subfigure 4.4b shows the function values $f(x(t))$ along solutions $x(\cdot)$ to* (MSD)*, for $t \in [1, 1\text{e}5]$ and different values of $p \in \{1\text{e}{-1}, 1\text{e}{-2}, 1\text{e}{-3}\}$. For $p = 1\text{e}{-1}$ and $p = 1\text{e}{-2}$, we observe decay in the function values. For the smallest values $p = 1\text{e}{-3}$, the function $t \mapsto f(x(t))$ appears nearly constant and decays extremely slowly, as we expect from* (4.9)*.*

From Theorem 4.1.5, we derive the following corollary.

Corollary 4.1.7. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f and let $x(\cdot)$ be a solution to* (SD) *in the sense of Definition 4.1.1. Then, the following are equivalent:*

i) *$x(\cdot)$ is bounded;*

ii) $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$.

Proof. Assume $x(\cdot)$ is bounded. Then $x(\cdot)$ possesses a weak sequential cluster point, i.e., there exists a sequence $(t_k)_{k\geq 0} \subset [t_0, +\infty)$ with $t_k \to +\infty$ and $x(t_k) \rightharpoonup x^\infty \in \mathcal{H}$ as $k \to +\infty$. Since the function f is convex and continuous it is weakly lower semicontinuous and we follow

$$f(x^\infty) \leq \liminf_{k\to+\infty} f(x(t_k)) = \lim_{t\to+\infty} f(x(t)) = \inf_{x\in\mathcal{H}} f(x),$$

where the existence of the limit follows from Theorem 4.1.5. Hence, $x^\infty \in \arg\min_{x\in\mathcal{H}} f(x)$.

Assume $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$ and let $x^* \in \arg\min_{x\in\mathcal{H}} f(x)$. In the proof of Theorem 4.1.5, it is shown that

$$\mathcal{E}_{x^*} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto (t - t_0)(f(x(t)) - f(x^*)) + \frac{1}{2}\|x(t) - x^*\|^2,$$

is monotonically decreasing. Since $f(x(t)) - f(x^*) \geq 0$, for all $t \geq t_0$

$$\frac{1}{2}\|x(t) - x^*\|^2 \leq \mathcal{E}_{x^*}(t) \leq \mathcal{E}_{x^*}(t_0) = \frac{1}{2}\|x(t_0) - x^*\|^2,$$

and hence $x(\cdot)$ is bounded. □

The convergence of solutions $x(\cdot)$ to (SD) can be very slow as Example 4.1.6 demonstrates. However, the slow convergence occurs primarily because the chosen objective function is intentionally constructed to cause slow convergence. The objective function f defined in (4.7) is bounded from below but does not posses a minimizer. In the next theorem, we show that solutions $x(\cdot)$ exhibit better asymptotic properties under the condition $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$.

Theorem 4.1.8. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f and let $x(\cdot)$ be a solution to* (SD) *in the sense of Definition 4.1.1. Further, assume $S := \arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$. Then, for all $t \geq t_0$*

$$f(x(t)) - \inf_{x\in\mathcal{H}} f(x) \leq \frac{\operatorname{dist}(x_0, S)^2}{2(t - t_0)}. \tag{4.10}$$

Proof. Fix $x^* \in \arg\min_{x\in\mathcal{H}} f(x)$. We differentiate $\frac{1}{2}\|x(t) - x^*\|^2$ with respect to t to obtain

$$\frac{d}{dt}\frac{1}{2}\|x(t) - x^*\|^2 = \langle x(t) - x^*, \dot{x}(t)\rangle = \langle x^* - x(t), \nabla f(x(t))\rangle \leq f(x^*) - f(x(t)),$$

where the last inequality follows by the convexity of f due to Proposition 2.1.24. Hence,

$$f(x(t)) - f(x^*) \leq -\frac{d}{dt}\frac{1}{2}\|x(t) - x^*\|^2. \tag{4.11}$$

Integrating inequality (4.11) from t_0 to $t \geq t_0$ gives

$$\int_{t_0}^{t} f(x(s)) - f(x^*)\, ds \leq \frac{1}{2}\|x(t_0) - x^*\|^2 - \frac{1}{2}\|x(t) - x^*\|^2 \leq \frac{1}{2}\|x(t_0) - x^*\|^2. \tag{4.12}$$

Proposition 4.1.3 states that $t \mapsto f(x(t))$ is monotonically decreasing and hence

$$(t - t_0)(f(x(t)) - f(x^*)) = \int_{t_0}^{t} f(x(t)) - f(x^*)\, ds \leq \int_{t_0}^{t} f(x(s)) - f(x^*)\, ds. \tag{4.13}$$

Combining (4.12) and (4.13), we have

$$(t - t_0)(f(x(t)) - f(x^*)) \leq \frac{1}{2}\|x(t_0) - x^*\|^2. \tag{4.14}$$

and therefore

$$f(x(t)) - \inf_{x \in \mathcal{H}} f(x) \leq \frac{\|x_0 - x^*\|^2}{2(t - t_0)}$$

Since this bound is uniform with respect to $x^* \in \arg\min_{x \in \mathcal{H}} f(x)$, we can apply the infimum to obtain (4.10). □

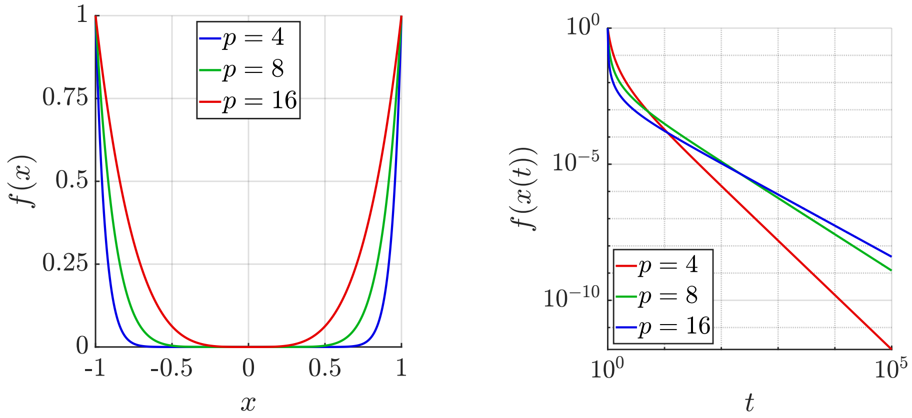
Example 4.1.9. *In the following example, we demonstrate the sharpness of Theorem 4.1.8. For $\mathcal{H} = \mathbb{R}$ with the euclidean inner product and norm, consider the optimization problem* (SOP) *with the objective function*

$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto |x|^p, \tag{4.15}$$

for $p > 2$. The function f is convex and continuously differentiable with locally Lipschitz continuous gradient. (Despite the fact that the Lipschitz continuity is merely local, the preceding theorems still apply because solutions to (SD) *remain in compact sets.) For the objective function f defined in* (4.15)*, the unique solution to* (SD) *with initial data with $t_0 = 1$ and $x(t_0) = 1$ is given by $x(t) = (1 + p(p-2)(t-1))^{\frac{1}{2-p}}$. Therefore,*

$$f(x(t)) - \inf_{x \in \mathcal{H}} f(x) = (1 + p(p-2)(t-1))^{\frac{p}{2-p}} = \mathcal{O}\left(\frac{1}{t^{\frac{p}{p-2}}}\right), \quad as \quad t \to +\infty. \tag{4.16}$$

We observe $\frac{p}{p-2} \to 1$ as $p \to +\infty$ and hence we get closer to the asymptoical rate $f(x(t)) = \mathcal{O}\left(\frac{1}{t}\right)$ for bigger values of p.



(a) (b)

Figure 4.2: Function f defined in (4.15) and function values $f(x(t))$ of solutions $x(\cdot)$ to (SD) for different values of $p \in \{4, 8, 16\}$.

The asymptotic convergence rates summarized in (4.16) *are visualized in Figure 4.2. Subfigure 4.2a shows the function f, defined in* (4.15), *for different values of $p \in \{4, 8, 16\}$. For larger values of p, the function f is flatter around the global minimum $x^* = 0$. This leads to smaller gradients near the optimum and hence slower convergence of $x(\cdot)$ to the optimum $x^* = 0$. This effect is illustrated in Subfigure 4.2b. For the largest value, $p = 16$, the asymptotic convergence is the slowest, as seen from the slope of $t \mapsto f(x(t))$ in the interval $t \in [1\text{e}1, 1\text{e}5]$ while for the smallest value, $p = 2$, the asymptotic decay is the fastest. Initially, in the interval $t \in [1, 1\text{e}1]$, this trend is reversed, but only because the function $x \mapsto |x|^p$ is steeper at the initial point $x_0 = 1$ for higher values of p. However, this does not affect the asymptotic convergence.*

Theorem 4.1.10. *Let $f : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f and let $x(\cdot)$ be a solution to* (SD) *in the sense of Definition 4.1.1. Further, assume $S := \arg\min_{x \in \mathcal{H}} f(x) \neq \emptyset$. Then*

$$x(t) \rightharpoonup x^\infty \in S, \quad \text{as} \quad t \to +\infty. \tag{4.17}$$

Proof. We apply Opial's Lemma (Lemma 2.1.6) to prove the weak convergence of $x(\cdot)$. Since $S \neq \emptyset$ holds by assumption, we only have to verify that each sequential cluster point of $x(\cdot)$ belongs to S and that for all $z \in S$ the limit $\lim_{t\to+\infty} \|x(t) - z\|$ exists.

Let x^∞ be a weak sequential cluster point of $x(\cdot)$. Hence, there exists a sequence $(t_k)_{k\geq 0}$ with $t_k \to +\infty$ and $x(t_k) \rightharpoonup x^\infty$ as $k \to +\infty$. By the weak lower semicontinuity of f we have

$$f(x^\infty) \leq \liminf_{k\to+\infty} f(x(t_k)) = \lim_{t\to+\infty} f(x(t)) = \inf_{x\in\mathcal{H}} f(x),$$

where the equalities follow from the monotonic decay of $t \mapsto f(x(t))$ and Theorem 4.1.5. Therefore, we conclude $x^\infty \in S$.

For all $z \in S$ it holds that

$$\frac{d}{dt}\frac{1}{2}\|x(t) - z\|^2 = \langle x(t) - z, \dot{x}(t)\rangle = \langle z - x(t), \nabla f(x(t))\rangle \leq 0,$$

where the last inequality follows by Proposition 2.1.24 and the convexity of f. Then, $\|x(t)-z\| \geq 0$ is monotonically decreasing in t and hence convergent.

Therefore, all conditions of Opial's Lemma (Lemma 2.1.6) are verified and in total we follow

$$x(t) \rightharpoonup x^\infty \in S = \operatorname*{arg\,min}_{x\in\mathcal{H}} f(x), \quad \text{as} \quad t \to +\infty.$$

□

Remark 4.1.11. *Under additional assumptions on the objective function, better convergence rates of the function values and improved convergence properties of $x(\cdot)$ can be established. For example, if the objective function f is strongly convex, the function values converge linearly to the optimal value, and the solution $x(\cdot)$ converges strongly to the unique minimizer of the optimization problem [199]. The contribution of this thesis focuses on the analysis of gradient dynamical systems for multiobjective optimization under the mere assumption of convexity. Therefore, we do not restate results that require stronger assumptions, but refer the reader to the literature cited in the beginning of this section.*

4.2 Review of existing gradient systems for multiobjective optimization

4.2.1 The multiobjective steepest descent dynamical system (MSD)

The first multiobjective gradient system discussed in this review is the *multiobjective steepest descent dynamical system*

$$\text{(MSD)} \qquad \dot{x}(t) + \operatorname*{proj}_{C(x(t))}(0) = 0, \quad \text{for} \quad t > t_0,$$

with initial data $t_0 > 0$ and $x(t_0) = x_0 \in \mathcal{H}$, and where $C(x) \coloneqq \operatorname{conv}(\{\nabla f_i(x) : i = 1, \ldots, m\})$ is the convex hull of the gradients $\nabla f_i(x)$ for $i = 1, \ldots, m$. This system can equivalently be written as $\dot{x}(t) = \phi(x(t))$, where $\phi : \mathcal{H} \to \mathcal{H}$, $x \mapsto \phi(x) \coloneqq -\operatorname{proj}_{C(x)}(0)$, is the multiobjective steepest descent direction, introduced in Subsection 2.3.4. Therefore, the system (MSD) can be seen as the multiobjective counterpart to the *steepest descent dynamical system*

$$\text{(SD)} \qquad \dot{x}(t) + \nabla f(x(t)) = 0,$$

for scalar optimization problems $\min_{x \in \mathcal{H}} f(x)$, with a smooth objective function $f : \mathcal{H} \to \mathbb{R}$. The system (SD) is extensively discussed in the previous section. In this section, we show that the results obtained for the system (SD) can be recovered in the context of multiobjective optimization for the system (MSD). On the other hand, the system (MSD) can be seen as the continuous-time counterpart of the *multiobjective gradient method* (MGD), which can be written as

$$\text{(MGD')} \qquad x^{k+1} = x^k - h \operatorname*{proj}_{C(x^k)}(0), \quad \text{for} \quad k \geq 0,$$

for an initial iterate $x^0 \in \mathcal{H}$ and step size $h > 0$. The scheme (MGD') can be obtained from an explicit discretization of (MSD). We show that we recover the results for the method (MGD') summarized in Subsection 2.3.4 in the continuous-time setting.

The first systems related to (MSD) were studied in the context of economics [127, 214], and in particular for resource allocation problems [79]. In [170, 207], the system (MSD) is analyzed in the context of multiobjective optimization where the existence of solutions to (MSD) and the convergence of solutions $x(\cdot)$ to (MSD) to Pareto optimal points of (MOP) are investigated. Further variants of (MSD) are proposed in [16, 17, 18], and we present these in the following subsections of this review.

In this subsection, we carry out the asymptotic analysis of the system (MSD), analogous to the analysis of the system (SD). This helps us to highlight the differences between gradient systems for scalar and multiobjective optimization. The techniques applied in this subsection are fundamental for the analysis of the novel multiobjective gradient system presented in this chapter.

Assumptions

In this subsection, we make the following assumptions on the objective functions.

($\mathcal{A}_1$) The objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and continuously differentiable with L-Lipschitz continuous gradients $\nabla f_i : \mathcal{H} \to \mathcal{H}$ with $L > 0$ for all $i = 1, \dots, m$.

($\mathcal{A}_2$) For all $x_0 \in \mathcal{H}$ and for all $x \in \mathcal{L}(F, F(x_0))$ it holds that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ and further

$$R := \sup_{F^* \in F(\mathcal{LP}_w(F,F(x_0)))} \inf_{z \in F^{-1}(\{F^*\})} \frac{1}{2}\|z - x_0\|^2 < +\infty. \tag{4.18}$$

At the end of this subsection, we discuss Assumption ($\mathcal{A}_2$) in the context of the system (MSD) to demonstrate that it is a natural assumption in the analysis ot the asymptotic properties of multiobjective gradient systems.

Existence and uniqueness of solutions

In this part, we present a proof of the existence of solutions to (MSD) and a discussion of their uniqueness. In contrast to the analogous analysis in Section 4.1 on the system (SD), the results obtained for the system (MSD) are weaker. Even when the gradients ∇f_i of the objective functions are Lipschitz continuous, the multiobjective steepest descent direction is in general only Hölder continuous (see Proposition 2.3.21 and Remark 2.3.22). Therefore, we cannot use the Cauchy–Lipschitz Theorem, which guarantees the existence of a unique global solution and which is applicable in arbitrary Hilbert spaces. Instead, we apply Peano's Theorem to obtain a solution, though this approach has certain limitations. First, it is applicable only in finite-dimensional Hilbert spaces. This limitation is restrictive, although the asymptotic analysis of solutions remains valid in general Hilbert spaces. Moreover, when we derive an optimization method from the gradient system by a numerical discretization scheme, we do not work with an ODE, and this limitation becomes irrelevant for the derivation of optimization algorithms. Second, Peano's Theorem only gives local solutions and we must put in additional effort to extend local to global solutions. This is not a major drawback, since it can be achieved using standard techniques based on Gronwall-type arguments and Zorn's Lemma. Third, the solutions are not necessarily unique. However, we argue that non-uniqueness is acceptable in our setting, since the asymptotic analysis, which is our main focus, applies to all solutions. In the following, we formally define a solution to (MSD) and then prove the main existence result, and conclude with a remark on the uniqueness.

Definition 4.2.1. *A function $x : [t_0, +\infty) \to \mathcal{H}$ is called a solution to* (MSD) *if it satisfies the following properties:*

i) $x \in C^1([t_0, +\infty), \mathcal{H})$*, i.e.,* $x(\cdot)$ *is continuously differentiable;*

ii) $x(t_0) = x_0$*;*

iii) For all $t > t_0$ *it holds that* $\dot{x}(t) + \operatorname{proj}_{C(x(t))}(0) = 0$*.*

Before proving the existence of solutions, we introduce a different description of the multiobjective steepest descent direction $-\operatorname{proj}_{C(x(t))}(0)$ in the following remark, which is more convenient for certain parts of the analysis.

Remark 4.2.2. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Then for all $t > t_0$, there exists $\theta(t) \in \Delta^m$ such that*

$$\dot{x}(t) = -\operatorname*{proj}_{C(x(t))}(0) = -\sum_{i=1}^{m} \theta_i(t) \nabla f_i(x(t)). \tag{4.19}$$

In the following subsection, whenever we use $\theta(t)$ we refer to the weights implicitly given by (4.19)*. The mapping*

$$\theta : [t_0, +\infty) \to \Delta^m, \quad t \mapsto \theta(t),$$

which is implicitly defined by (4.19)*, is in general measurable and has better regularity properties under additional assumptions. Since the analysis of* (MSD) *does not require $\theta(\cdot)$ to satisfy further conditions, we do not discuss these properties here.*

Theorem 4.2.3. *Assume $\mathcal{H}$ is finite dimensional. Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$. Then, for all $t_0 > 0$ and $x_0 \in \mathcal{H}$ there exists a solution $x(\cdot)$ to* (SD) *in the sense of Definition 4.2.1.*

Proof. By assumption $\dim(\mathcal{H}) < +\infty$ and by Proposition 2.3.21 the mapping $\phi : \mathcal{H} \to \mathcal{H}$, $x \mapsto -\operatorname{proj}_{C(x)}(0)$ is Hölder continuous. Therefore, we can use Peano's Theorem (Theorem 2.2.1) to conclude the existence of a local solution to (MSD), i.e., there exists $T > t_0$ and a function $x \in C^1([t_0, T), \mathcal{H})$ with $x(t_0) = x_0$ and $\dot{x}(t) + \operatorname{proj}_{C(x(t))}(0) = 0$ for all $t \in (t_0, T)$. We extend this solution to a global one using a standard technique based on Gronwall-type arguments and Zorn's Lemma.

First, we establish a growth property of $\phi(\cdot)$ which is necessary to guarantee that solutions $x(\cdot)$ do not become unbounded in finite time. Using the weights $\theta(t) \in \Delta^m$ defined in (4.19), we write for all $t > t_0$

$$\phi(x(t)) = -\sum_{i=1}^{m} \theta_i(t) \nabla f_i(x(t)).$$

Then, it holds that

$$\begin{aligned}
\|\phi(x(t))\| = \left\| \sum_{i=1}^{m} \theta_i(t) \nabla f_i(x(t)) \right\| &\leq \left\| \sum_{i=1}^{m} \theta_i(t) \nabla f_i(x(t_0)) \right\| + \left\| \sum_{i=1}^{m} \theta_i(t) \left(\nabla f_i(x(t)) - \nabla f_i(x(t_0)) \right) \right\| \\
&\leq \left\| \sum_{i=1}^{m} \theta_i(t) \nabla f_i(x(t_0)) \right\| + \sum_{i=1}^{m} \theta_i(t) \|\nabla f_i(x(t)) - \nabla f_i(x(t_0))\| \\
&\leq \sum_{i=1}^{m} \theta_i(t) \|\nabla f_i(x(t_0))\| + \sum_{i=1}^{m} \theta_i(t) L \|x(t) - x(t_0)\| \\
&\leq \max_{i=1,\dots,m} \|\nabla f_i(x(t_0))\| + L\|x(t) - x(t_0)\|,
\end{aligned}$$

and hence

$$\|\phi(x(t))\| \leq c + L\|x(t) - x(t_0)\|, \tag{4.20}$$

with $c := \max_{i=1,\dots,m} \|\nabla f_i(x(t_0))\|$.

With the growth condition (4.20) in place, we proceed to show the existence of a global solution using Zorn's Lemma. Define the set

$$\mathfrak{S} := \Big\{ (x,T) : T \in (t_0, +\infty] \quad \text{and} \quad x \in C^1([t_0,T), \mathcal{H}), \quad x(t_0) = x_0, \\ \text{and} \quad \dot{x}(t) + \operatorname*{proj}_{C(x(t))}(0) = 0 \quad \text{for all} \quad t \in [t_0, T) \Big\}.$$

(Note that $T \in (t_0, +\infty]$ in the definition of $\mathfrak{S}$ allows for the value $+\infty$ for T.) By Peano's Theorem, as we stated in the beginning of the proof, the set $\mathfrak{S}$ is nonempty. On $\mathfrak{S}$ we define the reflexive, transitive and antisymmetric partial order

$$(x_1, T_1) \preccurlyeq (x_2, T_2) \quad :\Longleftrightarrow \quad T_1 \leq T_2 \quad \text{and} \quad x_1(t) = x_2(t) \quad \text{for all} \quad t \in [t_0, T_1).$$

Next, we show that every nonempty, totally ordered subset of $\mathfrak{S}$ has an upper bound in $\mathfrak{S}$. Let $\mathfrak{C} \subseteq \mathfrak{S}$ be a nonempty, totally ordered subset of $\mathfrak{S}$. Define

$$T_{\mathfrak{C}} := \sup \left\{ T : (x,T) \in \mathfrak{C} \right\},$$

and

$$x_{\mathfrak{C}} : [t_0, T_{\mathfrak{C}}) \to \mathcal{H}, \quad t \mapsto x_{\mathfrak{C}}(t) := x(t) \quad \text{for} \quad t < \bar{t} < T_{\mathfrak{C}} \quad \text{and} \quad (x, \bar{t}) \in \mathfrak{C}.$$

By construction, $(x_{\mathfrak{C}}, T_{\mathfrak{C}}) \in \mathfrak{S}$ and $(x,T) \preccurlyeq (x, T_{\mathfrak{C}})$ for all $(x,T) \in \mathfrak{C}$. Hence, $(x_{\mathfrak{C}}, T_{\mathfrak{C}})$ is an upper bound of $\mathfrak{C}$ in $\mathfrak{S}$. Then, by Zorn's Lemma (see [61, 120]), there exists a maximal element $(x,T) \in \mathfrak{S}$. If $T = +\infty$, then $x(\cdot)$ is a solution to (MSD) in the sense of Definition 4.2.1. We show that $T = +\infty$ must hold by contradiction.

Assume $T < +\infty$. Define the function

$$h : [t_0, T) \to \mathbb{R}, \quad t \mapsto h(t) := \|x(t) - x(t_0)\|.$$

Then, by the chain rule and the Cauchy–Schwarz inequality, we derive

$$\frac{d}{dt}\frac{1}{2} h(t)^2 = \langle x(t) - x(t_0), \dot{x}(t) \rangle \leq \|x(t) - x(t_0)\| \, \|\dot{x}(t)\| = h(t) \|\phi(x(t))\|. \tag{4.21}$$

Starting from (4.21) and using (4.20), we derive the bound

$$\frac{d}{dt}\frac{1}{2} h(t)^2 \leq h(t) \left(c + L h(t) \right). \tag{4.22}$$

We follow the boundedness of $h(t)$ from (4.22) as in [16, Theorem 3.5]. Let $\varepsilon > 0$ and consider the function $h(\cdot)$ on the closed interval $[t_0, T - \varepsilon]$. Integrating inequality (4.22) from t_0 to $t > t_0$ and using $h(t_0) = 0$ gives for all $t \in [t_0, T - \varepsilon]$

$$\frac{1}{2} h(t)^2 \leq \int_{t_0}^{t} (c + L h(s)) h(s) \, ds.$$

The function $h(\cdot)$ is continuous and therefore $(c+Lh(\cdot)) \in L^1([t_0, T-\varepsilon], \mathbb{R})$. Then, we can apply Lemma 2.2.10 to follow for all $t \in [t_0, T-\varepsilon]$

$$h(t) \leq \int_{t_0}^{t} c + Lh(s)\, ds \leq cT + \int_{t_0}^{t} Lh(s)\, ds. \tag{4.23}$$

We apply Lemma 2.2.9 to (4.23) and get for all $t \in [t_0, T-\varepsilon]$

$$h(t) \leq cT \exp\left(L(t-t_0)\right) \leq cT \exp\left(L(T-t_0)\right) < +\infty.$$

Since this bound is uniform in $\varepsilon > 0$ and $t > t_0$, $h(\cdot)$ is bounded on $[t_0, T)$. Then, by (4.20) the velocity $\dot{x}(\cdot)$ is bounded as well on $[t_0, T)$, and by the continuity of $\dot{x}(\cdot)$, we follow the existence of

$$x_T := x_0 + \int_{t_0}^{T} \dot{x}(s)\, ds \in \mathcal{H}.$$

Using Peano's Theorem, as in the beginning of the proof, we conclude the existence of $\delta > 0$ and a solution $\hat{x} : [T, T+\delta) \to \mathcal{H}$ that satisfies $\hat{x} \in C^1([T, T+\delta)), \mathcal{H})$, $\hat{x}(t) = x_T$ and $\dot{\hat{x}}(t) + \operatorname{proj}_{C(\hat{x}(t))}(0) = 0$ for all $t \in (T, T+\delta)$. Then, we define

$$x^* : [t_0, T+\delta) \to \mathcal{H}, \quad t \mapsto \begin{cases} x(t), & \text{for } t \in [t_0, T), \\ \hat{x}(t), & \text{for } t \in [T, T+\delta), \end{cases}$$

and have $(x^*, T+\delta) \in \mathfrak{S}$. By construction it holds that $(x, T) \neq (x^*, T+\delta)$ and $(x, T) \preccurlyeq (x^*, T+\delta)$ which is a contradiction to the maximality of (x, T) in $\mathfrak{S}$. Hence, the assumption $T < +\infty$ leads to the desired contradiction, and therefore a global solution exists. □

Remark 4.2.4. *Under additional assumptions, we can prove the existence of unique solutions in infinite-dimensional Hilbert spaces. If the gradients $\nabla f_i(x_0)$ are linearly independent at $x_0 \in \mathcal{H}$, the mapping $\phi : \mathcal{H} \to \mathcal{H}$, $x \mapsto -\operatorname{proj}_{C(x)}(0)$ is locally Lipschitz continuous in x_0. Hence, one can apply the Cauchy–Lipschitz Theorem to conclude the existence of a unique local solution to* (MSD) *[17, Proposition 3.1]. We do not present this result in detail, as this technique cannot be applied to the novel multiobjective gradient dynamical systems discussed in this thesis.*

Preparatory results

The preparatory results for the system (MSD) are structurally analogous to the ones for the system (SD) presented in the previous section. The solutions $x(\cdot)$ to (MSD) produce monotonic decay in the objective function values $f_i(x(t))$ for all $i = 1 \ldots, m$, as stated in the following proposition.

Proposition 4.2.5. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \ldots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Then for all $i = 1, \ldots, m$ and all $t > t_0$*

$$\frac{d}{dt} f_i(x(t)) \leq -\|\dot{x}(t)\|^2. \tag{4.24}$$

Proof. We apply the chain rule to obtain for $t > t_0$

$$\frac{d}{dt} f_i(x(t)) = \langle \nabla f_i(x(t)), \dot{x}(t) \rangle. \tag{4.25}$$

By the definition of (MSD), we have for all $t > t_0$

$$-\dot{x}(t) = \operatorname*{proj}_{C(x(t))}(0).$$

Since $\nabla f_i(x(t)) \in C(x(t))$ for all $i = 1, \ldots, m$, by the variational characterization of the projection (Theorem 2.1.17), we obtain

$$\langle \nabla f_i(x(t)) - (-\dot{x}(t)), -\dot{x}(t) \rangle \geq 0, \quad \text{for all} \quad i = 1, \ldots, m,$$

and hence,

$$\langle \nabla f_i(x(t)), \dot{x}(t) \rangle \leq -\|\dot{x}(t)\|^2. \tag{4.26}$$

Together, (4.25) and (4.26) give

$$\frac{d}{dt} f_i(x(t)) \leq -\|\dot{x}(t)\|^2 \leq 0.$$

□

From Proposition 4.2.5, we derive the same integral bound for the velocity as for the system (SD). For this derivation, it is enough that one objective function f_i is bounded from below.

Corollary 4.2.6. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \ldots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. If there exists $j \in \{1, \ldots, m\}$ such that f_j is bounded from below, i.e., $\inf_{x \in \mathcal{H}} f_j(x) > -\infty$, then*

$$\int_{t_0}^{+\infty} \|\dot{x}(s)\|^2 \, ds < +\infty.$$

Proof. We integrate inequality (4.24) with $i = j$ from t_0 to $t > t_0$ to get

$$f_j(x(t)) \leq f_j(x(t_0)) + \int_{t_0}^{t} \frac{d}{ds} f_j(x(s)) \, ds \leq f_j(x(t_0)) - \int_{t_0}^{t} \|\dot{x}(s)\|^2 \, ds.$$

Since f_j is bounded from below, we derive the integral estimate

$$\int_{t_0}^{+\infty} \|\dot{x}(s)\|^2 \, ds \leq f_j(x(t_0)) - \inf_{x \in \mathcal{H}} f_j(x) < +\infty,$$

which completes the proof. □

Asymptotic analysis

The asymptotic analysis presented in this section differs from the analysis for the system (SD) provided in the previous section. For the system (SD) in the context of the convex, scalar optimization problem (SOP), the asymptotic convergence of the function values, $f(x(t)) \to \inf_{x\in\mathcal{H}} f(x)$, was established. Additionally, improved convergence rates of order $f(x(t)) - \inf_{x\in\mathcal{H}} f(x) = \mathcal{O}(t^{-1})$ were shown, under the assumption $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$.

However, these results cannot be directly translated to the context of multiobjective optimization, since there is no unique optimal function value. Furthermore, it is not meaningful to work with the respective optimal function values $\inf_{x\in\mathcal{H}} f_i(x)$ for $i = 1, \dots, m$ since there is not a single point $x^* \in \mathcal{H}$ that is optimal with respect to all objectives. Instead, to measure the convergence of the function values, we use the merit function $\varphi(\cdot)$ introduced in Subsection 2.3.3. While the function $\varphi(\cdot)$ is a suitable measure for the convergence speed, it introduces analytical challenges, as $\varphi(\cdot)$ is in general not differentiable, even if all objective functions are smooth. Consequently, directly working with the derivative $\frac{d}{dt}\varphi(x(t))$ to investigate convergence is inconvenient.

To address this, we introduce multiple anchored energy functions and derive the results from there. This approach is crucial, and the observations will play a significant role in the analysis of the novel multiobjective gradient systems we discuss in the main part of this chapter.

Overall, we generalize the results obtained for the system (SD) in a satisfactory way. For convex objective functions, we prove that $\varphi(x(t)) \to 0$ as $t \to +\infty$ in the general setting, and verify the rate $\varphi(x(t)) = \mathcal{O}(t^{-1})$ as $t \to +\infty$ under Assumption $(\mathcal{A}_2)$ which generalizes the condition $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$ for the problem (SOP). Under this condition, we also prove the weak convergence of $x(\cdot)$ to a weakly Pareto optimal point of (MOP). We close this section with an example that highlights the necessity of Assumption $(\mathcal{A}_2)$.

Theorem 4.2.7. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Then, $t \mapsto \varphi(x(t))$ is monotonically decreasing. If f_i is bounded from below for all $i = 1, \dots, m$, then* $\lim_{t\to+\infty} \varphi(x(t)) \in \mathbb{R}$ *exists.*

Proof. For $i = 1, \dots, m$ and $z \in \mathcal{H}$, define the anchored energy function

$$\mathcal{W}_{i,z} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{W}_{i,z}(t) := f_i(x(t)) - f_i(z),$$

and for $z \in \mathcal{H}$, define

$$\mathcal{W}_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{W}_z(t) := \min_{i=1,\dots,m} \mathcal{W}_{i,z}(t) = \min_{i=1,\dots,m} f_i(x(t)) - f_i(z).$$

The function $\mathcal{W}_z(\cdot)$ is absolutely continuous on every compact interval, differentiable almost everywhere and satisfies for all $t \geq t_0$,

$$\varphi(x(t)) = \sup_{z\in\mathcal{H}} \mathcal{W}_z(t).$$

By Proposition 4.2.5, we have for all $z \in \mathcal{H}$, $i = 1, \ldots, m$ and $t > t_0$

$$\frac{d}{dt}\mathcal{W}_{i,z}(t) \leq -\|\dot{x}(t)\|^2.$$

Then, by Lemma 2.2.14 for all $z \in \mathcal{H}$ and almost all $t > t_0$

$$\frac{d}{dt}\mathcal{W}_{z}(t) \leq -\|\dot{x}(t)\|^2. \tag{4.27}$$

Integrating this from $t_1 > t_0$ to $t_2 > t_1$ yields for all $z \in \mathcal{H}$ and all $t_2 > t_1 > t_0$

$$\mathcal{W}_z(t_2) \leq \mathcal{W}_z(t_1) - \int_{t_1}^{t_2} \|\dot{x}(s)\|^2 \, ds.$$

Applying the supremum over $z \in \mathcal{H}$ on both sides, gives for all $t_2 > t_1 > t_0$

$$\varphi(x(t_2)) \leq \varphi(x(t_1)) - \int_{t_1}^{t_2} \|\dot{x}(s)\|^2 \, ds \leq \varphi(x(t_1)).$$

Hence $t \mapsto \varphi(x(t))$ is monotonically decreasing. If f_i is bounded from below for all $i = 1, \ldots, m$ then $\varphi(x) < +\infty$ for all $x \in \mathcal{H}$. Then, since $\varphi(x) \geq 0$ for all $x \in \mathcal{H}$, we conclude $\lim_{t\to+\infty} \varphi(x(t)) \in \mathbb{R}$ exists. □

Theorem 4.2.8. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \ldots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Then, for all $z \in \mathcal{H}$*

$$\lim_{t\to+\infty} \min_{i=1,\ldots,m} f_i(x(t)) - f_i(z) \leq 0.$$

Proof. Let $x(\cdot)$ be a trajectory solution to (MSD). Define for $i = 1, \ldots, m$ and $z \in \mathcal{H}$ the function

$$\mathcal{E}_{i,z} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}_{i,z}(t) := (t - t_0)\left(f_i(x(t)) - f_i(z)\right) + \frac{1}{2}\|x(t) - z\|^2,$$

and for $z \in \mathcal{H}$ define

$$\mathcal{E}_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}_z(t) := \min_{i=1,\ldots,m} \mathcal{E}_{i,z}(t) = (t - t_0) \min_{i=1,\ldots,m} \left(f_i(x(t)) - f_i(z)\right) + \frac{1}{2}\|x(t) - z\|^2.$$

The function $\mathcal{E}_z(\cdot)$ is absolutely continuous on every compact interval and differentiable almost everywhere. We use the chain rule and $\theta(t) \in \Delta^m$ from (4.19) to get for almost all $t > t_0$

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_{i,z}(t) =& (t - t_0)\langle \nabla f(x(t)), \dot{x}(t)\rangle + f_i(x(t)) - f_i(z) + \langle x(t) - z, \dot{x}(t)\rangle \\
\leq& -(t - t_0)\|\dot{x}(t)\|^2 + f_i(x(t)) - f_i(z) + \left\langle \sum_{i=1}^m \theta_i(t)\nabla f_i(x(t)), z - x(t) \right\rangle \\
\leq& -(t - t_0)\|\dot{x}(t)\|^2 + f_i(x(t)) - f_i(z) + \sum_{i=1}^m \theta_i(t)\left(f_i(z) - f_i(x(t))\right) \\
\leq& -(t - t_0)\|\dot{x}(t)\|^2 + f_i(x(t)) - f_i(z) - \min_{i=1,\ldots,m} \left(f_i(x(t)) - f_i(z)\right).
\end{aligned}$$

By Lemma 2.2.14, we derive for all $z \in \mathcal{H}$ and almost all $t > t_0$

$$\frac{d}{dt}\mathcal{E}_z(t) \leq \min_{i=1,\dots,m} \frac{d}{dt}\mathcal{E}_{i,z}(t) \leq -(t-t_0)\|\dot{x}(t)\|^2. \tag{4.28}$$

We integrate (4.28) from t_0 to $t > t_0$ to get for all $z \in \mathcal{H}$ and all $t > t_0$

$$(t-t_0) \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) \leq \mathcal{E}_z(t) \leq \mathcal{E}_z(t_0) - \int_{t_0}^{t} (s-t_0)\|\dot{x}(s)\|^2 \, ds \leq \mathcal{E}_z(t_0). \tag{4.29}$$

By (4.29), we follow for all $z \in \mathcal{H}$ and all $t > t_0$

$$\min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) \leq \frac{\mathcal{E}_z(t_0)}{t-t_0} = \frac{\|x(t_0) - z\|^2}{2(t-t_0)}. \tag{4.30}$$

By Proposition 4.2.5, the function $t \mapsto f_i(x(t))$ is monotonically decreasing for all $i = 1, \dots, m$. Furthermore, the limit on the left-hand side of (4.30) exists and hence

$$\lim_{t \to +\infty} \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) \leq 0.$$

$\square$

Theorem 4.2.8 asserts that there does not exist a $z \in \mathcal{H}$ that strictly dominates $x(t)$ in the limit $t \to +\infty$. In the following theorem, we show that $\varphi(x(t)) \to 0$ as $t \to +\infty$ given that the functions f_i are bounded from below for $i = 1, \dots, m$.

Theorem 4.2.9. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Assume f_i is bounded from below for all $i = 1, \dots, m$. Then,*

$$\lim_{t \to +\infty} \varphi(x(t)) = 0.$$

Proof. By Theorem 4.2.7, $\varphi^\infty := \lim_{t \to +\infty} \varphi(x(t)) \in \mathbb{R} \cup \{+\infty\}$ exists as $t \mapsto \varphi(x(t))$ is monotonically decreasing. Since the functions f_i are bounded from below, we can conclude $\varphi^\infty \leq \varphi(x(t_0)) < +\infty$. We prove the statement by contraposition.

Assume $\varphi^\infty > 0$. By the definition of $\varphi(\cdot)$ in (2.23), we conclude for all $t \geq t_0$ the existence of an element $z(t) \in \mathcal{H}$ with

$$\min_{i=1,\dots,m} f_i(x(t)) - f_i(z(t)) \geq \frac{\varphi^\infty}{2} > 0. \tag{4.31}$$

By Proposition 4.2.5, the function $t \mapsto f_i(x(t))$ is monotonically decreasing for all $i = 1, \dots, m$. Since f_i is bounded from below by assumption, the limit $f_i^\infty := \lim_{t \to +\infty} f_i(x(t))$ exists for all $i = 1, \dots, m$. For all $a, b \in \mathbb{R}^m$ it holds that $\min_{i=1,\dots,m} a_i \leq \min_{i=1,\dots,m} (a_i - b_i) + \min_{i=1,\dots,m} b_i$. We apply this inequality in (4.31) to conclude

$$\frac{\varphi^\infty}{2} < \min_{i=1,\dots,m} f_i(x(t)) - f_i(z(t)) \leq \max_{i=1,\dots,m} f_i(x(t)) - f_i^\infty + \min_{i=1,\dots,m} f_i^\infty - f_i(z(t)). \tag{4.32}$$

Since $f_i(x(t)) \to f_i^\infty$ as $t \to +\infty$ for all $i = 1, \dots, m$, there exists $T > t_0$ such that

$$\max_{i=1,\dots,m} f_i(x(T)) - f_i^\infty \leq \frac{\varphi^\infty}{4}. \tag{4.33}$$

Together, (4.32) and (4.33) give

$$\frac{\varphi^\infty}{4} < \min_{i=1,\dots,m} f_i^\infty - f_i(z(T)) = \lim_{t\to+\infty} f_i(x(t)) - f_i(z(T)),$$

which contradicts Theorem 4.2.8. This completes the contraposition. $\square$

Theorem 4.2.10. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Assume Assumption $(\mathcal{A}_2)$ holds. Then, for all $t > t_0$*

$$\varphi(x(t)) \leq \frac{R}{t - t_0},$$

where $R > 0$ is defined in Assumption $(\mathcal{A}_2)$.

Proof. Let $z \in \mathcal{H}$. We use the chain rule and $\theta(t) \in \Delta^m$ defined in (4.19) to conclude for all $t > t_0$

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}\|x(t) - z\|^2 =& \langle x(t) - z, \dot{x}(t)\rangle \leq \left\langle z - x(t), \sum_{i=1}^m \theta_i(t)\nabla f_i(x(t))\right\rangle \\ \leq& \sum_{i=1}^m \theta_i(t)\left(f_i(z) - f_i(x(t))\right) \leq - \min_{i=1,\dots,m} f_i(x(t)) - f_i(z). \end{aligned} \tag{4.34}$$

Integrating (4.34) from t_0 to $t > t_0$ gives

$$\frac{1}{2}\|x(t) - z\|^2 - \frac{1}{2}\|x(t_0) - z\|^2 \leq - \int_{t_0}^t \min_{i=1,\dots,m} f_i(x(s)) - f_i(z)\, ds. \tag{4.35}$$

By Proposition 4.2.5, the function $t \mapsto f_i(x(t))$ is monotonically decreasing for all $i = 1, \dots, m$ and we conclude from (4.35) for all $t > t_0$

$$\begin{aligned} (t - t_0) \min_{i=1,\dots,m} f_i(x(t)) - f_i(z) &= \int_{t_0}^t \min_{i=1,\dots,m} f_i(x(t)) - f_i(z)\, ds \\ \leq \int_{t_0}^t \min_{i=1,\dots,m} f_i(x(s)) - f_i(z)\, ds &\leq \frac{1}{2}\|x(t_0) - z\|^2 - \frac{1}{2}\|x(t) - z\|^2 \leq \frac{1}{2}\|x_0 - z\|^2. \end{aligned}$$

Therefore, we get for all $z \in \mathcal{H}$ and all $t \geq t_0$

$$\min_{i=1,\dots,m} f_i(x(t)) - f_i(z) \leq \frac{\|x_0 - z\|^2}{2(t - t_0)}. \tag{4.36}$$

Applying the supremum and infimum as in Lemma 2.3.15 and using Assumption $(\mathcal{A}_2)$, we follow

$$\varphi(x(t)) \leq \frac{R}{t - t_0},$$

which completes the proof. $\square$

Theorem 4.2.11. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MSD) *in the sense of Definition 4.2.1. Assume $x(\cdot)$ is bounded. Then, $x(\cdot)$ converges weakly to a weakly Pareto optimal point of* (MOP).

Proof. We prove the weak convergence of $x(\cdot)$ using Opial's Lemma (Lemma 2.1.6). Since $x(\cdot)$ is bounded and the function f_i is convex and continuous it holds that $\inf_{t \geq t_0} f_i(x(t)) > -\infty$ for all $i = 1, \dots, m$. By Proposition 4.2.5, $t \mapsto f_i(x(t))$ is monotonically decreasing and hence $f_i^\infty \coloneqq \lim_{t \to +\infty} f_i(x(t)) = \inf_{t \geq t_0} f_i(x(t)) > -\infty$ for all $i = 1, \dots, m$. Then, we define the set

$$S \coloneqq \{ z \in \mathcal{H} \,:\, f_i(z) \leq f_i^\infty, \quad \text{for all} \quad i = 1, \dots, m \}.$$

Using Opial's Lemma, we show that $x(\cdot)$ converges weakly to an element $x^\infty \in S$ and prove the optimality of x^∞ in a subsequent step. To apply Opial's Lemma, we have to show that $S \neq \emptyset$, all weak sequential cluster points of $x(\cdot)$ belong to S and $\lim_{t \to +\infty} \|x(t) - z\|$ exists for all $z \in S$.

Since $x(\cdot)$ is bounded, it possesses at least one weak sequential cluster point $x^\infty \in \mathcal{H}$, i.e., there exists a sequence $(t_k)_{k \geq 0}$ with $t_k \to +\infty$ and $x(t_k) \rightharpoonup x^\infty$ as $k \to +\infty$. Since the functions f_i are convex and continuous, they are weakly lower semicontinuous for all $i = 1, \dots, m$ and we follow

$$f_i(x^\infty) \leq \liminf_{k \to +\infty} f_i(x(t_k)) = \lim_{t \to +\infty} f_i(x(t)) = f_i^\infty.$$

Therefore, $x^\infty \in S$ and hence $S \neq \emptyset$. By the same argument all weak sequential cluster points of $x(\cdot)$ belong to S.

For $z \in S$, we define

$$h_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \frac{1}{2}\|x(t) - z\|^2. \tag{4.37}$$

Using the chain rule to differentiate $h_z(\cdot)$ combined with $\theta(t) \in \Delta^m$ from (4.19) gives

$$\frac{d}{dt} h_z(t) = \langle x(t) - z, \dot{x}(t) \rangle = \left\langle z - x(t), \sum_{i=1}^m \theta_i(t) \nabla f_i(x(t)) \right\rangle = \sum_{i=1}^m \theta_i(t) \langle z - x(t), \nabla f_i(x(t)) \rangle.$$

Since the functions f_i are convex for all $i = 1, \dots, m$, we bound this using Proposition 2.1.24, to obtain

$$\leq \sum_{i=1}^m \theta_i(t) \left(f_i(z) - f_i(x(t)) \right) \leq - \min_{i=1,\dots,m} f_i(x(t)) - f_i(z) \leq - \min_{i=1,\dots,m} f_i^\infty - f_i(z) \leq 0,$$

where the last inequality follows by $z \in S$. Hence, the function $h_z(\cdot)$ is monotonically decreasing for all $z \in S$. Therefore, all conditions of Opial's Lemma are satisfied and we conclude

$$x(t) \rightharpoonup x^\infty \in S \quad \text{as} \quad t \to +\infty.$$

By Theorem 2.3.14, the function $\varphi(\cdot)$ is weakly lower semicontinuous and we conclude

$$\varphi(x^\infty) \leq \liminf_{k \to +\infty} \varphi(x(t_k)) = \lim_{t \to +\infty} \varphi(x(t)) = 0, \tag{4.38}$$

where the last equality follows by Theorem 4.2.9. Finally, from (4.38) we obtain that x^∞ is a weakly Pareto optimal point of (MOP) using Theorem 2.3.13. □

A discussion of Assumption $(\mathcal{A}_2)$ in the context of (MSD)

Initially, Assumption $(\mathcal{A}_2)$ might appear unnatural. In the following, we examine the assumption in detail by means of an illustrative example. We restate Assumption $(\mathcal{A}_2)$ as follows:

$(\mathcal{A}_2)$ For all $x_0 \in \mathcal{H}$ and for all $x \in \mathcal{L}(F, F(x_0))$ it holds that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ and further

$$R := \sup_{F^* \in F(\mathcal{LP}_w(F,F(x_0)))} \inf_{z \in F^{-1}(\{F^*\})} \frac{1}{2}\|z - x_0\|^2 < +\infty. \tag{4.39}$$

Assumption $(\mathcal{A}_2)$ not only asks for the existence of weakly Pareto optimal points, but also imposes a certain uniform boundedness condition. For the scalar optimization problem (SOP), a common assumption is $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$. This assumption is equivalent to the boundedness of solutions $x(\cdot)$ to (SD), as shown in Corollary 4.1.7. In the context of multiobjective optimization, we want to find weakly Pareto optimal points of (MOP). Therefore, it seems natural to extend the assumption $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$ from scalar optimization to the assumption $\mathcal{P}_w \neq \emptyset$ for multiobjective optimization. However, in the following example, we show that the assumption $\mathcal{P}_w \neq \emptyset$ is not sufficient to obtain the results proven in Theorems 4.2.10 and 4.2.11. We construct objective functions for which the problem (MOP) satisfies $\mathcal{P}_w \neq \emptyset$ but not Assumption $(\mathcal{A}_2)$. For this problem there exist unbounded solutions $x(\cdot)$ to (MSD) that do not converge to weakly Pareto optimal points of (MOP), and for which the convergence rate of the function values $\varphi(x(t)) = \mathcal{O}(t^{-1})$ as $t \to +\infty$ is not satisfied.

Define the function

$$g : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \begin{cases} \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2, & \text{if } |x_1| \leq 1, \quad x_2 + 1 \leq \sqrt{1 - x_1^2}, \\ |x_1| + \frac{1}{2}x_2^2 - \frac{1}{2}, & \text{if } |x_1| > 1, \quad x_2 + 1 \leq 0, \\ \sqrt{x_1^2 + (x_2+1)^2} - (x_2 + 1), & \text{else}, \end{cases} \tag{4.40}$$

which is convex and continuously differentiable with Lipschitz continuous gradients. A discussion of the properties of the function g can be found in Example 4.6.9 in Section 4.6 on the system (MTRIGS). We refer the reader to Section 4.6 for further details, as the function g was originally introduced in [49] where the system (MTRIGS) was introduced.

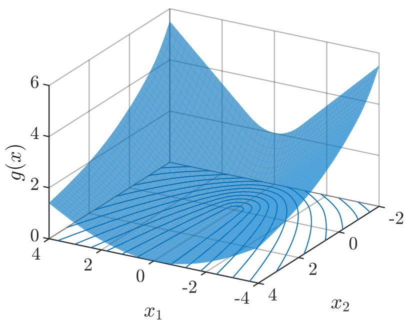


Figure 4.3: Function g defined in (4.40).

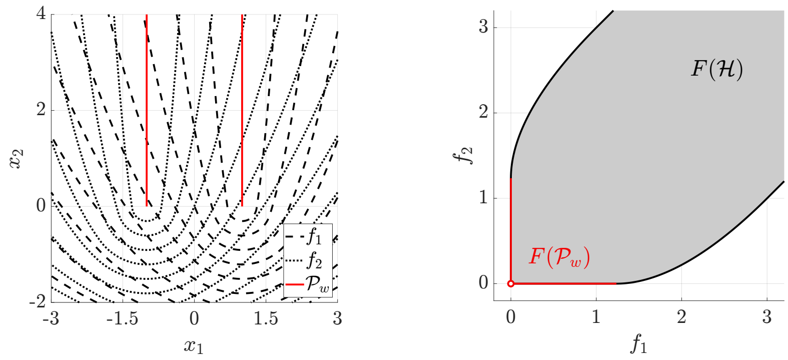
The function g is a so-called perspective function [32]. It satisfies $\arg\min_{x\in\mathcal{H}} g(x) = \{x \in \mathbb{R}^2 : x_1 = 0,\ x_2 \geq 0\}$ and for all $x_1 \in \mathbb{R}$ it holds that $g(x_1, x_2) \to 0$ as $x_2 \to +\infty$. Using the function g we define

$$f_1 : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto g\left(\begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix}\right) \quad \text{and} \quad f_2 : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto g\left(\begin{bmatrix} x_1 + 1 \\ x_2 \end{bmatrix}\right). \tag{4.41}$$

For the functions f_1 and f_2 defined in (4.41), we consider the multiobjective optimization problem

$$\text{(MOP-Ex)} \qquad \min_{x\in\mathbb{R}^2} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}.$$

The weak Pareto set of (MOP-Ex) is $\mathcal{P}_w := \left\{x \in \mathbb{R}^2 : x_1 \in \{-1, 1\},\ x_2 \geq 0\right\}$, which is shown in Subfigure 4.4a. Hence, $\mathcal{P}_w \neq \emptyset$ and furthermore, $\mathcal{LP}_w(F, F(x)) = \mathcal{P}_w \cap \mathcal{L}(F, F(x)) \neq \emptyset$ for all $x \in \mathbb{R}^2$. Nevertheless, Assumption $(\mathcal{A}_2)$ is not satisfied as there is no uniform bound of the weak Pareto set with respect to different optimal objective function values in the Pareto front.



(a) (b)

Figure 4.4: Subfigure 4.4a shows contour plots of objective functions f_1 and f_2 defined in (4.41) and the weak Pareto set corresponding to (MOP-Ex). The attainable set $F(\mathcal{H})$ and the weak Pareto front $F(\mathcal{P}_w)$ of problem (MOP-Ex) are illustrated in Subfigure 4.4b.

For the multiobjective optimization problem (MOP-Ex), there exist unbounded solutions $x(\cdot)$ to (MSD). We do not compute a solution in full detail, but sketch a way to show that $x(\cdot)$ does not converge. Consider $x(\cdot)$ a solution to (MSD) with initial data $t_0 = 1$ and $x(t_0) = (0, 0)^\top$. From the symmetry of the objective functions it can be deduced that $x_1(t) = 0$ and $x_2(t) \geq 0$ for all $t \geq t_0$. Then, by computing $\operatorname{proj}_{C(x(t))}(0)$, we see that the second component of the solution satisfies

$$\dot{x}_2(t) = 1 - \frac{x_2(t) + 1}{\sqrt{1 + (x_2(t) + 1)^2}} \geq \frac{1}{4(x_2(t) + 1)^2}.$$

From this, it follows that $x_2(t) \to +\infty$ as $t \to +\infty$. Hence $x(\cdot)$ is not bounded and not converging to an element in $\mathcal{P}_w$. Additionally, $\operatorname{dist}(x(t), \mathcal{P}_w) = 1$ for all $t \geq t_0$. Nevertheless, $\varphi(x(t)) \to 0$ as $t \to +\infty$ but at a slower asymptotic rate than $\mathcal{O}\left(\frac{1}{t}\right)$.

4.2.2 Adaption of (MSD) for constrained MOPs

So far, we have only considered unconstrained optimization problems, where the goal is to find weakly Pareto optimal points in the entire space $\mathcal{H}$. In [18], the system (MSD) is adapted to address constrained multiobjective optimization problems of the form

$$\text{(CMOP)} \qquad \min_{x \in K} \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

where $K \subset \mathcal{H}$ is a closed, convex and nonempty set. Under the assumptions that $f_i : \mathcal{H} \to \mathbb{R}$ is quasi-convex and continuously differentiable with Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$, the following *constrained multiobjective steepest descent system* is proposed in [18]:

$$\text{(CMSD)} \qquad \dot{x}(t) + \operatorname*{proj}_{N_K(x(t)) + C(x(t))}(0) = 0, \quad \text{for} \quad t > t_0,$$

with initial data $t_0 > 0$ and $x(t_0) = x_0 \in K$. Here

$$N_K : K \rightrightarrows \mathcal{H}, \quad x \mapsto N_K(x) := \{z \in \mathcal{H} : \langle z, y - x \rangle \leq 0, \quad \text{for all} \quad y \in K\},$$

is the normal cone mapping which models the contact forces of the constrained in the system (CMSD), while

$$C : \mathcal{H} \rightrightarrows \mathcal{H}, \quad x \mapsto C(x) := \operatorname{conv}\left(\{\nabla f_i(x) : \quad \text{for all} \quad i = 1, \dots, m\}\right),$$

is the convex hull of the gradients, which describes the driving forces. In the interior of K, i.e., for $x(t) \in \operatorname{int}(K)$, it holds that $N_K(x(t)) = \{0\}$, and therefore the system (CMSD) is equivalent to the multiobjective steepest descent system (MSD). For the system (CMSD) a solution is given by the following definition.

Definition 4.2.12. *A function $x : [t_0, +\infty) \to K \subset \mathcal{H}$ is called a solution to* (CMSD) *if it satisfies the following properties:*

i) $x(\cdot)$ is continuous and absolutely continuous on each compact interval $[t_0, T]$ for $t_0 < T < +\infty$;

ii) There exist $v, w : [t_0, +\infty) \to \mathcal{H}$ such that:

a) $v, w \in L^2([t_0, T], \mathcal{H})$ for all $t_0 < T < +\infty$;

b) $v(t) \in N_K(x(t))$, $w(t) \in C(x(t))$ for almost all $t > t_0$;

c) $v(t) + w(t) = \operatorname{proj}_{N_K(x(t))+C(x(t))}(0)$ for almost all $t > t_0$;

d) $\dot{x}(t) + v(t) + w(t) = 0$ for almost all $t > t_0$.

The normal cone mapping $N_K(\cdot)$ introduces discontinuities in equation (CMSD). As a result, solutions $x(\cdot)$ to (CMSD) are not continuously differentiable but less regular, and the differential equation is satisfied only almost everywhere. The following existence result can be established for (CMSD).

Theorem 4.2.13. *Assume $\mathcal{H}$ is finite dimensional. Let $K \subset \mathcal{H}$ be closed, convex and nonempty. Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$. Then, for all $t_0 > 0$ and all $x_0 \in K$ there exists a solution to* (CMSD) *in the sense of Definition 4.2.12.*

Remark 4.2.14. *The proof of Theorem 4.2.13 given in [18] is more general and applies to equations of the form*

$$\dot{x}(t) + \operatorname*{proj}_{\partial\Phi(x(t)) - B(x(t))}(0) = 0, \tag{4.42}$$

where $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous and $B : \mathcal{H} \rightrightarrows \mathcal{H}$ is a continuous set-valued operator which satisfies a certain growth property. The existence proof relies on a Yosida approximation of the maximal monotone operator $\partial\Phi$ and Peano's Theorem to conclude existence of solutions to a regularized version of (4.42)*. Then a solution to the original equation is derived by letting the regularization parameter in the Yosida approximation tend to zero and conclude the existence of a limit that satisfies* (4.42) *by compactness arguments and the closedness of the involved operators.*

For the system (CMSD), the following asymptotic results can be derived.

Theorem 4.2.15. *Let $f_i : \mathcal{H} \to \mathbb{R}$ be continuously differentiable with Lipschitz continuous gradient ∇f_i and assume f_i is bounded from below on K for $i = 1, \dots, m$. Let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (CMSD) *in the sense of Definition 4.2.12. Then:*

i) For all $i = 1, \dots, m$ and for almost all $t > t_0$

$$\frac{d}{dt} f_i(x(t)) \leq -\|\dot{x}(t)\|^2;$$

ii) It holds that

$$\int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty;$$

iii) If $x(\cdot)$ is bounded, then $x(\cdot)$ converges weakly in $\mathcal{H}$, i.e., $x(t) \rightharpoonup x^\infty \in \mathcal{H}$ as $t \to +\infty$.

4.2.3 Adaption of (MSD) for constrained and nonsmooth MOPs

A further adaption of (MSD) is presented in [17], where an extension of (CMSD) to nonsmooth multiobjective optimization problems is proposed. The paper investigates constrained multiobjective optimization problems

(CMOP')
$$\min_{x \in K} \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

where $K \subset \mathcal{H}$ is a closed, convex and nonempty set and the objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and Lipschitz continuous on bounded sets for $i = 1, \dots, m$. Under these assumptions, the following *constrained multiobjective steepest descent system* is introduced:

$$\text{(CMSD')} \qquad \dot{x}(t) + \operatorname*{proj}_{N_K(x(t))+\tilde{C}(x(t))}(0) = 0, \quad \text{for} \quad t > t_0,$$

with initial data $t_0 > 0$ and $x(t_0) \in K$. Similar to (CMSD), the mapping

$$N_K : K \rightrightarrows \mathcal{H}, \quad x \mapsto N_K(x) := \{z \in \mathcal{H} \,:\, \langle z, y - x\rangle \leq 0, \quad \text{for all} \quad y \in K\},$$

is the normal cone mapping, which models the contact forces of the trajectory $x(\cdot)$ and the boundary of K in (CMSD'). The set-valued operator

$$\tilde{C} : \mathcal{H} \rightrightarrows \mathcal{H}, \quad x \mapsto \tilde{C}(x) := \overline{\operatorname{conv}}\left(\bigcup_{i=1}^{m} \partial f_i(x)\right),$$

maps $x \in \mathcal{H}$ to the closure of the convex hull of the union of the convex subdifferentials $\partial f_i(x)$ of the respective objective functions. The mapping $\tilde{C}(\cdot)$ extends the multiobjective steepest descent direction to nonsmooth convex objective functions. A solution to (CMSD') is given by the following definition.

Definition 4.2.16. *A function $x : [t_0, +\infty) \to K \subset \mathcal{H}$ is called a solution to* (CMSD') *if it satisfies the following properties:*

i) $x(\cdot)$ is continuous and absolutely continuous on each compact interval $[t_0, T]$ for $t_0 < T < +\infty$;

ii) There exist $v, \xi_i : [t_0, +\infty) \to \mathcal{H}$ and $\theta_i : [t_0, +\infty) \to [0, 1]$ for $i = 1, \dots, m$ such that:

- *a) $\theta_i \in L^\infty([0, +\infty), \mathbb{R})$ and $\theta(t) \in \Delta^m$ for almost all $t > t_0$;*
- *b) $v \in L^\infty([0, T], \mathcal{H})$ and $\xi_i \in L^\infty([0, T], \mathcal{H})$ for all $T > t_0$ and all $i = 1, \dots, m$;*
- *c) $\eta(t) \in N_K(x(t))$ and $\xi_i(t) \in \partial f_i(x(t))$ for all $i = 1, \dots, m$ and almost all $t > t_0$;*
- *d) $\dot{x}(t) + \eta(t) + \sum_{i=1}^m \theta_i(t)\xi_i(t) = 0$ for almost all $t > t_0$;*
- *e) $\dot{x}(t) + \operatorname{proj}_{N_K(x(t))+\tilde{C}(x(t))} = 0$ for almost all $t > t_0$.*

Similar to (CMSD), solutions $x(\cdot)$ to (CMSD') are not continuously differentiable, but merely absolutely continuous on compact intervals, as a consequence of the discontinuities in (CMSD') introduced by the normal cone mapping $N_K(\cdot)$. Therefore, the equation (CMSD') is only satisfied almost everywhere in $[t_0, +\infty)$.

Theorem 4.2.17. *Assume $\mathcal{H}$ is finite dimensional. Let $K \subset \mathcal{H}$ be closed, convex and nonempty and let $f_i : \mathcal{H} \to \mathbb{R}$ be convex, Lipschitz continuous on bounded sets and bounded from below for all $i = 1, \dots, m$. Then, for all $t_0 > 0$ and all $x_0 \in K$ there exists a solution to* (CMSD') *in the sense of Definition 4.2.16.*

Remark 4.2.18. *Similar to the proof of existence of solutions to* (CMSD)*, Theorem 4.2.17 relies on a Yosida approximation of the maximal monotone operator* $N_K(\cdot)$ *and a Moreau–Yosida approximation of the convex functions* f_i *for* $i = 1, \dots, m$*. A solution to* (CMSD') *can be obtained by letting the regularization parameters converge to zero and showing that there exists a limit which is a solution in the sense of Definition 4.2.16, using the closedness of the associated operators and compactness arguments. As the analysis in this thesis does not make use of this technique, we do not present these ideas in detail. We summarize the asymptotic properties of* (CMSD') *which are analogous to Theorem 4.2.15 for the system* (CMSD).

Theorem 4.2.19. *Let* $K \subset \mathcal{H}$ *be closed, convex and nonempty and let* $f_i : \mathcal{H} \to \mathbb{R}$ *be convex, Lipschitz continuous on bounded sets and bounded from below for all* $i = 1, \dots, m$*. Let* $x : [t_0, +\infty) \to \mathcal{H}$ *be a solution to* (CMSD') *in the sense of Definition 4.2.16. Then:*

i) For all $i = 1, \dots, m$ *and for almost all* $t > t_0$

$$\frac{d}{dt} f_i(x(t)) \le -\|\dot{x}(t)\|^2;$$

ii) It holds that

$$\int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty;$$

iii) If $x(\cdot)$ *is bounded, then* $x(\cdot)$ *converges weakly to a weakly Pareto optimal point of* (CMOP')*, i.e.,* $x(t) \rightharpoonup x^\infty \in \mathcal{P}_w$ *as* $t \to +\infty$.

4.2.4 The inertial multiobjective gradient system (IMOG)

The final system we introduce in this literature review is the *inertial multiobjective gradient system*

$$\text{(IMOG)} \qquad \mu\ddot{x}(t) + \gamma\dot{x}(t) + \operatorname*{proj}_{C(x(t))}(0) = 0, \quad \text{for} \quad t > t_0,$$

with positive constants $\mu, \gamma > 0$, and initial data $t_0 > 0$ and $x(t_0) = x_0$, $\dot{x}(t_0) = v_0 \in \mathcal{H}$. Here, $C(x) \coloneqq \operatorname{conv}\left(\{\nabla f_i(x) \,:\, i = 1, \dots, m\}\right)$ is the convex hull of the gradients. The system (IMOG) was introduced in [16] to incorporate inertial effects into multiobjective gradient dynamics. It combines the *multiobjective steepest descent dynamical system*

$$\text{(MSD)} \qquad \dot{x}(t) + \operatorname*{proj}_{C(x(t))}(0) = 0,$$

which is discussed in detail in Subsection 4.2.1, with the so-called *heavy ball with friction system*

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

with $\gamma > 0$. The system (HBF) was studied in [196] in the context of accelerating iterative schemes for scalar optimization. Its name originates from [19], where it was derived as a model for a heavy ball rolling down the graph of the function f and analyzed in the context of global optimization. The system (IMOG) serves as an important starting point for the analysis developed in this chapter. We provide further references and a more detailed discussion on the system (HBF) in Section 4.4, where we introduce a related system to (IMOG) that addresses some limitations of the original inertial multiobjective gradient system. In this section, we briefly summarize the theoretical results related to (IMOG).

A solution to (IMOG) is formally defined in the following way.

Definition 4.2.20. *A function $x : [t_0, +\infty) \to \mathcal{H}$ is called a solution to* (IMOG) *if it satisfies the following properties:*

i) $x \in C^2([t_0, +\infty), \mathcal{H})$, *i.e.,* $x(\cdot)$ *is twice continuously differentiable;*

ii) $x(t_0) = x_0$ *and* $\dot{x}(t_0) = v_0$*;*

iii) $\mu\ddot{x}(t) + \gamma\dot{x}(t) + \mathrm{proj}_{C(x(t))}(0) = 0$ *for all* $t > t_0$.

By Peano's Theorem and the regularity of the objective functions, the existence of global solutions to (IMOG) follows, as the following theorem states.

Theorem 4.2.21. *Assume that $\mathcal{H}$ is finite dimensional. Let f_i be continuously differentiable with Lipschitz continuous gradient ∇f_i for all $i = 1, \ldots, m$. Then, for all $t_0 > 0$ and $x_0, v_0 \in \mathcal{H}$ there exists a solution to* (IMOG) *in the sense of Definition 4.2.20.*

The proof of Theorem 4.2.21 is similar to the proof of existence of solutions to (MSD) (Theorem 4.2.3). It also relies on Peano's Theorem and therefore applies only in finite-dimensional spaces, and does not guarantee the uniqueness of solutions. The following proposition in [16] further addresses the question of uniqueness.

Proposition 4.2.22. *Let f_i be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \ldots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (IMOG) *in the sense of Definition 4.2.20. If the gradients $\{\nabla f_i(x(t)) : i = 1, \ldots, m\}$ are linearly independent for all $t > t_0$, then $x(\cdot)$ is the unique solution to* (IMOG).

The multiobjective steepest descent direction $x \mapsto -\mathrm{proj}_{C(x)}(0)$ is locally Lipschitz continuous in $x \in \mathcal{H}$, if the gradients $\{\nabla f_i(x) : i = 1, \ldots, m\}$ are linearly independent in x. Then, the uniqueness of solutions in Proposition 4.2.22 follows by the Cauchy–Lipschitz Theorem (Theorem 2.2.2).

The energy functions introduced in the following proposition are an important component in the analysis of (IMOG).

Proposition 4.2.23. *Let f_i be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \ldots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (IMOG) *in the sense of Definition 4.2.20. Define for all $i = 1, \ldots, m$,*

$$\mathcal{E}_i : [t_0, +\infty) \to \mathcal{H}, \quad t \mapsto \mathcal{E}_i(t) := f_i(x(t)) + \frac{\mu}{\gamma}\frac{d}{dt} f_i(x(t)) + \mu\|\dot{x}(t)\|^2.$$

For all $i = 1, \dots, m$, the function $\mathcal{E}_i(\cdot)$ is differentiable almost everywhere. Then, for all $i = 1, \dots, m$ and almost all $t > t_0$,

$$\frac{d}{dt}\mathcal{E}_i(t) \le -\frac{\mu^2}{\gamma}\|\ddot{x}(t)\|^2 - \frac{1}{\gamma}\left(\gamma^2 - \mu L\right)\|\dot{x}(t)\|^2.$$

Proposition 4.2.23 shows monotonic decay of the energy function $\mathcal{E}_i(\cdot)$ given that $\gamma^2 > \mu L$. This condition will be necessary to derive further asymptotic results on (IMOG).

Proposition 4.2.24. *Let f_i be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (IMOG) *in the sense of Definition 4.2.20 and assume $\gamma^2 > \mu L$. Then:*

i) For all $i = 1, \dots, m$, it holds that $\lim_{t\to+\infty} \mathcal{E}_i(t) = \mathcal{E}_i^\infty \in \mathbb{R}$ exists;

ii) $\dot{x} \in L^2([t_0, +\infty), \mathcal{H}) \cap L^\infty([t_0, +\infty), \mathcal{H})$ and $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$;

iii) $\ddot{x}(t) \in L^\infty([t_0, +\infty), \mathcal{H}) \cap L^2([t_0, +\infty), \mathcal{H})$ and $\operatorname{liminfess}_{t\to+\infty}\|\ddot{x}(t)\| = 0$;

iv) For all $i = 1, \dots, m$, it holds that $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{E}_i^\infty$;

v) There exists a measurable function $\theta : [t_0, +\infty) \to \Delta^m$, $t \mapsto \theta(t)$ such that for all $t \in [t_0, +\infty)$

$$\mu\ddot{x}(t) + \gamma\dot{x}(t) + \sum_{i=1}^{m} \theta_i(t)\nabla f_i(x(t)) = 0.$$

The asymptotic results stated in Proposition 4.2.24 are followed by a theorem showing that solutions $x(\cdot)$ to (IMOG) converge to Pareto optimal points of (MOP).

Theorem 4.2.25. *Let f_i be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f_i for $i = 1, \dots, m$ and assume $\gamma^2 > \mu L$. Let $x : [t_0, +\infty) \to \mathcal{H}$ be a bounded solution to* (IMOG) *in the sense of Definition 4.2.20. Then, $x(\cdot)$ converges weakly to a weakly Pareto optimal point of* (MOP)*, i.e., $x(t) \rightharpoonup x^\infty \in \mathcal{P}_w$ as $t \to +\infty$.*

4.3 Existence results for a generalized differential equation

In the course of this chapter, we introduce three novel gradient dynamical systems, which are connected to the multiobjective optimization problem (MOP), namely (IMOG'), (MAVD) and (MTRIGS). To unify the discussion on the existence of solutions, in this section, we propose the generalized differential equation (D). We prove existence of solutions to this equation for finite dimensional spaces, i.e., $\dim(\mathcal{H}) < +\infty$. A uniqueness result is not included in this chapter, but moved to the following sections, where special instances of (D) get discussed. The proof on existence of solutions makes use of existence results for a related differential inclusion and uses techniques developed in our papers [49, 216, 217].

4.3.1 The generalized differential equation (D)

Let $d_i : (0, +\infty) \times \mathcal{H} \to \mathcal{H}$, $(t, u) \mapsto d_i(t, u)$ be continuous for $i = 1, \ldots, m$ and define the set-valued map $D : (0, +\infty) \times \mathcal{H} \rightrightarrows \mathcal{H}$, $(t, u) \mapsto D(t, u) := \operatorname{conv}(\{d_i(t, u) : i = 1, \ldots, m\})$ and let $\gamma : (0, +\infty) \to [0, +\infty)$, $t \mapsto \gamma(t)$ be a monotonically decreasing and continuous function. We define the generalized differential equation

$$\text{(D)} \qquad \gamma(t)\dot{x}(t) + \operatorname*{proj}_{D(t,x(t))+\ddot{x}(t)}(0) = 0, \quad \text{for} \quad t > t_0,$$

with initial data $t_0 > 0$, $x(t_0) = x_0 \in \mathcal{H}$ and $\dot{x}(t_0) = v_0 \in \mathcal{H}$.

In this section, we prove under additional conditions on the functions $d_i(\cdot, \cdot)$ the existence of global solutions to (D). The implicit structure of (D) does not allow for application of Peano's Theorem (Theorem 2.2.1) or the Cauchy–Lipschitz Theorem (Theorem 2.2.2) to prove existence of solutions. Instead, we show that the system (D) possesses a solution if there exists a solution to a related differential inclusion. This way, we do not have to treat the implicit equation (D) directly, but can employ existence results for differential inclusions. By this approach, we do not obtain solutions $x(\cdot)$ which are twice continuously differentiable but less regular. We give a precise definition of a solution to (D) after the discussion of the announced differential inclusion.

4.3.2 The associated differential inclusion (DI-D)

In the following definition, we introduce the set valued map $H : (0, +\infty) \times \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$, which specifies the differential inclusion that is central to this subsection.

Definition 4.3.1. *Let $d_i : (0, +\infty) \times \mathcal{H} \to \mathcal{H}$, $(t, u) \mapsto d_i(t, u)$ be continuous for $i = 1, \ldots, m$, define the set-valued map $D : (0, +\infty) \times \mathcal{H} \rightrightarrows \mathcal{H}$, $(t, u) \mapsto D(t, u) := \operatorname{conv}(\{d_i(t, u) : i = 1, \ldots, m\})$ and let $\gamma : (0, +\infty) \to [0, +\infty)$, $t \mapsto \gamma(t)$ be a monotonically decreasing and continuous function. Define the set-valued map*

$$\begin{aligned} &H : (0, +\infty) \times \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}, \\ &(t, u, v) \mapsto H(t, u, v) := \{v\} \times \left(-\gamma(t)v - \operatorname*{arg\,min}_{g \in D(t,u)} \langle g, -v \rangle \right). \end{aligned} \tag{4.43}$$

The main object of interest in this subsection is the differential inclusion

$$\text{(DI-D)} \qquad \left| \begin{array}{ll} (\dot{u}(t), \dot{v}(t)) \in H(t, u(t), v(t)), \quad \text{for} \quad t > t_0, \\ \\ (u(t_0), v(t_0)) = (u_0, v_0), \end{array} \right.$$

given initial data $t_0 > 0$ and $u_0, v_0 \in \mathcal{H}$, where the set-valued map $H : (0, +\infty) \times \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}$ is given by Definition 4.3.1.

We start by discussing the properties of the set-valued mapping $H(\cdot, \cdot, \cdot)$ given by Definition 4.3.1. To this end, we introduce the following auxiliary lemma. Lemma 4.3.2 states that the set-valued map $(t, u, v) \mapsto \arg\min_{g \in D(t,u)} \langle g, -v \rangle$ is upper semicontinuous (see Definition 2.2.3).

Lemma 4.3.2. *Let $d_i : (0, +\infty) \times \mathcal{H} \to \mathcal{H}$ for $i = 1, \ldots, m$ and $D : (0, +\infty) \times \mathcal{H} \rightrightarrows \mathcal{H}$ be given by Definition 4.3.1. Let $(\overline{t}, \overline{u}, \overline{v}) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ be fixed. Then, for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $(t, u, v) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ with $\|(t, u, v) - (\overline{t}, \overline{u}, \overline{v})\|_{\mathbb{R} \times \mathcal{H} \times \mathcal{H}} < \delta$ and for all $g \in \arg\min_{g \in D(t,u)} \langle g, -v \rangle$ there exists $\overline{g} \in \arg\min_{\overline{g} \in D(\overline{t}, \overline{u})} \langle \overline{g}, -\overline{v} \rangle$ with $\|g - \overline{g}\| < \varepsilon$.*

Proof.
Let $(\overline{t}, \overline{u}, \overline{v}) \in [t_0, +\infty) \times \mathcal{H} \times \mathcal{H}$ be fixed. We can describe the set $\arg\min_{\overline{g} \in D(\overline{t}, \overline{u})} \langle \overline{g}, -\overline{v} \rangle$ using the vertices of $D(\overline{t}, \overline{u})$ since the set $D(\overline{t}, \overline{u})$ is a convex polyhedron and the objective function $\overline{g} \mapsto \langle \overline{g}, -\overline{v} \rangle$ is linear. A minimum of $\min_{\overline{g} \in D(\overline{t}, \overline{u})} \langle \overline{g}, -\overline{v} \rangle$ is attained at a vertex of $D(\overline{t}, \overline{u})$ and since this set is compact it exists at least one $i \in \{1, \ldots, m\}$ such that $\langle d_i(\overline{t}, \overline{u}), -\overline{v} \rangle = \min_{\overline{g} \in D(\overline{t}, \overline{u})} \langle \overline{g}, -\overline{v} \rangle$. The same can be done for any $(t, u, v) \in [t_0, +\infty) \times \mathcal{H} \times \mathcal{H}$. Define the index sets of optimal and non-optimal vertices

$$\begin{aligned} \mathcal{A}(t, u, v) &:= \left\{ i \in \{1, \ldots, m\} \,:\, \langle d_i(t, u), -v \rangle = \min_{g \in D(t,u)} \langle g, -v \rangle \right\}, \quad \text{and} \\ \mathcal{I}(t, u, v) &:= \{1, \ldots, m\} \setminus \mathcal{A}(t, u, v), \end{aligned}$$

and fix the notation $\overline{\mathcal{A}} := \mathcal{A}(\overline{t}, \overline{u}, \overline{v})$ and $\overline{\mathcal{I}} := \mathcal{I}(\overline{t}, \overline{u}, \overline{v})$. By the arguments mentioned above $\mathcal{A}(t, u, v) \neq \emptyset$ holds for all $(t, u, v) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$. (Note that the following argument also works in the case $\overline{\mathcal{I}} = \emptyset$.) There exists $M \in \mathbb{R}$ such that for all $i \in \overline{\mathcal{A}}$ and $j \in \overline{\mathcal{I}}$ it holds that

$$\langle d_i(\overline{t}, \overline{u}), -\overline{v} \rangle < M < \langle d_j(\overline{t}, \overline{u}), -\overline{v} \rangle.$$

Then by the continuity of $(t, u, v) \mapsto \langle d_i(t, u), v \rangle$ we can choose $\delta > 0$ such that for all $(t, u, v) \in [t_0, +\infty) \times \mathcal{H} \times \mathcal{H}$ with $\|(t, u, v) - (\overline{t}, \overline{u}, \overline{v})\|_{\mathbb{R} \times \mathcal{H} \times \mathcal{H}} < \delta$ and all $i \in \overline{\mathcal{A}}$ and $j \in \overline{\mathcal{I}}$

$$\langle d_i(t, u), -v \rangle < M < \langle d_j(t, u), -v \rangle.$$

Hence, for all $(t, u, v) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ with $\|(t, u, v) - (\overline{t}, \overline{u}, \overline{v})\|_{\mathbb{R} \times \mathcal{H} \times \mathcal{H}} < \delta$ it holds that $\mathcal{A}(t, u, v) \subset \overline{\mathcal{A}}$. Now, the remainder of the proof follows from the continuity of the functions $d_i(\cdot, \cdot)$ for $i = 1, \ldots, m$. Let $g \in \arg\min_{g \in D(t,u)} \langle g, -v \rangle$ be arbitrary. Write $g = \sum_{i \in \mathcal{A}(t,u,v)} \theta_i d_i(t, u)$ as a convex combination of the optimal vertices of $D(t, u)$ with $\theta \in \Delta^m$. From $\mathcal{A}(t, u, v) \subset \overline{\mathcal{A}}$, it follows that $\overline{g} = \sum_{i\in\mathcal{A}(t,u,v)} \theta_i d_i(\overline{t},\overline{u})$ is a solution to $\min_{\overline{g}\in D(\overline{t},\overline{u})}\langle \overline{g}, -\overline{v}\rangle$. Since the functions $d_i(\cdot,\cdot)$ are continuous for all $i = 1,\dots,m$, we can choose $\delta > 0$ such that

$$\|g - \overline{g}\| = \left\| \sum_{i\in\mathcal{A}(t,u,v)} \theta_i \left(d_i(t,u) - d_i(\overline{t},\overline{u})\right)\right\| \leq \max_{i=1,\dots,m} \|d_i(t,u) - d_i(\overline{t},\overline{u})\| < \varepsilon.$$

□

Proposition 4.3.3. *Let $d_i : (0,+\infty)\times\mathcal{H}\to\mathcal{H}$ for $i = 1,\dots,m$, $D : (0,+\infty)\times\mathcal{H}\rightrightarrows\mathcal{H}$ and $\gamma : (0,+\infty)\to[0,+\infty)$ be given by Definition 4.3.1. Then, the set-valued map $H(\cdot,\cdot,\cdot)$ defined in* (4.43) *has the following properties:*

i) For all $(t,u,v)\in(0,+\infty)\times\mathcal{H}\times\mathcal{H}$, the set $H(t,u,v)\subset\mathcal{H}\times\mathcal{H}$ is convex and compact;

ii) $H(\cdot,\cdot,\cdot)$ is upper semicontinuous;

iii) The map

$$\psi : (0,+\infty)\times\mathcal{H}\times\mathcal{H}\to\mathcal{H}\times\mathcal{H}, \quad (t,u,v)\mapsto\psi(t,u,v) := \operatorname*{proj}_{H(t,u,v)}(0)$$

is locally compact if and only if $\dim(\mathcal{H}) < +\infty$*;*

iv) Let $t_0 > 0$. Assume the functions $d_i(\cdot,\cdot)$ are uniformly L-Lipschitz continuous in the second component on $[t_0+\infty)\times\mathcal{H}$ with $L > 0$, i.e., $\|d_i(t,u_1) - d_i(t,u_2)\| \leq L\|u_1 - u_2\|$ for all $t\in[t_0,+\infty)$, $u_1,u_2\in\mathcal{H}$ and $i = 1,\dots,m$. Then, there exists $c > 0$ such that for all $(t,u,v)\in[t_0,+\infty)\times\mathcal{H}\times\mathcal{H}$, it holds that

$$\sup_{\xi\in H(t,u,v)} \|\xi\|_{\mathcal{H}\times\mathcal{H}} \leq c(1 + \|(u,v)\|_{\mathcal{H}\times\mathcal{H}}).$$

Proof.

i) Fix $(t,u,v)\in(0,+\infty)\times\mathcal{H}\times\mathcal{H}$. The set $D(t,u) := \operatorname{conv}(\{d_i(t,u) : i = 1,\dots,m\})$ is convex and compact. Then $\arg\min_{g\in D(t,u)}\langle g, -v\rangle$ is also convex and compact and the statement follows since sums and Cartesian products of convex and compact sets are convex and compact.

ii) We show that $H(\cdot,\cdot,\cdot)$ is upper semicontinuous in the ε sense (see Definition 2.2.4) using Lemma 4.3.2. Then, we use Proposition 2.2.5 together with *i)* to conclude $H(\cdot,\cdot,\cdot)$ is upper semicontinuous as well. Using Lemma 4.3.2 we will show that for all $\varepsilon > 0$ there exists $\delta > 0$ satisfying

$$H(B_\delta((\overline{t},\overline{u},\overline{v}))) \subset H(\overline{t},\overline{u},\overline{v}) + B_\varepsilon((0,0)),$$

where $B_\delta((\overline{t},\overline{u},\overline{v}))) \subset \mathbb{R}\times\mathcal{H}\times\mathcal{H}$ and $B_\varepsilon((0,0)) \subset \mathcal{H}\times\mathcal{H}$ are open balls with radius δ and ε, respectively. To this end, we show that for all $(t,u,v)\in(0,+\infty)\times\mathcal{H}\times\mathcal{H}$ with $\|(\overline{t},\overline{u},\overline{v}) - (t,u,v)\|_{\mathbb{R}\times\mathcal{H}\times\mathcal{H}} < \delta$ and for all $(x,y)\in H(t,u,v)$ there exists an element $(\overline{x},\overline{y})\in H(\overline{t},\overline{u},\overline{v})$ with $\|(\overline{x},\overline{y}) - (x,y)\|_{\mathcal{H}\times\mathcal{H}} < \varepsilon$.

Let $(\bar{t}, \bar{u}, \bar{v}) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ and $\varepsilon > 0$ be arbitrary but fixed. For all $(t, u, v) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ the relation $(x, y) \in H(t, u, v)$ is equivalent to

$$\begin{aligned} &x = v, \\ &y = -\gamma(t)v - g, \text{ with} \\ &g \in \underset{g \in D(t,u)}{\arg\min} \langle g, -v \rangle. \end{aligned} \tag{4.44}$$

By Lemma 4.3.2, there exists $\delta_1 > 0$ such that for all $(t, u, v) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ with $\|(t, u, v) - (\bar{t}, \bar{u}, \bar{v})\|_{\mathcal{H}\times\mathcal{H}} < \delta_1$ and all $g \in \arg\min_{g \in D(t,u)} \langle g, -v \rangle$ there exists $\bar{g} \in \arg\min_{\bar{g} \in D(\bar{t},\bar{u})} \langle \bar{g}, -\bar{v} \rangle$ with

$$\|g - \bar{g}\| < \frac{\varepsilon}{3}. \tag{4.45}$$

By continuity of $\gamma(\cdot)$, there exists $\delta_2 > 0$ such that for all $t \in (0, +\infty)$ with $|t - \bar{t}| < \delta_2$ it holds that

$$\left|\gamma(t) - \gamma(\bar{t})\right| \|\bar{v}\| < \frac{\varepsilon}{3}. \tag{4.46}$$

Fix $\delta = \min \left\{ \delta_1, \delta_2, \frac{\varepsilon}{3(1+\gamma(t_0))} \right\}$ and let $(t, u, v) \in (0, +\infty) \times \mathcal{H} \times \mathcal{H}$ with $\|(\bar{t}, \bar{u}, \bar{v}) - (t, u, v)\|_{\mathbb{R}\times\mathcal{H}\times\mathcal{H}} < \delta$ and let $(x, y) = (v, -\gamma(t)v - g) \in H(t, u, v)$ with $g \in \arg\min_{g \in D(t,u)} \langle g, -v \rangle$. By the choice of δ there exists

$$(\bar{x}, \bar{y}) = \left(\bar{v}, -\gamma(\bar{t})\bar{v} - \bar{g}\right) \in H(\bar{t}, \bar{u}, \bar{v}),$$

with $\bar{g} \in \arg\min_{\bar{g} \in D(\bar{t},\bar{u})} \langle \bar{g}, -\bar{v} \rangle$ satisfying (4.45) and (4.46). Then, it follows that

$$\begin{aligned} \|(x, y) - (\bar{x}, \bar{y})\|_{\mathcal{H}\times\mathcal{H}} &\leq \|v - \bar{v}\| + \left\|-\gamma(t)v - g + \gamma(\bar{t})\bar{v} + \bar{g}\right\| \\ &\leq (1 + \gamma(t_0)) \|v - \bar{v}\| + \left|\gamma(t) - \gamma(\bar{t})\right| \|\bar{v}\| + \|g - \bar{g}\| < \varepsilon, \end{aligned}$$

which completes the proof.

iii) If $\dim(\mathcal{H}) < +\infty$, the proof follows from *ii)*. On the other hand, from ψ being locally compact, we follow that $v \mapsto v$ is locally compact which is equivalent to $\mathcal{H}$ being finite-dimensional.

iv) Recall the following inequality between the norm $\|(\cdot, \cdot)\|_{\mathcal{H}\times\mathcal{H}}$ and the norm $\|\cdot\|$. For all $x, y \in \mathcal{H}$, it holds that

$$\|(x, y)\|_{\mathcal{H}\times\mathcal{H}} \leq \|x\| + \|y\| \leq \sqrt{2} \|(x, y)\|_{\mathcal{H}\times\mathcal{H}}. \tag{4.47}$$

Let $(t, u, v) \in [t_0, +\infty) \times \mathcal{H} \times \mathcal{H}$ and $\xi \in H(t, u, v)$. Then, $\xi = (v, -\gamma(t)v - g)$ with $g \in \arg\min_{g \in D(t,u)} \langle g, -v \rangle$. Using the definition of ξ and the first inequality of (4.47) we get

$$\|\xi\|_{\mathcal{H}\times\mathcal{H}} \leq \|v\| + \|\gamma(t)v + g\|.$$

Bounding $\|g\|$ by the element with maximum norm in $D(t, u)$ and using the triangle inequality gives

$$\begin{aligned} &\leq (1 + \gamma(t)) \|v\| + \max_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i d_i(t, u) \right\| \\ &\leq (1 + \gamma(t_0)) \|v\| + \max_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i \left(d_i(t, u) - d_i(t_0, 0)\right) \right\| + \max_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i d_i(t_0, 0) \right\|. \end{aligned}$$

In the next step, we use the Lipschitz continuity of $d_i(\cdot,\cdot)$ and get

$$\leq (1+\gamma(t_0))\,\|v\| + L\|u\| + \max_{i=1,\ldots,m} \|d_i(t_0,0)\|\,.$$

Finally, we use the second inequality from (4.47) and have

$$\leq c\left(1 + \|(u,v)\|_{\mathcal{H}\times\mathcal{H}}\right),$$

with $c = \sqrt{2}\max\left(\{1+\gamma(t_0), L, \max_{i=1,\ldots,m}\|d_i(t_0,0)\|\}\right)$. ☐

The next theorem states an existence result for the differential inclusion (DI-D).

Theorem 4.3.4. *Assume $\mathcal{H}$ is finite-dimensional. Then, for all $(u_0,v_0)\in\mathcal{H}\times\mathcal{H}$ there exists $T>t_0$ and and a solution to the differential inclusion* (DI-D) *on $[t_0,T]$, i.e., there exists an absolutely continuous function $(u,v):[t_0,T]\to\mathcal{H}\times\mathcal{H}$, $t\mapsto(u(t),v(t))$ with $(u(t_0),v(t_0)) = (u_0,v_0)$ and which satisfies*

$$(\dot{u}(t),\dot{v}(t)) \in H(t,u(t),v(t)),$$

for almost all $t\in(t_0,T)$.

Proof. The proof follows immediately from Proposition 4.3.3 *i) - iii)* which shows that the set-valued map $H(\cdot,\cdot,\cdot)$ given by Definition 4.3.1 satisfies all conditions required to apply Theorem 2.2.7. ☐

Theorem 4.3.4 states the existence of local solutions to (DI-D). In the following theorem, we extend local solutions to global solutions using a standard technique, which relies on Zorn's Lemma.

Theorem 4.3.5. *Assume $\mathcal{H}$ is finite dimensional and assume the function $d_i(\cdot,\cdot)$ are uniformly L-Lipschitz continuous in the second component on $[t_0,+\infty)\times\mathcal{H}$ with $L>0$, i.e., $\|d_i(t,u_1)-d_i(t,u_2)\|\leq L\|u_1-u_2\|$ for all $t\in[t_0,+\infty)$, $u_1,u_2\in\mathcal{H}$ and $i=1,\ldots,m$. Then, for all $(u_0,v_0)\in\mathcal{H}\times\mathcal{H}$ there exists a solution to the differential inclusion* (DI-D) *on $[t_0,+\infty)$, i.e., there exists a continuous function $(u,v):[t_0,+\infty)\to\mathcal{H}\times\mathcal{H}$, $t\mapsto(u(t),v(t))$ with $(u(t_0),v(t_0))=(u_0,v_0)$ which is absolutely continuous on every compact interval $[t_0,T]\subset[t_0,+\infty)$ and which satisfies*

$$(\dot{u}(t),\dot{v}(t)) \in H(t,u(t),v(t)),$$

for almost all $t\in(t_0,+\infty)$.

Proof. Define the set

$$\begin{aligned}\mathfrak{S} := \big\{(u,v,T) : T\in(t_0,+\infty] \text{ and } (u,v):[t_0,T)\to\mathcal{H}\times\mathcal{H} \text{ is absolutely continuous on every}\\ \text{compact interval contained in } [t_0,T) \text{ and is a solution of (DI-D) on } [t_0,T)\big\}.\end{aligned}$$

(Note that $T\in(t_0,+\infty]$ in the definition of $\mathfrak{S}$ allows for the value $+\infty$ for T.) By Theorem 4.3.4, the set $\mathfrak{S}$ is not empty. On $\mathfrak{S}$ we define the partial order $\preccurlyeq$ the following way. For $(u_1,v_1,T_1),(u_2,v_2,T_2)\in\mathfrak{S}$, define

$$(u_1,v_1,T_1)\preccurlyeq(u_2,v_2,T_2) \quad :\Longleftrightarrow \quad T_1\leq T_2 \text{ and } (u_1(t),v_1(t)) = (u_2(t),v_2(t)) \text{ for all } t\in[t_0,T_1).$$

The partial order $\preccurlyeq$ is reflexive, transitive and antisymmetric. We show that any nonempty, totally ordered subset of $\mathfrak{S}$ has an upper bound in $\mathfrak{S}$. Let $\mathfrak{C} \subseteq \mathfrak{S}$ be a totally ordered nonempty subset of $\mathfrak{S}$. We define

$$T_{\mathfrak{C}} := \sup \{T : (u, v, T) \in \mathfrak{C}\},$$

and

$$(u_{\mathfrak{C}}, v_{\mathfrak{C}}) : [t_0, T_{\mathfrak{C}}) \to \mathcal{H} \times \mathcal{H},\ (u_{\mathfrak{C}}, v_{\mathfrak{C}})(t) := (u(t), v(t)) \text{ for } t < \bar{t} < T_{\mathfrak{C}} \text{ and } (u, v, \bar{t}) \in \mathfrak{C}.$$

By construction, $(u_{\mathfrak{C}}, v_{\mathfrak{C}}, T_{\mathfrak{C}}) \in \mathfrak{S}$ and $(u, v, T) \preccurlyeq (u_{\mathfrak{C}}, v_{\mathfrak{C}}, T_{\mathfrak{C}})$, hence there exists an upper bound of $\mathfrak{C}$ in $\mathfrak{S}$. According to Zorn's Lemma (see [61, 120]), there exists a maximal element in $\mathfrak{S}$, which we denote by (u, v, T). If $T = +\infty$, the proof is complete. Assume that $T < +\infty$. We show that this contradicts the maximality of (u, v, T) in $\mathfrak{S}$. Define on $[t_0, T)$ the function

$$h : [t_0, T) \to \mathbb{R}, t \mapsto h(t) := \|(u(t), v(t)) - (u(t_0), v(t_0))\|_{\mathcal{H} \times \mathcal{H}}.$$

Using the Cauchy–Schwarz inequality, we get for almost all $t \in [t_0, T)$

$$\frac{d}{dt}\left(\frac{1}{2}h^2(t)\right) = \langle (\dot{u}(t), \dot{v}(t)), (u(t), v(t)) - (u(t_0), v(t_0)) \rangle_{\mathcal{H} \times \mathcal{H}} \leq \|(\dot{u}(t), \dot{v}(t))\|_{\mathcal{H} \times \mathcal{H}}\, h(t). \quad (4.48)$$

Proposition 4.3.3 *iv)* guarantees the existence of a constant $c > 0$ with

$$\|(\dot{u}(t), \dot{v}(t))\|_{\mathcal{H} \times \mathcal{H}} \leq c(1 + \|(u(t), v(t))\|_{\mathcal{H} \times \mathcal{H}}), \quad (4.49)$$

for almost all $t \in [t_0, T)$. Define $\tilde{c} := c\,(1 + \|(u(t_0), v(t_0))\|_{\mathcal{H} \times \mathcal{H}})$. We apply the triangle inequality and get for almost all $t \in [t_0, T)$

$$\|(\dot{u}(t), \dot{v}(t))\|_{\mathcal{H} \times \mathcal{H}} \leq \tilde{c}\,(1 + \|(u(t), v(t)) - (u(t_0), v(t_0))\|_{\mathcal{H} \times \mathcal{H}}). \quad (4.50)$$

Combining (4.48) and (4.50) gives

$$\frac{d}{dt}\left(\frac{1}{2}h^2(t)\right) \leq \tilde{c}\,(1 + h(t))\, h(t). \quad (4.51)$$

Using a Gronwall-type argument (see Lemma 2.2.9, Lemma 2.2.10 and Theorem 3.5 in [16]), we conclude from (4.51) that for all $t \in [t_0, T)$

$$h(t) \leq \tilde{c}T \exp(\tilde{c}T).$$

Therefore, h is bounded on $[t_0, T)$. Then, u and v are also bounded on $[t_0, T)$ and from (4.49) we deduce that $\dot{u}$ and $\dot{v}$ are essentially bounded. This and the fact that $\dot{u}$ and $\dot{v}$ are absolutely continuous guarantees that

$$u_T := u_0 + \int_{t_0}^{T} \dot{u}(s)\, ds \in \mathcal{H} \quad \text{and} \quad v_T := v_0 + \int_{t_0}^{T} \dot{v}(s)\, ds \in \mathcal{H}$$

are well-defined. In the next step we extend the solution from (u_T, v_T). Considering the differential inclusion

$$\left| \begin{array}{ll} (\dot{u}(t), \dot{v}(t)) \in H(t, u(t), v(t)), & \text{for} \quad t > T, \\ \\ (u(T), v(T)) = (u_T, v_T), & \end{array} \right. \tag{4.52}$$

and using Theorem 4.3.4, we obtain that there exist $\delta > 0$ and a solution $(\hat{u}, \hat{v}) : [T, T+\delta] \to \mathcal{H} \times \mathcal{H}$ of (4.52) which is absolutely continuous on compact intervals of $[T, T+\delta]$. Defining

$$(u^*, v^*) : [t_0, T+\delta) \to \mathcal{H} \times \mathcal{H},\ t \mapsto \begin{cases} (u(t), v(t)), & \text{for } t \in [t_0, T), \\ (\hat{u}(t), \hat{v}(t)), & \text{for } t \in [T, T+\delta), \end{cases}$$

we obtain an element $(u^*, v^*, T+\delta) \in \mathfrak{S}$ with the property that $(u, v, T) \neq (u^*, v^*, T+\delta)$ and $(u, v, T) \preccurlyeq (u^*, v^*, T+\delta)$. This is a contradiction to the fact that (u, v, T) is a maximal element in $\mathfrak{S}$. $\square$

4.3.3 Existence of solutions to (D)

Building on the preparatory work carried out in the preceding subsection, we are able to define a solution to (D) and formulate the final existence result.

Definition 4.3.6. *We call a function $x : [t_0, +\infty) \to \mathcal{H}$ a solution to* (D) *with initial data $t_0 > 0$, $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ if it satisfies the following conditions:*

i) $x \in C^1([t_0, +\infty), \mathcal{H})$, *i.e., x is continuously differentiable on $[t_0, +\infty)$;*

ii) $\dot{x}$ *is absolutely continuous on $[t_0, T]$ for all $T \geq t_0$;*

iii) There exists a (Bochner) measurable function $\ddot{x} : [t_0, +\infty) \to \mathcal{H}$ with $\dot{x}(t) = \dot{x}(t_0) + \int_{t_0}^{t} \ddot{x}(s)\, ds$ for all $t \geq t_0$;

iv) $\dot{x}$ *is differentiable almost everywhere and $\frac{d}{dt}\dot{x}(t) = \ddot{x}(t)$ holds for almost all $t \in [t_0, +\infty)$;*

v) $\gamma(t)\dot{x}(t) + \operatorname{proj}_{D(t, x(t)) + \ddot{x}(t)}(0) = 0$ *holds for almost all $t \in [t_0, +\infty)$;*

vi) $x(t_0) = x_0$ *and $\dot{x}(t_0) = v_0$ hold.*

Remark 4.3.7. *Conditions iii) and iv) are merely consequences of ii) (see [76, 87]), since $\dot{x}$ is absolutely continuous on every compact interval $[t_0, T]$ with values in a Hilbert space (which satisfies the Radon-Nikodym property). The (Bochner) measurability of $\ddot{x}$ will be of importance in the analysis of the trajectories in the following subsections.*

In this subsection, we construct trajectory solutions of (D) starting from solutions of the differential inclusion (DI-D). To this end, we use Lemma 2.1.19 to show that solutions of (DI-D) give solutions to (D).

Lemma 4.3.8. *Let $x_0, v_0 \in \mathcal{H}$ and $t_0 > 0$. Assume $(u(\cdot), v(\cdot)) : [t_0, +\infty) \to \mathcal{H} \times \mathcal{H}$ is a solution to* (DI-D) *with $(u(t_0), v(t_0)) = (x_0, v_0)$. Then, it follows that $x(t) \coloneqq u(t)$ satisfies*

$$\gamma(t)\dot{x}(t) + \operatorname*{proj}_{D(t,x(t))+\ddot{x}(t)}(0) = 0,$$

for almost all $t \in [t_0, +\infty)$ and $x(t_0) = x_0$, $\dot{x}(t_0) = v_0$.

Proof. Since $(u(\cdot), v(\cdot))$ is a solution to (DI-D), the relations

$$\begin{aligned} &\dot{u}(t) = v(t) \text{ and} \\ &\dot{v}(t) \in -\gamma(t)v(t) - \operatorname*{arg\,min}_{g \in D(t,u(t))} \langle g, -v(t) \rangle, \end{aligned} \tag{4.53}$$

hold for almost all $t \in [t_0, +\infty)$. Using $\gamma(t) > 0$, we can write the second line as $\dot{v}(t) \in -\gamma(t)v(t) - \arg\min_{g \in D(t,u(t))} \langle g, -\gamma(t)v(t) \rangle$. Using Lemma 2.1.19 with $\eta = -\gamma(t)v(t)$, $C = D(t, u(t))$ and $\xi = \dot{v}(t)$, the second line in (4.53) gives for almost all $t > t_0$

$$-\gamma(t)v(t) = \operatorname*{proj}_{D(t,u(t))+\dot{v}(t)}(0).$$

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 $x(t_0) = u(t_0) = x_0$ and $\dot{x}(t_0) = v(t_0) = v_0$ yields the desired result. □

Finally, we can state the full existence theorem for the system (D).

Theorem 4.3.9. *Assume $\mathcal{H}$ is finite-dimensional and and assume the function $d_i(\cdot, \cdot)$ are uniformly L-Lipschitz continuous in the second component on $[t_0, +\infty) \times \mathcal{H}$ with $L > 0$, i.e., $\|d_i(t, u_1) - d_i(t, u_2)\| \le L\|u_1 - u_2\|$ for all $t \in [t_0, +\infty)$, $u_1, u_2 \in \mathcal{H}$ and $i = 1, \ldots, m$. Then, for all $x_0, v_0 \in \mathcal{H}$, there exists a function $x(\cdot)$ which is a solution to equation* (D) *in the sense of Definition 4.3.6.*

Proof. The proof follows immediately combining Theorem 4.3.4 and Lemma 4.3.8. □

Remark 4.3.10. *In Theorem 4.3.9, we assume that the functions $d_i(\cdot, \cdot)$ are uniformly Lipschitz continuous in the second component on $[t_0, +\infty) \times \mathcal{H}$. One can relax this condition and only require Lipschitz continuity on bounded sets given that the solutions remain bounded.*

4.4 The inertial multiobjective gradient system (IMOG')

In this section, we present the *inertial multiobjective gradient system*

$$\text{(IMOG')} \qquad \alpha\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0, \quad \text{for} \quad t > t_0,$$

with $\alpha > 0$, $C(x) \coloneqq \operatorname{conv}(\{\nabla f_i(x) : i = 1, \dots, m\})$ and initial data $t_0 > 0, x(t_0) = x_0 \in \mathcal{H}$ and $\dot{x}(t_0) = v_0 \in \mathcal{H}$. The system (IMOG') is a further development of the multiobjective steepest descent system (MSD), and does not use the steepest descent direction $\operatorname{proj}_{C(x(t))}(0)$ directly but enhances it by second-order information. In the following, we motivate the derivation of the system (IMOG'). We choose the designation (IMOG') to emphasize its relation to the differential equation

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

with $\mu, \gamma > 0$, which gets introduced in [16] in the context of multiobjective optimization and which forms the foundational step in defining inertial gradient systems for multiobjective optimization. Both systems (IMOG') and (IMOG) reduce in the single objective setting ($m = 1$), to the *heavy ball with friction system*

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

with $\gamma > 0$. The system (HBF) is well studied in various contexts. In the setting of scalar optimization it gets discussed in [196] as a general approach to accelerate iterative methods. In the context of convex optimization, further contributions can be found in [4, 19, 114]. In scalar optimization, it is shown that this system improves the steepest descent dynamical system and gives faster convergence. A discretization of (HBF) leads to a first-order method with momentum which improves the steepest descent method [7, 9]. Similar ideas were applied to solve monotone inclusions [5].

The system (IMOG) can be obtained from (HBF) by replacing the gradient $\nabla f(x(t))$ by the multiobjective steepest descent direction $\operatorname{proj}_{C(x(t))}(0)$. This is a natural approach to an inertial system for multiobjective optimization. In [16, Theorem 4.7] it is shown that the solutions $x(t)$ of (IMOG) in fact converge to solutions of (MOP) as we recite in the following.

Theorem 4.4.1. *Let f_i be convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i with $L > 0$ for $i = 1, \dots, m$. Let $x : [t_0, +\infty) \to \mathcal{H}$ be a bounded solution to* (IMOG) *and assume that $\gamma^2 \geq \mu L$. Then $x(t)$ converges weakly to a weakly Pareto optimal point of* (MOP).

While this result shows that solutions to (IMOG) converge to Pareto optimal solutions of (MOP), there is no theoretical evidence that these solutions converge faster than the ones obtained from the multiobjective steepest descent system (MSD). From an optimization point of view, it would be desirable to use a time dependent damping coefficient $\gamma = \frac{\alpha}{t}$ to obtain gradient systems for multiobjective optimization which behave similar to the *inertial gradient system with asymptotic vanishing damping* for scalar optimization

$$\text{(AVD)} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0,$$

with $\alpha \geq 3$, which gets discussed in [218] in connection to Nesterov's accelerated gradient method [182]. The authors of [16] conjectured that the system (IMOG) with time dependent damping $\gamma = \frac{\alpha}{t}$ gives the same improvement in the context of multiobjective optimization and can lead to fast first-order methods for multiobjective optimization. However, for an adaption of (IMOG) with asymptotically vanishing damping the analysis laid out in [16] breaks down because of the condition $\gamma^2 \geq \mu L$.

When trying to incorporate time dependent damping in (IMOG), it turns out that the generalization of (HBF) to the multiobjective optimization setting might be the problem. We obtain the generalization (IMOG') by investigating the following energy estimate. For the system (HBF), we can prove decaying energy in the form

$$\frac{d}{dt}\left(f(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2\right) = -\alpha\|\dot{x}(t)\|^2. \tag{E}$$

An estimate analogous to this does not hold for the system (IMOG). This observation was the starting point in the derivation of (IMOG'). Generalizing (E) to the multiobjective setting leads to the inequalities

$$\frac{d}{dt}\left(f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2\right) \leq -\alpha\|\dot{x}(t)\|^2,$$

for $i = 1, \dots, m$. Writing out the left-hand side using the chain rule and rewriting this as a variational inequality we obtain (IMOG') from these energy estimates by interpreting the variational inequality as a projection (see Proposition 4.4.5).

From an analytical point of view (IMOG') looks disadvantageous in comparison to (IMOG) since it is an implicit differential equation. And in fact, this makes it harder to prove existence of solutions as we cannot invoke standard results like Peano's Theorem or the Cauchy–Lipschitz Theorem. However, we see in the following sections and chapters that this is the key observation to develop more sophisticated fast gradient dynamics and first-order methods for multiobjective optimization.

This section contains two main results. In the first part, we prove existence of solutions to (IMOG') by using results prepared in Subsection 4.3. This is followed by an asymptotical analysis of the solutions to (IMOG') and a prove on the weak converge to weakly Pareto optimal points of (MOP).

In this section, we make the following standing assumption:

($\mathcal{A}_1$) The objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and continuously differentiable with L-Lipschitz continuous gradients $\nabla f_i : \mathcal{H} \to \mathcal{H}$ for all $i = 1, \dots, m$.

The content of this section was already published in the following paper:

[217] Sonntag, K. and Peitz, S. *Fast Multiobjective Gradient Methods with Nesterov Acceleration via Inertial Gradient-Like Systems.* In: *Journal of Optimization Theory and Applications* 201 (2024), pp. 539–582. doi: 10.1007/s10957-024-02389-3.

4.4.1 Discussion of existence and uniqueness of solutions

The discussion of existence and uniqueness of solutions to (IMOG') is based on Section 4.3. We begin by properly defining a solution to (IMOG').

Definition 4.4.2. *We call a function $x : [t_0, +\infty) \to \mathcal{H}$ a solution to* (IMOG') *with initial data $t_0 > 0$, $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ if it satisfies the following conditions:*

i) $x \in C^1([t_0, +\infty), \mathcal{H})$, *i.e., $x(\cdot)$ is continuously differentiable on $[t_0, +\infty)$;*

ii) *$\dot{x}(\cdot)$ is absolutely continuous on $[t_0, T]$ for all $T \geq t_0$;*

iii) *There exists a (Bochner) measurable function $\ddot{x} : [t_0, +\infty) \to \mathcal{H}$ with $\dot{x}(t) = \dot{x}(t_0) + \int_{t_0}^{t} \ddot{x}(s)\, ds$ for all $t \geq t_0$;*

iv) *$\dot{x}(\cdot)$ is differentiable almost everywhere and $\frac{d}{dt}\dot{x}(t) = \ddot{x}(t)$ holds for almost all $t \in [t_0, +\infty)$;*

v) *$\alpha\dot{x}(t) + \operatorname{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0$ holds for almost all $t \in [t_0, +\infty)$.*

vi) *$x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ hold.*

Theorem 4.4.3. *Assume $\mathcal{H}$ is finite-dimensional and and assume the gradients of the objective function ∇f_i are uniformly L-Lipschitz continuous for all $i = 1, \dots, m$. Then, for all $t_0 > 0$ and $x_0, v_0 \in \mathcal{H}$, there exists a function $x(\cdot)$ which is a solution to equation* (IMOG') *in the sense of Definition 4.4.2.*

Proof. The proof follows from Theorem 4.3.9. We show (IMOG') is a special instance of (D) for appropriate choices of $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$. Define the functions

$$\begin{aligned} &\gamma : (0, +\infty) \to [0, +\infty), \quad t \mapsto \gamma(t) := \alpha, \\ &d_i : (0, +\infty) \times \mathcal{H} \to \mathcal{H}, \quad (t, u) \mapsto \nabla f_i(u), \end{aligned} \tag{4.54}$$

and let $D(\cdot, \cdot)$ be as defined in Definition 4.3.1. By (4.54) the functions $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$ have the following properties. The function $\gamma(\cdot)$ is continuous and monotonically decreasing. The functions $d_i(\cdot, \cdot)$ are continuous on $(0, +\infty) \times \mathcal{H}$ and uniformly L-Lipschitz continuous in the second component on $[t_0, +\infty) \times \mathcal{H}$ for all $i = 1, \dots, m$. Further, for all $(t, u) \in (0, +\infty) \times \mathcal{H}$ it holds that

$$D(t, u) = C(u).$$

For this choice of $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$ equation (D) reduces to (IMOG') and we conclude the existence of a solution to (D) in the sense of Definition 4.3.6 by Theorem 4.3.9. □

Remark 4.4.4. *It remains unclear whether solutions to* (IMOG') *are unique. There are two main challenges in establishing uniqueness. First, the multiobjective steepest descent direction is only Hölder continuous, not Lipschitz continuous (see Proposition 2.3.21 and Remark 2.3.22). Therefore, even for simpler gradient dynamics such as the multiobjective steepest descent dynamical system $\dot{x}(t) + \operatorname{proj}_{C(x(t))}(0) = 0$, we cannot conclude uniqueness without additional assumptions. Second,* (IMOG') *is not an ordinary differential equation, but rather an implicit differential equation. This prevents the application of standard tools like the Cauchy–Lipschitz Theorem to prove uniqueness. Nevertheless, the asymptotic analysis of solutions presented in the following subsections holds independently of uniqueness.*

4.4.2 Preparatory results

In this subsection, we omit the assumption $\dim(\mathcal{H}) < +\infty$. We show that trajectories of the differential equation (IMOG') converge weakly to weakly Pareto optimal points of (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 4.4.5. *Let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (IMOG') *in the sense of Definition 4.4.2. For $i = 1, \dots, m$ define the function*

$$\mathcal{W}_i : [t_0, T) \to \mathbb{R}, \quad t \mapsto \mathcal{W}_i(t) := f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2.$$

Then, for all $i = 1, \dots, m$ and almost all $t \in (t_0, +\infty)$ it holds that $\frac{d}{dt}\mathcal{W}_i(t) \leq -\alpha\|\dot{x}(t)\|^2$.

Proof. By Definition 4.4.2 the function $x(\cdot)$ is continuously differentiable and $\dot{x}(\cdot)$ is absolutely continuous on each compact interval $[t_0, T]$ and therefore differentiable almost everywhere with derivative $\ddot{x}(t)$. Hence, for almost all $t > t_0$ the function $\mathcal{W}_i(\cdot)$ is differentiable and by the chain rule we have

$$\frac{d}{dt}\mathcal{W}_i(t) = \langle \nabla f_i(x(t)), \dot{x}(t)\rangle + \langle \dot{x}(t), \ddot{x}(t)\rangle. \tag{4.55}$$

By the variational characterization of the convex projection $\alpha\dot{x}(t) = \operatorname{proj}_{C(x(t))+\ddot{x}(t)}(0)$, we get for all $i = 1, \dots, m$ and almost all $t > t_0$

$$\langle \alpha\dot{x}(t) + \nabla f_i(x(t)) + \ddot{x}(t), \alpha\dot{x}(t)\rangle \leq 0, \tag{4.56}$$

Together (4.55) and (4.56) imply the desired result. □

Corollary 4.4.6. *Let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (IMOG') *with initial data $t_0 > 0$, $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ in the sense of Definition 4.4.2. Assume there exists $i \in \{1, \dots, m\}$ such that the lower level set*

$$\mathcal{L}\left(f_i, f_i(x(t_0)) + \frac{1}{2}\|\dot{x}(t_0)\|^2\right) = \left\{z \in \mathcal{H} \,:\, f_i(z) \leq f_i(x(t_0)) + \frac{1}{2}\|\dot{x}(t_0)\|^2\right\},$$

is bounded. Then $x(\cdot)$ is bounded.

Proof. By Proposition 4.4.5 the function $\mathcal{W}_i(\cdot)$ is monotonically decreasing for all $i = 1, \dots, m$ and therefore

$$f_i(x(t)) \leq \mathcal{W}_i(t) \leq \mathcal{W}_i(t_0) = f_i(x(t_0)) + \frac{1}{2}\|\dot{x}(t_0)\|^2.$$

Hence

$$x(t) \in \mathcal{L}\left(f_i, f_i(x(t_0)) + \frac{1}{2}\|\dot{x}(t_0)\|^2\right),$$

for all $i = 1, \dots, m$. If one of these sets is bounded then $x(t)$ is bounded as well. □

Proposition 4.4.7. *Let $x : [t_0, +\infty) \to \mathcal{H}$ be a bounded solution of* (IMOG') *in the sense of Definition 4.4.2 and let further ∇f_i be Lipschitz continuous on bounded sets. Then, for all $i = 1, \dots, m$ it holds that:*

i) $\mathcal{W}_i^\infty := \lim_{t\to+\infty} \mathcal{W}_i(t) > -\infty$;

ii) $\dot{x} \in L^2([t_0, +\infty), \mathcal{H}) \cap L^\infty([t_0, +\infty), \mathcal{H})$;

iii) $\ddot{x} \in L^\infty([t_0, +\infty), \mathcal{H})$, $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$ *and* $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{W}_i^\infty$;

iv) There exists a monotonically increasing unbounded sequence $(t_k)_{k\geq 0}$ with $\operatorname{proj}_{C(x(t_k))}(0) \to 0$ *as $k \to +\infty$.*

Proof. i) From Proposition 4.4.5, we immediately get that $\mathcal{W}_i(\cdot)$ is monotonically decreasing and therefore $\mathcal{W}_i^\infty := \lim_{t\to+\infty} \mathcal{W}_i(t) \in \mathbb{R} \cup \{-\infty\}$ exists. Next, we show $\mathcal{W}_i^\infty > -\infty$. 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 $x(\cdot)$ remains bounded by assumption, we conclude that $f_i(x(t))$ is uniformly bounded from below for all $t \geq t_0$, and hence

$$\mathcal{W}_i^\infty \geq \inf_{t\geq t_0} f_i(x(t)) > -\infty.$$

ii) We know that $f_i(x(t))$ is bounded. Then, by the definition of $\mathcal{W}_i(\cdot)$ and the fact that $\mathcal{W}_i(\cdot)$ is monotonically decreasing, we immediately get that $\dot{x}(\cdot)$ is bounded for all $t \geq t_0$. Since it is continuous, it follows that $\dot{x} \in L^\infty([t_0, +\infty), \mathcal{H})$. Using Proposition 4.4.5, we bound

$$\alpha \int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 \, dt \leq -\int_{t_0}^{+\infty} \frac{d}{dt}\mathcal{W}_i(s) \, ds = \mathcal{W}_i(t_0) - \mathcal{W}_i^\infty < +\infty,$$

and therefore $\dot{x} \in L^2([t_0, +\infty), \mathcal{H})$.

iii) By Lemma 2.1.18 the solution $x(\cdot)$ satisfies for almost all $t > t_0$,

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

Since $\dot{x}(t)$ and $\nabla f_i(x(t))$ remain bounded for all $t \geq t_0$ it follows that $\ddot{x}(t) = -\alpha\dot{x}(t) - \operatorname{proj}_{C(x(t))}(-\ddot{x}(t))$ remains bounded for almost all $t \geq t_0$. By Definition 4.4.2, the function $\ddot{x}(\cdot)$ is measurable and we follow $\ddot{x} \in L^\infty([t_0, +\infty), \mathcal{H})$. Then, from $\dot{x} \in L^2([t_0, +\infty), \mathcal{H})$ being absolutely continuous and $\ddot{x} \in L^\infty([t_0, +\infty), \mathcal{H})$ it follows that $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$. We Conclude $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{W}_i^\infty$ by $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$ and part *i)*.

iv) Assume the negation of statement *iv)* holds, i.e., there exists $M > 0$ and $T \geq t_0$ such that for all $t \geq T$ it holds that

$$\left\| \operatorname*{proj}_{C(x(t))}(0) \right\| \geq M. \tag{4.57}$$

Fix an arbitrary $\delta > 0$. Since $\dot{x}(t) \to 0$ and ∇f_i is Lipschitz continuous on a set containing $x(t)$ it follows that there exists $T_\delta > T$ such that for all $t > T_\delta$ and $s \in [t, t+\delta)$

$$\|\nabla f_i(x(s)) - \nabla f_i(x(t))\| < \frac{M}{4} \quad \text{and} \quad \|\alpha \dot{x}(s)\| < \frac{M}{4}. \tag{4.58}$$

Let $t \geq T_\delta$ be arbitrary. Define $v(t) \coloneqq \operatorname{proj}_{C(x(t))} / \|\operatorname{proj}_{C(x(t))}\|$. From (4.57) it follows that

$$\langle \xi(t), v(t) \rangle \geq M \quad \text{for all} \quad \xi(t) \in C(x(t)). \tag{4.59}$$

Next, we combine (4.58) and (4.59). Let $s \in [t, t+\delta)$ and $\xi(s) \in C(x(s))$. There exists $\theta(s) \in \Delta^m$ with $\xi(s) = \sum_{i=1}^m \theta_i(s) \nabla f_i(x(s))$. Then

$$\begin{aligned} \langle \xi(s) + \alpha \dot{x}(s), v(t) \rangle &= \left\langle \sum_{i=1}^m \theta_i(s) \nabla f_i(x(s)), v(t) \right\rangle + \langle \alpha \dot{x}(s), v(t) \rangle \\ = \left\langle \sum_{i=1}^m \theta_i(s) \nabla f_i(x(t)), v(t) \right\rangle &+ \sum_{i=1}^m \theta_i(s) \langle \nabla f_i(x(s)) - \nabla f_i(x(t)), v(t) \rangle + \langle \alpha \dot{x}(s), v(t) \rangle. \end{aligned} \tag{4.60}$$

Since $\sum_{i=1}^m \theta_i(s) \nabla f_i(x(t)) \in C(x(t))$, we can use (4.59) to bound this by

$$\geq M + \sum_{i=1}^m \theta_i(s) \langle \nabla f_i(x(s)) - \nabla f_i(x(t)), v(t) \rangle + \langle \alpha \dot{x}(s), v(t) \rangle. \tag{4.61}$$

Now, we apply the Cauchy–Schwarz inequality and get

$$\geq M - \sum_{i=1}^m \theta_i(s) \|\nabla f_i(x(s)) - \nabla f_i(x(t))\| \, \|v(t)\| - \|\alpha \dot{x}(s)\| \, \|v(t)\|. \tag{4.62}$$

By definition we have $\|v(t)\| = 1$. Then, we use (4.58) to bound (4.62) by

$$\geq M - \frac{M}{4} - \frac{M}{4} = \frac{M}{2}. \tag{4.63}$$

Inequalities (4.60) – (4.63) give

$$\langle -\ddot{x}(s), v(t) \rangle \geq \frac{M}{2} \quad \text{for almost all} \quad s \in [t, t+\delta).$$

Using the Cauchy–Schwarz inequality once more, we get

$$\begin{aligned} \|\dot{x}(t) - \dot{x}(t+\delta)\| &\geq \langle \dot{x}(t) - \dot{x}(t+\delta), v(t) \rangle = \left\langle \int_t^{t+\delta} -\ddot{x}(s) \, ds, v(t) \right\rangle \\ &= \int_t^{t+\delta} \langle -\ddot{x}(s), v(t) \rangle \, ds \geq \int_t^{t+\delta} \frac{M}{2} \, ds = \frac{M\delta}{2}. \end{aligned}$$

Since this holds for all $t \geq T_\delta$ the functions $\dot{x}(\cdot)$ is not Cauchy and therefore not converging which contradicts $\lim_{t \to +\infty} \dot{x}(t) = 0$. □

We use part *iv)* of Proposition 4.4.7 to show that each weak limit point of the trajectory $x(t)$ is Pareto critical and hence weakly Pareto optimal.

4.4.3 Asymptotic analysis

If we can show that the trajectories of (IMOG') converge weakly, Proposition 4.4.7 together with Lemma 2.3.24 guarantees that the limit points are Pareto critical and therefore weakly Pareto optimal. To prove convergence of solutions, we require Opial's Lemma (Lemma 2.1.6).

Proposition 4.4.8. *Let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution of* (IMOG') *in the sense of Definition 4.4.2 and assume $x(\cdot)$ is bounded. Then, the set*

$$S \coloneqq \{z \in \mathcal{H} \,:\, f_i(z) \leq \mathcal{W}_i^\infty \text{ for all } i = 1, \ldots, m\}, \tag{4.64}$$

is nonempty.

Proof. Part *iii)* of Proposition 4.4.7 states that $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{W}_i^\infty$ for all $i = 1, \ldots, m$. Since $x(\cdot)$ is bounded, it possesses at least one weak sequential cluster point x^∞, i.e., there exists a sequence $(t_k)_{k\geq 0}$ with $x(t_k) \rightharpoonup x^\infty$ as $k \to +\infty$. The objective functions f_i are convex and continuous and therefore weakly lower semicontinuous for all $i = 1, \ldots, m$. Therefore

$$f_i(x^\infty) \leq \liminf_{k\to+\infty} f_i(x(t_k)) = \lim_{t\to+\infty} f_i(x(t)) = \mathcal{W}_i^\infty,$$

and we conclude $x^\infty \in S$. □

For the set S defined in (4.64) and a bounded solution $x(\cdot)$ of (IMOG'), the first part of Opial's Lemma is easy to obtain. It follows analogously to the proof of Proposition 4.4.8 where it is shown that S is nonempty. To show the second part of Opial's Lemma, we verify that $h_z(t) \coloneqq \frac{1}{2}\|x(t) - z\|^2$ satisfies a differential inequality. Then, the convergence can be deduced from Lemma 2.2.11. With these ingredients, we can formulate the main convergence theorem of this subsection.

Theorem 4.4.9. *Assume the gradients ∇f_i are Lipschitz continuous on bounded sets for $i = 1, \ldots, m$. Let $x : [t_0, +\infty) \to \mathcal{H}$ be a solution of* (IMOG') *in the sense of Definition 4.4.2 and assume $x(\cdot)$ is bounded. Then, $x(\cdot)$ converges weakly to a weakly Pareto optimal point of* (MOP).

Proof. Let S be as defined in (4.64). For $z \in S$ define

$$h_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto h_z(t) \coloneqq \frac{1}{2}\|x(t) - z\|^2.$$

By Definition 4.4.2 the function $h_z(\cdot)$ is continuously differentiable with absolutely continuous derivative. Using the chain rule, we compute the first and the second derivative of $h_z(\cdot)$ as

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

where the equation for $\ddot{h}_z(t)$ only holds almost everywhere. For almost all $t \in (t_0, +\infty)$

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

Using the definition of (IMOG'), we write $\ddot{x}(t) + \alpha\dot{x}(t) = -\sum_{i=1}^{m} \theta_i(t)\nabla f_i(x(t))$ for almost all $t \in (t_0, +\infty)$ with weights $\theta(t) \in \Delta_m$. Then, for almost all $t \in (t_0, +\infty)$

$$\alpha\dot{h}_z(t) + \ddot{h}_z(t) = \sum_{i=1}^{m} \theta_i(t)\langle \nabla f_i(x(t)), z - x(t)\rangle + \|\dot{x}(t)\|^2. \tag{4.65}$$

By Proposition 4.4.5 $\mathcal{W}_i(\cdot)$ is monotonically decreasing for all $i = 1, \dots, m$ and we get

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

and therefore

$$\sum_{i=1}^{m} \theta_i(t)\langle \nabla f_i(x(t)), z - x(t)\rangle \leq \frac{1}{2}\|\dot{x}(t)\|^2. \tag{4.66}$$

Combining inequalities (4.65) and (4.66) gives for almost all $t \in (t_0, +\infty)$

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

By Proposition 4.4.7, we know $\|\dot{x}(\cdot)\|^2 \in L^1([t_0, +\infty), \mathbb{R})$. Then, Lemma 2.2.11 states 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 for $i = 1, \dots, m$. Then, we can use Opial's Lemma (Lemma 2.1.6) to prove that $x(t)$ converges weakly to an element in S. Let x^{∞} be the weak limit of $x(t)$. Then, by Proposition 4.4.7, there exists a monotonically increasing unbounded sequence $(t_k)_{k\geq 0}$ with $\operatorname{proj}_{C(x(t_k))}(0) \to 0$ for $k \to +\infty$. Since $x(t_k)$ converges weakly to x^{∞}, Lemma 2.3.24 states that x^{∞} is Pareto critical. Since all objective functions are convex, we conclude that x^{∞} is weakly Pareto optimal. □

4.5 The multiobjective gradient system with asymptotic vanishing damping (MAVD)

In this section, we discuss the *multiobjective gradient system with asymptotic vanishing damping*

$$\text{(MAVD)} \qquad \frac{\alpha}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0,$$

with $\alpha > 0$, $C(x) \coloneqq \operatorname{conv}(\{\nabla f_i(x) : i = 1, \dots, m\})$ and initial data $t_0 > 0, x(t_0) = x_0 \in \mathcal{H}$ and $\dot{x}(t_0) = v_0 \in \mathcal{H}$. Our interest in the system (MAVD) is motivated by the active research in dynamical systems for fast minimization and their relationship with numerical optimization methods. In Section 4.4, we described the development of the *inertial multiobjective gradient system*

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

as a generalization of the *heavy ball with friction system*

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

to multiobjective optimization. Adapting the system (HBF) by time dependent damping leads to the *inertial gradient system with asymptotic vanishing damping*

$$\text{(AVD)} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0,$$

which is introduced in [218] in connection with *Nesterov's accelerated gradient method* [182] and analyzed further in [13, 58, 59]. For $\alpha > 0$, every solution $x(\cdot)$ of (AVD) satisfies $\lim_{t\to+\infty} f(x(t)) = \min_{x\in\mathcal{H}} f(x)$. For $\alpha \geq 3$, it holds that $f(x(t)) - \min_{x\in\mathcal{H}} f(x) = \mathcal{O}(t^{-2})$ [218]. For $\alpha > 3$, the trajectories experience an improved converge rate of order $f(x(t)) - \min_{x\in\mathcal{H}} f(x) = o(t^{-2})$, and every solution $x(\cdot)$ converges weakly to a minimizer of f given that the set of minimizers is nonempty [13, 166].

It was an open question whether similar results can be obtained for multiobjective optimization problems [16]. This question is answered positively in [216] by introducing the system (MAVD), which this section is dedicated to. Similar to (AVD) the system (MAVD) is obtained by enhancing (IMOG') using an asymptotically vanishing damping coefficient $\gamma = \frac{\alpha}{t}$. In this section, we show that this system significantly improves on the multiobjective steepest descent dynamic (MSD). We prove that the function values of a solution $x(\cdot)$ to (MAVD) with $\alpha \geq 3$ decay faster with rate $\varphi(x(t)) = \mathcal{O}(t^{-2})$ measured with the merit function $\varphi(\cdot)$. Additionally, we prove weak convergence of the trajectories to weakly Pareto optimal points of (MOP) under the condition $\alpha > 3$. In the single objective case ($m = 1$) these findings reduce to the known results for the system (AVD) [13, 15, 58, 59, 166, 218].

The content of this section was already published in the following paper:

[216] Sonntag, K. and Peitz, S. *Fast convergence of inertial multiobjective gradient-like systems with asymptotic vanishing damping.* In: *SIAM Journal on Optimization* 34 (3) (2024), pp. 2259–2286. doi: 10.1137/23M1588512.

4.5.1 Assumptions

($\mathcal{A}_1$) The objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and continuously differentiable with L-Lipschitz continuous gradients $\nabla f_i : \mathcal{H} \to \mathcal{H}$ with $L > 0$ for all $i = 1, \dots, m$.

($\mathcal{A}_2$) For all $x_0 \in \mathcal{H}$ and for all $x \in \mathcal{L}(F, F(x_0))$ it holds that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ and further

$$R := \sup_{F^* \in F(\mathcal{LP}_w(F,F(x_0)))} \inf_{z \in F^{-1}(\{F^*\})} \frac{1}{2}\|z - x_0\|^2 < +\infty. \tag{4.67}$$

Discussion of Assumption ($\mathcal{A}_2$)

In the literature, Assumption ($\mathcal{A}_2$) is a typical assumption made in the asymptotic analysis of gradient systems and first-order methods for multiobjective optimization [156, 221, 222, 223, 224]. In the following remark, we compare Assumption ($\mathcal{A}_2$) with common assumptions made in scalar optimization.

Remark 4.5.1. *Assumption* ($\mathcal{A}_2$) *is satisfied in the following cases.*

i) For singleobjective optimization problems, Assumption ($\mathcal{A}_2$) *is satisfied if the optimization problem has at least one optimal solution. In this setting, for all* $x_0 \in \mathcal{H}$ *the weak Pareto set* $\mathcal{P}_w = \mathcal{LP}_w(F, F(x_0))$ *coincides with the solution set* $\arg\min_{x \in \mathcal{H}} f(x)$ *and* $\inf_{x \in P_w} \frac{1}{2}\|x - x_0\|^2 < +\infty$ *holds.*

ii) Assumption ($\mathcal{A}_2$) *is valid, if the level set* $\mathcal{L}(F, F(x_0))$ *is bounded. For example, this is the case when for at least one* $i \in \{1, \dots, m\}$ *the set* $\{x \in \mathcal{H} : f_i(x) \leq f_i(x_0)\}$ *is bounded.*

We close the discussion on Assumption ($\mathcal{A}_2$) by two examples giving further context to Remark 4.5.1.

Example 4.5.2. *We discuss Assumption* ($\mathcal{A}_2$) *and Remark 4.5.1 by the means of two examples.*

i) Assumption ($\mathcal{A}_2$) *cannot hold when the weak Pareto set is empty. Since any scalar optimization problem is also a multiobjective optimization problem, we can consider*

$$\min_{x \in \mathbb{R}} F(x) := \exp(x). \tag{MOP-Ex$_1$}$$

For all $x_0 \in \mathbb{R}$ *the set* $\mathcal{LP}_w(F, F(x_0)) \subset \mathcal{P}_w = \arg\min_{x \in \mathbb{R}} F(x) = \emptyset$ *is empty and hence Assumption* ($\mathcal{A}_2$) *does not hold.*

ii) In the second example the weak Pareto set is nonempty but the supremum defining R *is not bounded. Consider the multiobjective optimization problem*

$$\min_{x \in \mathbb{R}^2} F(x) := \begin{bmatrix} x_1^2 \\ \exp(x_2) \end{bmatrix}, \tag{MOP-Ex$_2$}$$

with two objective functions defined on $\mathbb{R}^2$*. For* (MOP-Ex$_2$) *the weak Pareto set is* $\mathcal{P}_w = \{0\} \times \mathbb{R}$*. For all* $x_0 \in \mathbb{R}^2$ *it holds that* $\mathcal{LP}_w(F, F(x_0)) = \{0\} \times (-\infty, (x_0)_2] \neq \emptyset$*, but for this problem* $R = +\infty$*.*

For the problems (MOP-Ex$_1$) *and* (MOP-Ex$_2$) *all objective functions have unbounded level sets. As stated in Remark 4.5.1 ii), a bounded level set of one objective functions is a sufficient condition for Assumption* ($\mathcal{A}_2$) *to hold.*

4.5.2 Discussion of existence and uniqueness of solutions

Similar to the last section where the system (IMOG') gets discussed, the discussion of existence and uniqueness of solutions to (MAVD) is based on Section 4.3. In the following definition, we describe what is to be understood under a solution to (MAVD).

Definition 4.5.3. *We call a function $x : [t_0, +\infty) \to \mathcal{H}$ a solution to* (MAVD) *with initial data $t_0 > 0$, $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ if it satisfies the following conditions:*

i) $x \in C^1([t_0, +\infty), \mathcal{H})$, i.e., $x(\cdot)$ is continuously differentiable on $[t_0, +\infty)$;

ii) $\dot{x}(\cdot)$ is absolutely continuous on $[t_0, T]$ for all $T \geq t_0$;

iii) There exists a (Bochner) measurable function $\ddot{x} : [t_0, +\infty) \to \mathcal{H}$ with $\dot{x}(t) = \dot{x}(t_0) + \int_{t_0}^t \ddot{x}(s)ds$ for all $t \geq t_0$;

iv) $\dot{x}(\cdot)$ is differentiable almost everywhere and $\frac{d}{dt}\dot{x}(t) = \ddot{x}(t)$ holds for almost all $t \in [t_0, +\infty)$;

v) $\frac{\alpha}{t}\dot{x}(t) + \operatorname{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0$ holds for almost all $t \in [t_0, +\infty)$;

vi) $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ hold.

Theorem 4.5.4. *Assume $\mathcal{H}$ is finite-dimensional and and assume the gradients of the objective function ∇f_i are L-Lipschitz continuous with $L > 0$ for all $i = 1, \dots, m$. Then, for all $t_0 > 0$ and $x_0, v_0 \in \mathcal{H}$, there exists a function $x(\cdot)$ which is a solution to equation* (IMOG') *in the sense of Definition 4.5.3.*

Proof. The proof follows from Theorem 4.3.9. We show (MAVD) is a special instance of (D) for appropriate choices of $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$. Define the functions

$$\begin{aligned} &\gamma : (0, +\infty) \to [0, +\infty), \quad t \mapsto \gamma(t) := \frac{\alpha}{t}, \\ &d_i : (0, +\infty) \times \mathcal{H} \to \mathcal{H}, \quad (t, u) \mapsto d_i(t, u) := \nabla f_i(u), \end{aligned} \tag{4.68}$$

and let $D(\cdot, \cdot)$ be as given in Definition 4.3.1. By (4.68) the functions $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$ have the following properties. The function $\gamma(\cdot)$ is continuous and monotonically decreasing. The functions $d_i(\cdot, \cdot)$ are continuous on $(0, +\infty) \times \mathcal{H}$ and uniformly L-Lipschitz continuous in the second component on $[t_0, +\infty) \times \mathcal{H}$ for all $i = 1, \dots, m$. Further, for all $(t, u) \in (0, +\infty) \times \mathcal{H}$ it holds that

$$D(t, u) = C(u).$$

For this choice of $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$ equation (D) reduces to (MAVD) and the existence of a solution to (D) in the sense of Definition 4.3.6 holds by Theorem 4.3.9. □

Remark 4.5.5. *The discussion of the uniqueness of solutions to* (MAVD) *is analogous to the discussion for* (IMOG'). *The multiobjective steepest descent direction is not Lipschitz continuous, but merely Hölder continuous (see Proposition 2.3.21 and Remark 2.3.22). As with the system* (IMOG'), *the second difficulty arises from the implicit structure of the equation* (MAVD), *which precludes the application of standard results like the Cauchy–Lipschitz Theorem to establish uniqueness.*

4.5.3 Preparatory results

In this section, we collect some preliminary results on the solutions to (MAVD). We show that the trajectories $x(\cdot)$ of (MAVD) minimize the function values. In Theorem 4.5.12, we show for $\alpha > 0$ that $\varphi(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2 \to 0$ as $t \to +\infty$. By this, it follows that every weak accumulation point of $x(\cdot)$ is weakly Pareto optimal.

We omit the assumption $\dim(\mathcal{H}) < +\infty$ from this point on. Throughout this subsection, we fix a solution $x : [t_0, +\infty) \to \mathcal{H}$ to (MAVD) in the sense of Definition 4.5.3 with initial velocity $\dot{x}(t_0) = 0$. Setting the initial velocity to zero has the advantage that the trajectories $x(\cdot)$ remain in the level set $\mathcal{L}(F, F(x_0))$ as stated in Corollary 4.5.7. Hence, if the level set $\mathcal{L}(F, F(x_0))$ is bounded, the solution $x(\cdot)$ remains bounded. The results can be generalized to the case $\dot{x}(t_0) \neq 0$.

Proposition 4.5.6. *For $i = 1, \dots, m$, define the component-wise energy functions*

$$\mathcal{W}_i : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{W}_i(t) := f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2. \tag{4.69}$$

Then, for all $i = 1, \dots, m$ and almost all $t \in [t_0, +\infty)$, it holds that $\frac{d}{dt}\mathcal{W}_i(t) \leq -\frac{\alpha}{t}\|\dot{x}(t)\|^2$. Hence, $\mathcal{W}_i(\cdot)$ is monotonically decreasing, and $\mathcal{W}_i^\infty := \lim_{t\to+\infty} \mathcal{W}_i(t) \in \mathbb{R} \cup \{-\infty\}$ exists. If f_i is bounded from below, then $\mathcal{W}_i^\infty \in \mathbb{R}$.

Proof. The function $\mathcal{W}_i$ is differentiable almost everywhere in $[t_0, +\infty)$ with derivative

$$\frac{d}{dt}\mathcal{W}_i(t) = \frac{d}{dt}\left[f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2\right] = \langle \nabla f_i(x(t)), \dot{x}(t)\rangle + \langle \dot{x}(t), \ddot{x}(t)\rangle . \tag{4.70}$$

Using the variational representation of $-\frac{\alpha}{t}\dot{x}(t) = \operatorname{proj}_{C(x(t))+\ddot{x}(t)}(0)$ and the fact that $\nabla f_i(x(t)) \in C(x(t))$, we get for all $i = 1, \dots, m$

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

and hence,

$$\langle \nabla f_i(x(t)), \dot{x}(t)\rangle + \langle \ddot{x}(t), \dot{x}(t)\rangle \leq -\frac{\alpha}{t}\|\dot{x}(t)\|^2. \tag{4.71}$$

Combining (4.70) and (4.71) gives the desired results. □

Due to the inertial effects in (MAVD), there is in general no monotone descent for the objective values along the trajectories. The following corollary guarantees that the function values along the trajectories are bounded from above by the initial function values given $\dot{x}(t_0) = 0$.

Corollary 4.5.7. *For all $i = 1, \dots, m$ and all $t \in [t_0, +\infty)$, it holds that*

$$f_i(x(t)) \leq f_i(x_0),$$

i.e., $x(t) \in \mathcal{L}(F, F(x_0))$ for all $t \geq t_0$.

Proof. From Proposition 4.5.6, we follow for all $t \in [t_0, +\infty)$

$$f_i(x_0) = \mathcal{W}_i(t_0) \geq \mathcal{W}_i(t) = f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2 \geq f_i(x(t)).$$

□

In the following analysis, we use the weights $\theta(t) \in \Delta^m$ which are implicitly given by

$$-\frac{\alpha}{t}\dot{x}(t) = \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = \sum_{i=1}^m \theta_i(t)\nabla f_i(x(t)) + \ddot{x}(t), \tag{4.72}$$

for almost all $t \in [t_0, +\infty)$. To evaluate the integral over the function $\theta(\cdot)$, we have to guarantee that we can find a measurable selection $t \mapsto \theta(t) \in \Delta^m$ satisfying (4.72).

Lemma 4.5.8. *There exists a measurable function*

$$\theta : [t_0, +\infty) \to \Delta^m, \quad t \mapsto \theta(t),$$

with

$$\operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = \sum_{i=1}^m \theta_i(t)\nabla f_i(x(t)) + \ddot{x}(t),$$

for all $t \in [t_0, +\infty)$.

Proof. Our proof is based on the proof of Proposition 4.6 in [16]. Rewrite $\theta(t)$ as a solution to the problem

$$\theta(t) \in \operatorname*{arg\,min}_{\theta \in \Delta^m} j(t, \theta), \quad \text{with} \quad j(t, \theta) := \frac{1}{2}\left\|\sum_{i=1}^m \theta_i \nabla f_i(x(t)) + \ddot{x}(t)\right\|^2.$$

We show that $j(\cdot, \cdot)$ is a Carathéodory integrand. Then, the proof follows from Theorem 14.37 in [204], which guarantees the existence of a measurable selection $\theta : [t_0, +\infty) \to \Delta^m$, $t \mapsto \theta(t) \in \operatorname{arg\,min}_{\theta \in \Delta^m} j(t, \theta)$. For all $t \in [t_0, +\infty)$, the function $\theta \mapsto j(t, \theta)$ is continuous. By Definition 4.5.3, $\ddot{x}(\cdot)$ is (Bochner) measurable. Then, for all $\theta \in \Delta^m$ the function $t \mapsto j(\theta, t)$ is measurable as a composition of a measurable and a continuous function. This implies that $j(\cdot, \cdot)$ is indeed a Carathéodory integrand which completes the proof. □

In the following, whenever we write $\theta(\cdot)$, we refer to the measurable function described in Lemma 4.5.8.

Lemma 4.5.9. *For* $z \in \mathcal{H}$, *define*

$$h_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto h_z(t) := \frac{1}{2}\|x(t) - z\|^2. \tag{4.73}$$

The function $h_z(\cdot)$ *is continuously differentiable and its derivative* $\dot{h}_z(\cdot)$ *is absolutely continuous. For almost all* $t \in [t_0, +\infty)$, *it holds that*

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \sum_{i=1}^m \theta_i(t)(f_i(x(t)) - f_i(z)) \leq \|\dot{x}(t)\|^2. \tag{4.74}$$

Proof. The regularity of $h_z(\cdot)$ is a consequence of Definition 4.5.3. By the chain rule, we have for almost all $t \in [t_0, +\infty)$

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

We combine these expressions with (4.72) to get

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) = \|\dot{x}(t)\|^2 + \left\langle x(t) - z, -\sum_{i=1}^m \theta_i(t)\nabla f_i(x(t))\right\rangle. \tag{4.75}$$

The objective functions f_i are convex and hence $\langle x(t) - z, \nabla f_i(x(t))\rangle \geq f_i(x(t)) - f_i(z)$ for $i = 1, \dots, m$. Using this inequality in (4.75) gives for almost all $t \in [t_0, +\infty)$

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \sum_{i=1}^m \theta_i(t)(f_i(x(t)) - f_i(z)) \leq \|\dot{x}(t)\|^2.$$

□

From Lemma 4.5.9, we derive the following relation between $h_z(\cdot)$ and $\mathcal{W}_i(\cdot)$.

Lemma 4.5.10. *Let $z \in \mathcal{H}$ and let $\mathcal{W}_i(\cdot)$ and $h_z(\cdot)$ be defined as in* (4.69) *and* (4.73)*, respectively. Then, for all $t \in [t_0, +\infty)$, it holds that*

$$\int_{t_0}^t \frac{1}{s}\sum_{i=1}^m \theta_i(s)\left(\mathcal{W}_i(s) - f_i(z)\right) ds + \frac{3}{2\alpha}\sum_{i=1}^m \theta_i(t)\left(\mathcal{W}_i(t) - f_i(z)\right) \leq C_z - \frac{1}{t}\dot{h}_z(t),$$

with $C_z := (\alpha + 1)\frac{1}{t_0^2}h_z(t_0) + \frac{3}{2\alpha}\max_{i=1,\dots,m}\left(f_i(x_0) - f_i(z)\right)$.

Proof. Adding $\frac{1}{2}\|\dot{x}(t)\|^2$ to inequality (4.74) and dividing by t, we get for almost all $t \in [t_0, +\infty)$

$$\frac{1}{t}\ddot{h}_z(t) + \frac{\alpha}{t^2}\dot{h}_z(t) + \frac{1}{t}\sum_{i=1}^m \theta_i(t)(\mathcal{W}_i(t) - f_i(z)) \leq \frac{3}{2t}\|\dot{x}(t)\|^2. \tag{4.76}$$

We reorder the terms in inequality (4.76) and integrate from t_0 to $t > t_0$, to obtain

$$\begin{aligned}&\int_{t_0}^t \frac{1}{s}\sum_{i=1}^m \theta_i(s)(\mathcal{W}_i(s) - f_i(z))ds \\ &\leq -\int_{t_0}^t \left(\frac{1}{s}\ddot{h}_z(s) + \frac{\alpha}{s^2}\dot{h}_z(s)\right) ds + \int_{t_0}^t \frac{3}{2s}\|\dot{x}(s)\|^2 ds.\end{aligned}$$

Integration by parts on the first integral on the right-hand side and using $\dot{h}_z(t_0) = 0$ gives

$$\leq -\frac{1}{t}\dot{h}_z(t) - (\alpha + 1)\int_{t_0}^t \frac{1}{s^2}\dot{h}_z(s)ds + \int_{t_0}^t \frac{3}{2s}\|\dot{x}(s)\|^2 ds. \tag{4.77}$$

By Proposition 4.5.6, we have for all $t \in [t_0, +\infty)$

$$\int_{t_0}^t \frac{3}{2s}\|\dot{x}(s)\|^2 ds \leq \frac{3}{2\alpha}\sum_{i=1}^m \theta_i(t)\left(\mathcal{W}_i(t_0) - \mathcal{W}_i(t)\right). \tag{4.78}$$

Applying inequality (4.78) to (4.77) and using $\mathcal{W}_i(t_0) = f_i(x_0)$ yields for all $t \in [t_0, +\infty)$

$$\begin{aligned} &\int_{t_0}^{t} \frac{1}{s} \sum_{i=1}^{m} \theta_i(s)(\mathcal{W}_i(s) - f_i(z))ds \\ \leq &-\frac{1}{t}\dot{h}_z(t) - (\alpha+1)\int_{t_0}^{t} \frac{1}{s^2}\dot{h}_z(s)ds + \frac{3}{2\alpha}\sum_{i=1}^{m} \theta_i(t)\left(f_i(x_0) - \mathcal{W}_i(t)\right). \end{aligned} \tag{4.79}$$

Using integration by parts one more time gives

$$\int_{t_0}^{t} \frac{1}{s^2}\dot{h}_z(s)ds = \frac{1}{t^2}h_z(t) - \frac{1}{t_0^2}h_z(t_0) + \int_{t_0}^{t} \frac{2}{s^3}h_z(s)ds \geq -\frac{1}{t_0^2}h_z(t_0). \tag{4.80}$$

Combining (4.79) and (4.80), we derive

$$\begin{aligned} &\int_{t_0}^{t} \frac{1}{s} \sum_{i=1}^{m} \theta_i(s)(\mathcal{W}_i(s) - f_i(z))ds \\ \leq &-\frac{1}{t}\dot{h}_z(t) + (\alpha+1)\frac{1}{t_0^2}h_z(t_0) + \frac{3}{2\alpha}\sum_{i=1}^{m} \theta_i(t)\left(f_i(x_0) - \mathcal{W}_i(t)\right) \\ \leq &-\frac{1}{t}\dot{h}_z(t) + (\alpha+1)\frac{1}{t_0^2}h_z(t_0) + \frac{3}{2\alpha}\sum_{i=1}^{m} \theta_i(t)\left(f_i(x_0) - f_i(z)\right) \\ &+\frac{3}{2\alpha}\sum_{i=1}^{m} \theta_i(t)\left(f_i(z) - \mathcal{W}_i(t)\right) \\ \leq &C_z - \frac{1}{t}\dot{h}_z(t) - \frac{3}{2\alpha}\sum_{i=1}^{m} \theta_i(t)\left(\mathcal{W}_i(t) - f_i(z)\right), \end{aligned}$$

with

$$C_z := (\alpha+1)\frac{1}{t_0^2}h_z(t_0) + \frac{3}{2\alpha}\max_{i=1,\dots,m}\left(f_i(x_0) - f_i(z)\right), \tag{4.81}$$

which completes the proof. □

Lemma 4.5.11. *Let $z \in \mathcal{H}$ and let $\mathcal{W}_i(\cdot), h_z(\cdot)$ and C_z be defined as in* (4.69), (4.73) *and* (4.81), *respectively. Then, for all $\tau > t_0$ it holds that*

$$\min_{i=1,\dots,m}\left(\mathcal{W}_i(\tau) - f_i(z)\right)\left[\tau\ln\tau + A\tau + B\right] \leq C_z(\tau - t_0) + \frac{h_z(t_0)}{t_0},$$

with constants $A, B \in \mathbb{R}$ which are independent of z.

Proof. Let $z \in \mathcal{H}$ and $\tau \geq t > t_0$. Proposition 4.5.6 states that the functions $\mathcal{W}_i(\cdot)$ are monotonically decreasing for all $i = 1, \dots, m$. Therefore, we have for all $s \in [t_0, t]$, that $\mathcal{W}_i(\tau) - f_i(z) \leq \mathcal{W}_i(s) - f_i(z)$ and hence

$$\begin{aligned} &\min_{i=1,\dots,m}\left(\mathcal{W}_i(\tau) - f_i(z)\right)\int_{t_0}^{t}\frac{1}{s}ds + \frac{3}{2\alpha}\min_{i=1,\dots,m}\left(\mathcal{W}_i(\tau) - f_i(z)\right) \\ \leq &\int_{t_0}^{t}\frac{1}{s}\min_{i=1,\dots,m}\left(\mathcal{W}_i(s) - f_i(z)\right)ds + \frac{3}{2\alpha}\min_{i=1,\dots,m}\left(\mathcal{W}_i(t) - f_i(z)\right). \end{aligned} \tag{4.82}$$

Using Lemma 4.5.10, we get

$$\begin{aligned}&\int_{t_0}^{t}\frac{1}{s}\min_{i=1,\dots,m}\left(\mathcal{W}_i(s)-f_i(z)\right)ds+\frac{3}{2\alpha}\min_{i=1,\dots,m}\left(\mathcal{W}_i(t)-f_i(z)\right)\\ \leq&\int_{t_0}^{t}\frac{1}{s}\sum_{i=1}^{m}\theta_i(s)\left(\mathcal{W}_i(s)-f_i(z)\right)ds+\frac{3}{2\alpha}\sum_{i=1}^{m}\theta_i(t)\left(\mathcal{W}_i(t)-f_i(z)\right)\\ \leq& C_z-\frac{1}{t}\dot{h}_z(t).\end{aligned} \tag{4.83}$$

Together, inequalities (4.82) and (4.83) give

$$\min_{i=1,\dots,m}\left(\mathcal{W}_i(\tau)-f_i(z)\right)\left[\ln t-\ln t_0+\frac{3}{2\alpha}\right]\leq C_z-\frac{1}{t}\dot{h}_z(t). \tag{4.84}$$

Integrating inequality (4.84) from $t=t_0$ to $t=\tau$, we have

$$\begin{aligned}&\min_{i=1,\dots,m}\left(\mathcal{W}_i(\tau)-f_i(z)\right)\left[\tau\ln\tau-\tau-t_0\ln t_0+t_0+\left(\frac{3}{2\alpha}-\ln t_0\right)(\tau-t_0)\right]\\ \leq& C_z(\tau-t_0)-\int_{t_0}^{\tau}\frac{1}{t}\dot{h}_z(t)dt.\end{aligned} \tag{4.85}$$

Integration by parts yields

$$\int_{t_0}^{\tau}\frac{1}{t}\dot{h}_z(t)dt=\frac{h_z(\tau)}{\tau}-\frac{h_z(t_0)}{t_0}+\int_{t_0}^{\tau}\frac{h_z(t)}{t^2}dt\geq-\frac{h_z(t_0)}{t_0}. \tag{4.86}$$

Using inequality (4.86) in (4.85), we write

$$\begin{aligned}&\min_{i=1,\dots,m}\left(\mathcal{W}_i(\tau)-f_i(z)\right)\left[\tau\ln\tau-\tau-t_0\ln t_0+t_0+\left(\frac{3}{2\alpha}-\ln t_0\right)(\tau-t_0)\right]\\ \leq& C_z(\tau-t_0)+\frac{h_z(t_0)}{t_0}.\end{aligned}$$

Introducing suitable constants $A, B\in\mathbb{R}$, this gives the desired result. □

The next theorem is the main result of this subsection. Theorem 4.5.12 states that the function values of the trajectory $F(x(t))\in\mathbb{R}^m$ converge to an element of the Pareto front. As a consequence of Theorem 4.5.12, every weak limit point of the trajectory $x(\cdot)$ is weakly Pareto optimal. This is important for proving the weak convergence of the trajectories to weakly Pareto optimal points in Subsection 4.5.4.

Theorem 4.5.12. *Define the energy function*

$$\begin{aligned}\mathcal{W}:[t_0,+\infty)\to\mathbb{R},\quad t\mapsto\mathcal{W}(t)\coloneqq&\sup_{z\in\mathcal{H}}\min_{i=1,\dots,m}\left(\mathcal{W}_i(t)-f_i(z)\right)\\ =&\sup_{z\in\mathcal{H}}\min_{i=1,\dots,m}\left(f_i(x(t))-f_i(z)\right)+\frac{1}{2}\|\dot{x}(t)\|^2\\ =&\varphi(x(t))+\frac{1}{2}\|\dot{x}(t)\|^2.\end{aligned}$$

Assume the functions f_i are bounded from below for $i=1,\dots,m$ and Assumption $(\mathcal{A}_2)$ holds. Then, $\lim_{t\to+\infty}\mathcal{W}(t)=0$. Hence, $\lim_{t\to+\infty}\|\dot{x}(t)\|=0$ and $\lim_{t\to+\infty}\varphi(x(t))=0$ and by Theorem 2.3.13 every weak limit point of $x(\cdot)$ is Pareto critical.

Proof. Lemma 4.5.11 states for all $\tau > t_0$

$$\min_{i=1,\dots,m} (\mathcal{W}_i(\tau) - f_i(z)) [\tau \ln \tau + A\tau + B] \leq C_z(\tau - t_0) + \frac{h_z(t_0)}{t_0}. \tag{4.87}$$

We cannot directly take the supremum on both sides since C_z might be unbounded w.r.t. $z \in \mathcal{H}$. For $z \in \mathcal{L}(F, F(x_0))$ we have $\max_{i=1,\dots,m} (f_i(x_0) - f_i(z)) \leq \max_{i=1,\dots,m} (f_i(x_0) - \inf_{z\in\mathcal{H}} f_i(z)) =: M$. Since all f_i are bounded from below by assumption, we have $M < +\infty$. Fix $F^* \in F(\mathcal{LP}_w(F, F(x_0)))$. Using the definition of C_z given in (4.81), we get for all $z \in F^{-1}(\{F^*\})$

$$\begin{aligned} C_z(\tau - t_0) + \frac{h_z(t_0)}{t_0} &\leq \left((\alpha+1)\frac{1}{t_0^2} h_z(t_0) + \frac{3}{2\alpha} M \right)(\tau - t_0) + \frac{h_z(t_0)}{t_0} \\ &= \left(\frac{\alpha+1}{t_0^2}(\tau - t_0) + \frac{1}{t_0} \right) h_z(t_0) + \frac{3}{2\alpha} M(\tau - t_0). \end{aligned} \tag{4.88}$$

By Assumption $(\mathcal{A}_2)$ $\sup_{F^*\in F(\mathcal{LP}_w(F,F(x_0)))} \inf_{z\in F^{-1}(\{F^*\})} h_z(t_0) = R < +\infty$. Applying this infimum and supremum to (4.88) we have

$$\begin{aligned} &\sup_{F^*\in F(\mathcal{LP}_w(F,F(x_0)))} \inf_{z\in F^{-1}(\{F^*\})} \left(C_z(\tau - t_0) + \frac{h_z(t_0)}{t_0} \right) \\ &\leq \left(\frac{(\alpha+1)R}{t_0^2} + \frac{3}{2\alpha} M \right)(\tau - t_0) + \frac{R}{t_0}. \end{aligned} \tag{4.89}$$

Combining Lemma 2.3.15 with (4.87) and (4.89), we get for all $\tau > t_0$

$$\mathcal{W}(\tau) [\tau \ln \tau + A\tau + B] \leq C\tau + D, \tag{4.90}$$

with $A, B, D \in \mathbb{R}$ and $C > 0$. Since $\mathcal{W}(t)$ is nonnegative, $\lim_{t\to+\infty} \mathcal{W}(t) = 0$ holds. □

Remark 4.5.13. *From the proof of Theorem 4.5.12, we can deduce a slightly stronger result. There is not only convergence* $\lim_{t\to+\infty} \mathcal{W}(t) = 0$*, but from inequality* (4.90) *we get convergence of order*

$$\mathcal{W}(t) = \mathcal{O}\left(\frac{1}{\ln t}\right), \quad \textit{as} \quad t \to +\infty.$$

Since this is a rather slow rate of convergence, which is not used in the following proofs, we do not point it out in Theorem 4.5.12.

We can derive some additional facts on the function values $f_i(x(t))$ along the trajectories from Theorem 4.5.12.

Theorem 4.5.14. *Assume all assumptions of Theorem 4.5.12 are met. Then, for all* $i = 1, \dots, m$

$$\lim_{t\to\infty} f_i(x(t)) = f_i^\infty \in \mathbb{R}$$

exists.

Proof. Theorem 4.5.12 states $\lim_{t\to\infty} \mathcal{W}(t) = 0$. By definition, we have $\mathcal{W}(t) = \varphi(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2$. Theorem 2.3.13 guarantees $\varphi(x(t)) \geq 0$ for all $t \geq t_0$ and obviously $\frac{1}{2}\|\dot{x}(t)\|^2 \geq 0$. Then, from $\lim_{t\to+\infty} \mathcal{W}(t) = 0$ it follows that $\lim_{t\to+\infty} \frac{1}{2}\|\dot{x}(t)\|^2 = 0$. By Proposition 4.5.6 $\lim_{t\to+\infty} \mathcal{W}_i(t) = \lim_{t\to+\infty} f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2$ exists and hence limit $\lim_{t\to+\infty} f_i(x(t))$ exists, which completes the proof. □

4.5.4 Asymptotic analysis

For $\alpha \geq 3$, we prove fast convergence for the function values with rate $\varphi(x(t)) = \mathcal{O}(t^{-2})$ as $t \to +\infty$, as we show in Theorem 4.5.18. In Theorem 4.5.20, we prove that for $\alpha > 3$ the trajectories $x(\cdot)$ of (MAVD) converge weakly to weakly Pareto optimal points using Opial's Lemma.

Fast convergence of function values

In this part, we show that solutions of (MAVD) have good properties with respect to multiobjective optimization. Along the trajectories of (MAVD) the function values converge with order $\mathcal{O}(t^{-2})$ to an optimal value, given $\alpha \geq 3$. The convergence has to be understood in terms of the merit function $\varphi(\cdot)$ which is introduced in Subsection 2.3.3. We prove this result using Lyapunov type energy functions similar to the analysis of the scalar case laid out in [13, 218]. To this end, we introduce two important auxiliary functions in Definition 4.5.15 and discuss their basic properties in the following lemmas. The main result of this subsection on the convergence of the function values is stated in Theorem 4.5.18.

Definition 4.5.15. *Let $\lambda \geq 0$, $\xi \geq 0$, $z \in \mathcal{H}$ and $x : [t_0, +\infty) \to \mathcal{H}$ be a solution to* (MAVD) *in the sense of Definition 4.5.3. Define for all $i = 1, \dots, m$ the component-wise energy functions*

$$\begin{aligned} \mathcal{E}_{i,\lambda,\xi,z} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}_{i,\lambda,\xi,z}(t) :=& t^2 (f_i(x(t)) - f_i(z)) + \frac{1}{2}\|\lambda(x(t) - z) + t\dot{x}(t)\|^2 \\ &+ \frac{\xi}{2}\|x(t) - z\|^2. \end{aligned} \tag{4.91}$$

Using the functions $\mathcal{E}_{i,\lambda,\xi,z}(\cdot)$, define the energy function

$$\begin{aligned} \mathcal{E}_{\lambda,\xi,z} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}_{\lambda,\xi,z}(t) :=& \min_{i=1,\dots,m} \mathcal{E}_{i,\lambda,\xi,z}(t) = t^2 \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) \\ &+ \frac{1}{2}\|\lambda(x(t) - z) + t\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t) - z\|^2. \end{aligned} \tag{4.92}$$

Lemma 4.5.16. *Let $\lambda > 0$, $\alpha \geq \lambda + 1$, fix $\xi^* = \lambda(\alpha - 1 - \lambda) > 0$ and let $z \in \mathcal{H}$. Then, for all $i = 1, \dots, m$ and almost all $t \in [t_0, +\infty)$*

$$\frac{d}{dt}\mathcal{E}_{i,\lambda,\xi^*,z}(t) \leq 2t(f_i(x(t)) - f_i(z)) - t\lambda \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) + t(\lambda + 1 - \alpha)\|\dot{x}(t)\|^2.$$

Proof. The function $\mathcal{E}_{i,\lambda,\xi^*,z}(\cdot)$ is differentiable almost everywhere since f_i is continuously differentiable and $x(\cdot)$ is continuously differentiable with absolutely continuous derivative $\dot{x}(\cdot)$ by Definition 4.5.3. Compute $\frac{d}{dt}\mathcal{E}_{i,\lambda,\xi^*,z}(t)$ using the chain rule on (4.91)

$$\begin{aligned} \frac{d}{dt}\mathcal{E}_{i,\lambda,\xi^*,z}(t) &= 2t(f_i(x(t)) - f_i(z)) + t^2 \left\langle \dot{x}(t), \nabla f_i(x(t)) + \ddot{x}(t) \right\rangle \\ &+ t \left\langle x(t) - z, \frac{\lambda(\lambda + 1) + \xi^*}{t}\dot{x}(t) + \lambda\ddot{x}(t) \right\rangle + t(\lambda + 1)\|\dot{x}(t)\|^2. \end{aligned} \tag{4.93}$$

Using Proposition 4.5.6 on the second summand in (4.93), we bound this by

$$\begin{aligned} \leq & 2t(f_i(x(t)) - f_i(z)) + t\left\langle x(t) - z, \frac{\lambda(\lambda+1)+\xi^*}{t}\dot{x}(t) + \lambda\ddot{x}(t)\right\rangle \\ & + t(\lambda+1-\alpha)\|\dot{x}(t)\|^2. \end{aligned} \tag{4.94}$$

We rewrite (4.94) into

$$= 2t(f_i(x(t)) - f_i(z)) + t\lambda\left\langle x(t) - z, \frac{\alpha}{t}\dot{x}(t) + \ddot{x}(t)\right\rangle + t(\lambda+1-\alpha)\|\dot{x}(t)\|^2,$$

using $\lambda(\lambda+1) + \xi^* = \lambda\alpha$. The definition of (MAVD) together with Lemma 4.5.8 implies

$$= 2t(f_i(x(t)) - f_i(z)) - t\lambda\sum_{i=1}^{m}\theta_i(t)\left\langle x(t) - z, \nabla f_i(x(t))\right\rangle + t(\lambda+1-\alpha)\|\dot{x}(t)\|^2.$$

The objective functions f_i are convex and hence $f_i(z) - f_i(x(t)) \geq \langle \nabla f_i(x(t)), z - x(t)\rangle$ and therefore,

$$\leq 2t(f_i(x(t)) - f_i(z)) - t\lambda\sum_{i=1}^{m}\theta_i(t)\left(f_i(x(t)) - f_i(z)\right) + t(\lambda+1-\alpha)\|\dot{x}(t)\|^2.$$

We bound the convex combination using the minimum to get

$$\leq 2t(f_i(x(t)) - f_i(z)) - t\lambda\min_{i=1,\dots,m}\left(f_i(x(t)) - f_i(z)\right) + t(\lambda+1-\alpha)\|\dot{x}(t)\|^2.$$

□

To retrieve a result similar to Lemma 4.5.16 for the function $\mathcal{E}_{\lambda,\xi,z}(\cdot)$ defined in (4.92), we use Lemma 2.2.14 which helps us to treat the derivative of $\mathcal{E}_{\lambda,\xi,z}(\cdot)$.

Lemma 4.5.17. *Let $\lambda > 0$, $\alpha \geq \lambda + 1$, fix $\xi^* = \lambda(\alpha - 1 - \lambda) > 0$ and let $z \in \mathcal{H}$. The energy function $\mathcal{E}_{\lambda,\xi^*,z}(\cdot)$ satisfies the following conditions:*

i) The function $\mathcal{E}_{\lambda,\xi^*,z}(\cdot)$ is differentiable in almost all $t \in [t_0, +\infty)$;

ii) For almost all $t \in [t_0, +\infty)$, it holds that

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi^*,z}(t) \leq (2-\lambda)t\min_{i=1,\dots,m}\left(f_i(x(t)) - f_i(z)\right) - (\alpha - \lambda - 1)t\|\dot{x}(t)\|^2; \tag{4.95}$$

iii) For all $t \in [t_0, +\infty)$, it holds that

$$\begin{aligned} \mathcal{E}_{\lambda,\xi^*,z}(t) - \mathcal{E}_{\lambda,\xi^*,z}(t_0) \leq & (2-\lambda)\int_{t_0}^{t} t\min_{i=1,\dots,m}\left(f_i(x(t)) - f_i(z)\right)dt \\ & + \int_{t_0}^{t} t(\lambda+1-\alpha)\|\dot{x}(t)\|^2 dt. \end{aligned}$$

Proof. i) The functions $t \mapsto f_i(x(t))$ are continuously differentiable for all $i = 1, \dots, m$. Then, by Lemma 2.2.14 the function $t \mapsto \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z))$ is differentiable in t for almost all $t \in [t_0, +\infty)$. Since $x(\cdot)$ is a solution to (MAVD) in the sense of Definition 4.5.3, we know that $\|\lambda(x(t) - z) + t\dot{x}(t)\|^2$ and $\frac{\xi}{2}\|x(t) - z\|^2$ are differentiable in t for almost all $t \in [t_0, +\infty)$. In total we get that $\mathcal{E}_{\lambda,\xi^*,z}(t)$ is differentiable in t for almost all $t \in [t_0, +\infty)$.

ii) We need the derivative of $\min_{i=1,\dots,m} (f_i(x(t)) - f_i(z))$ in order to compute the derivative of $\mathcal{E}_{\lambda,\xi^*,z}(t)$. By Lemma 2.2.14 for almost all $t \in [t_0, +\infty)$ there exists $j \in \{1, \dots, m\}$ with

$$\begin{aligned} \frac{d}{dt} \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) &= \frac{d}{dt} (f_j(x(t)) - f_j(z)), \text{ and} \\ \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) &= f_j(x(t)) - f_j(z). \end{aligned} \tag{4.96}$$

For the remainder of the proof fix $t \in [t_0, +\infty)$ and $j \in \{1, \dots, m\}$ satisfying equation (4.96). From the first part of (4.96), we immediately get

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi^*,z}(t) = \frac{d}{dt}\mathcal{E}_{j,\lambda,\xi^*}(t). \tag{4.97}$$

Applying Lemma 4.5.16, we bound (4.97) by

$$\leq 2t(f_j(x(t)) - f_j(z)) - t\lambda \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) + t(\lambda + 1 - \alpha)\|\dot{x}(t)\|^2.$$

Then, the second equation in (4.96) gives

$$=(2 - \lambda)t \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) + t(\lambda + 1 - \alpha)\|\dot{x}(t)\|^2.$$

Statement *iii)* follows immediately from *ii)* by integrating inequality (4.95) from t_0 to $t > t_0$. □

The term $\min_{i=1,\dots,m} (f_i(x(t)) - f_i(z))$ will not remain nonnegative in general. Hence, we cannot guarantee that $\mathcal{E}_{\lambda,\xi^*,z}(t)$ is nonnegative. Therefore, the function $\mathcal{E}_{\lambda,\xi^*,z}(t)$ is not suitable for convergence analysis and we cannot directly retrieve results on the convergence rates. We are still able to get convergence results using Lemma 2.3.15.

Theorem 4.5.18. *Let $\alpha \geq 3$ and assume Assumption ($\mathcal{A}_2$) holds. Then*

$$t^2\varphi(x(t)) \leq t_0^2\varphi(x_0) + 2(\alpha - 1)R + (3 - \alpha) \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds,$$

and hence $\varphi(x(t)) \leq \frac{t_0^2\varphi(x_0)+2(\alpha-1)R}{t^2}$ for all $t \in [t_0, +\infty)$.

Proof. We consider the energy function $\mathcal{E}_{\lambda,\xi^*,z}(\cdot)$ with parameter $\lambda = 2$. From the definition of $\mathcal{E}_{2,\xi^*,z}(\cdot)$ and part *iii)* of Lemma 4.5.17, we deduce

$$t^2 \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) \leq \mathcal{E}_{2,\xi^*,z}(t_0) + (3 - \alpha) \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds.$$

Writing out the definition of $\mathcal{E}_{2,\xi^*,z}(t_0)$ and using $\lambda = 2$ and $\xi^* = \lambda(\alpha - 1 - \lambda) = 2(\alpha - 3)$, we have

$$\begin{aligned} t^2 \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z)) \leq & t_0^2 \min_{i=1,\dots,m} (f_i(x_0) - f_i(z)) + (\alpha - 1)\|x_0 - z\|^2 \\ & + (3 - \alpha) \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds. \end{aligned} \tag{4.98}$$

We want to apply the supremum and infimum in accordance with Lemma 2.3.15. Let $F^* = (f_1^*, \dots, f_m^*) \in F(\mathcal{LP}_w(F, F(x_0)))$, then

$$\begin{aligned} & \inf_{z \in F^{-1}(\{F^*\})} \left[t_0^2 \min_{i=1,\dots,m} (f_i(x_0) - f_i(z)) + (\alpha - 1)\|x_0 - z\|^2 \right] \\ = & t_0^2 \min_{i=1,\dots,m} (f_i(x_0) - f_i^*) + (\alpha - 1) \inf_{z \in F^{-1}(\{F^*\})} \|x_0 - z\|^2. \end{aligned} \tag{4.99}$$

Now, we can apply the supremum to inequality (4.99) and get

$$\begin{aligned} & \sup_{F^* \in \mathcal{LP}_w(F, F(x_0))} \inf_{z \in F^{-1}(\{F^*\})} \left[t_0^2 \min_{i=1,\dots,m} (f_i(x_0) - f_i(z)) + (\alpha - 1)\|x_0 - z\|^2 \right] \\ \leq & t_0^2 \sup_{F^* \in \mathcal{LP}_w(F, F(x_0))} \inf_{z \in F^{-1}(\{F^*\})} \min_{i=1,\dots,m} (f_i(x_0) - f_i^*) \\ & + (\alpha - 1) \sup_{F^* \in \mathcal{LP}_w(F, F(x_0))} \inf_{z \in F^{-1}(\{F^*\})} \inf_{z \in F^{-1}(\{F^*\})} \|x_0 - z\|^2. \end{aligned} \tag{4.100}$$

By Assumption $(\mathcal{A}_2)$ and the definition of $\varphi(\cdot)$, this is equal to

$$= t_0^2 \varphi(x_0) + 2(\alpha - 1)R. \tag{4.101}$$

Now, by applying $\sup_{F^* \in \mathcal{LP}_w(F, F(x_0))} \inf_{z \in F^{-1}(\{F^*\})}$ to $t^2 \min_{i=1,\dots,m} (f_i(x(t)) - f_i(z))$ and using (4.98) - (4.101), we get

$$t^2 \varphi(x(t)) \leq t_0^2 \varphi(x_0) + 2(\alpha - 1)R + (3 - \alpha) \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds.$$

□

Corollary 4.5.19. *Let $\alpha > 3$ and assume Assumption $(\mathcal{A}_2)$ holds. Then*

$$\int_{t_0}^{+\infty} s\|\dot{x}(s)\|^2 ds < +\infty,$$

i.e., $(t \mapsto t\|\dot{x}(t)\|^2) \in L^1([t_0, +\infty), \mathbb{R})$.

Weak convergence of trajectories

In this part, we show that bounded solutions of (MAVD) converge weakly to weakly Pareto optimal points of (MOP), given $\alpha > 3$. We prove this in Theorem 4.5.20 using Opial's Lemma (Lemma 2.1.6). Since we need to apply Theorem 4.5.12 and Theorem 4.5.14, we assume in this subsection that the functions f_i are bounded from below for all $i = 1, \dots, m$ and that Assumption $(\mathcal{A}_2)$ holds. In order to utilize Opial's Lemma we need Lemma 2.2.12 .

Theorem 4.5.20. *Let* $\alpha > 3$ *and let* $x : [t_0, +\infty) \to \mathcal{H}$ *be a bounded solution to* (MAVD). *Assume that the functions* f_i *are bounded from below and that Assumption* $(\mathcal{A}_2)$ *holds. Then,* $x(t)$ *converges weakly to a weakly Pareto optimal point of* (MOP).

Proof. Define the set

$$S \coloneqq \{z \in \mathcal{H} : f_i(z) \leq f_i^\infty \text{ for all } i = 1, \ldots, m\},$$

where $f_i^\infty = \lim_{t \to +\infty} f_i(x(t))$. This limit exists due to Theorem 4.5.14. Since $x(t)$ is bounded it posses a weak sequential cluster point $x^\infty \in \mathcal{H}$. Hence, there exists a sequence $(x(t_k))_{k \geq 0}$ with $t_k \to +\infty$ and $x(t_k) \rightharpoonup x^\infty$ as $k \to +\infty$. Because the objective functions are lower semicontinuous in the weak topology, we get for all $i = 1, \ldots, m$

$$f_i(x^\infty) \leq \liminf_{k \to +\infty} f_i(x(t_k)) = \lim_{k \to +\infty} f_i(x(t_k)) = f_i^\infty.$$

Therefore, we can conclude that $x^\infty \in S$. Hence, S is nonempty and each weak sequenial cluster point of $x(t)$ belongs to S. Let $z \in S$ and define

$$h_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto h_z(t) \coloneqq \frac{1}{2}\|x(t) - z\|^2.$$

By Definition 4.5.3, the function $h_z(\cdot)$ is continuously differentiable with absolutely continuous derivative $\dot{h}_z(\cdot)$. The first and second derivative of $h_z(t)$ is given by

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

for almost all $t \in [t_0, +\infty)$. Multiplying $\dot{h}_z(t)$ with $\frac{\alpha}{t}$ and adding it to $\ddot{h}_z(t)$ gives

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) = \left\langle x(t) - z, \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t)\right\rangle + \|\dot{x}(t)\|^2. \tag{4.102}$$

Using the equation (MAVD) together with the weights $\theta(t) \in \Delta^m$ from Lemma 4.5.8, we get from (4.102) the equation

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) = \sum_{i=1}^{m} \theta_i(t) \langle z - x(t), \nabla f_i(x(t))\rangle + \|\dot{x}(t)\|^2. \tag{4.103}$$

We want to bound the inner products $\langle z - x(t), \nabla f_i(x(t))\rangle$. Since $\mathcal{W}_i(\cdot)$ is monotonically decreasing by Proposition 4.5.6 and converging to f_i^∞ by Theorem 4.5.14, we get

$$f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2 = \mathcal{W}_i(t) \geq f_i^\infty, \tag{4.104}$$

for all $i = 1, \ldots, m$. From $z \in S$ and the convexity of the functions f_i, we conclude for all $i = 1, \ldots, m$

$$f_i^\infty \geq f_i(z) \geq f_i(x(t)) + \langle \nabla f_i(x(t)), z - x(t)\rangle. \tag{4.105}$$

Together, (4.104) and (4.105) imply

$$\langle \nabla f_i(x(t)), z - x(t)\rangle \leq \frac{1}{2}\|\dot{x}(t)\|^2, \tag{4.106}$$

for all $i = 1, \dots, m$. Now, we combine (4.103) and (4.106) and multiply with t, to conclude

$$t\ddot{h}_z(t) + \alpha\dot{h}_z(t) \leq \frac{3t}{2}\|\dot{x}(t)\|^2. \tag{4.107}$$

Theorem 4.5.18 states that $(t \mapsto t\|\dot{x}(t)\|^2) \in L^1([t_0, +\infty), \mathbb{R})$ for $\alpha > 3$. Then, Lemma 2.2.12 applied to equation (4.107) guarantees that $h_z(t)$ converges and by Opial's Lemma (Lemma 2.1.6) we conclude that $x(t)$ converges weakly to an element in S. By Theorem 4.5.12, we know that every weak accumulation point of $x(t)$ is weakly Pareto optimal. □

4.5.5 Numerical experiments

In this subsection, we conduct numerical experiments to verify the convergence rates we prove in the previous subsection. In particular, we show that the convergence of $\varphi(x(t))$ with rate $\mathcal{O}(t^{-2})$ as stated in Theorem 4.5.18 holds. Since we cannot calculate analytical solutions to (MAVD) for a general multiobjective optimization problem in closed form, we compute the approximation to a solution $x(\cdot)$ using a discretization. We do not discuss the quality of the discretization we use. For all experiments we use initial time $t_0 = 1$, set a fixed initial state $x(t_0) = x_0$ and use initial velocity $\dot{x}(t_0) = 0$. We use equidistant time steps $t_k = t_0 + kh$, with $h = 1\mathrm{e}{-3}$. We use the scheme $x(t_k) \approx x^k$, $\dot{x}(t_k) \approx \frac{x^{k+1}-x^k}{h}$ and $\ddot{x}(t_k) \approx \frac{x^{k+1}-2x^k+x^{k-1}}{h^2}$ to compute the discretization $(x^k)_{k\geq 0}$ of the trajectory $x(\cdot)$ for 100 000 time steps. We look at two examples with instances of the multiobjective optimization problem (MOP). Both problem instances use two convex and smooth objective functions $f_i : \mathbb{R}^2 \to \mathbb{R}$ for $i = 1, 2$. In Subsection 4.5.5, we look at a quadratic multiobjective optimization problem and in Subsection 4.5.5, we consider a convex optimization problem with objective functions that are not strongly convex. For both examples we plot approximations of the solution $x(\cdot)$ and plot the function $\varphi(x(t))$ to show that the inequality $\varphi(x(t)) \leq \frac{t_0^2\varphi(x_0)+2(\alpha-1)R}{t^2}$ holds for $t \geq t_0$. To compute $\varphi(x(t))$ we have to solve the optimization problem $\varphi(x^k) = \sup_{z\in\mathcal{H}} \min_{i=1,\dots,m} f_i(x^k) - f_i(z)$ for every of the 100 000 iterations with adequate accuracy. Therefore, we restrict ourselves to problems where the Pareto set of (MOP) can be explicitly computed. For these problems $\varphi(\cdot)$ can be evaluated more efficiently using Lemma 2.3.15.

A quadratic multiobjective optimization problem

We begin with an instance of (MOP) with two quadratic objective functions

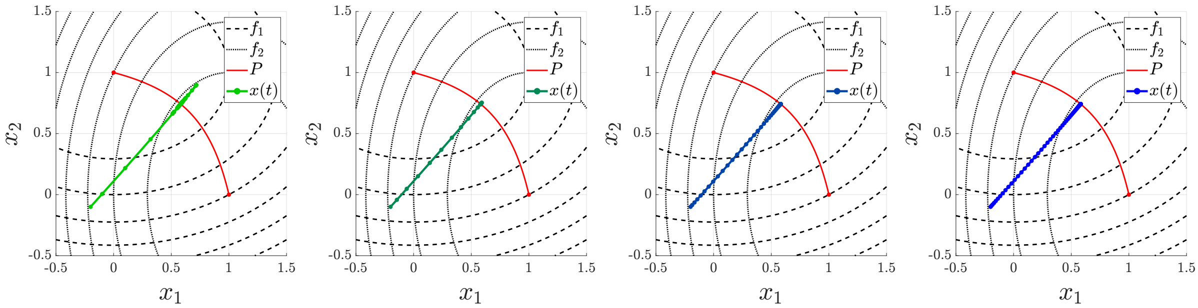
$$f_i : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \frac{1}{2}(x - x^i)^\top Q_i (x - x^i),$$

for $i = 1, 2$, given matrices and vectors

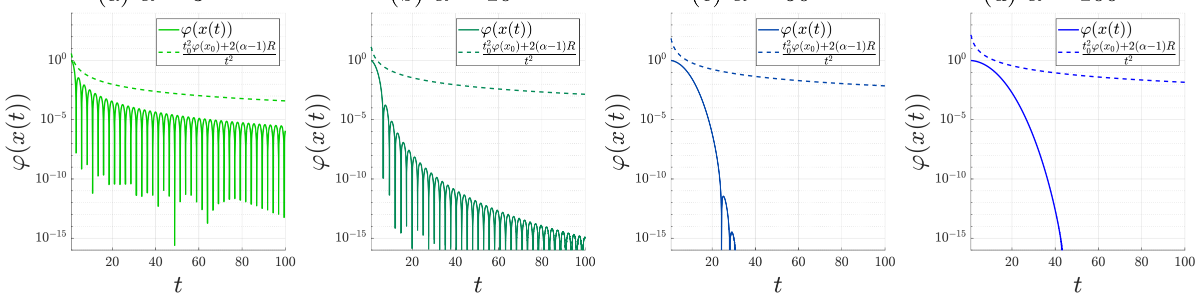
$$Q_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad x^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad x^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

For this problem the Pareto set is

$$\mathcal{P} = \left\{ x \in \mathbb{R}^2 : x = \begin{pmatrix} 2\lambda/(1+\lambda) \\ 2(1-\lambda)/(2-\lambda) \end{pmatrix}, \text{ for } \lambda \in [0,1] \right\}.$$



(a) $\alpha = 3$ (b) $\alpha = 10$ (c) $\alpha = 50$ (d) $\alpha = 100$



(e) $\alpha = 3$ (f) $\alpha = 10$ (g) $\alpha = 50$ (h) $\alpha = 100$

Figure 4.5: Trajectories $x(\cdot)$ and inequalities $\varphi(x(t)) \leq \frac{t_0^2 \varphi(x_0) + 2(\alpha-1)R}{t^2}$ for different values of $\alpha \in \{3, 10, 50, 100\}$.

In our first experiment, we use the initial value $x_0 = (-.2, -.1)^\top$. We compute an approximation of a solution to (MAVD) for different values of $\alpha \in \{3, 10, 50, 100\}$ as described in the introduction of this subsection. The results can be seen in Figure 4.5. Subfigures 4.5a - 4.5d contain plots of the trajectories $x(\cdot)$ for different values of α. In the plots of the trajectories we added a circle every 500 iterations to visualize the velocities. In Subfigures 4.5e - 4.5h the values of $\varphi(x(t))$ and the bounds $\frac{t_0^2 \varphi(x_0) + 2(\alpha-1)R}{t^2}$ for different values of α are shown. The inequality $\varphi(x(t)) \leq \frac{t_0^2 \varphi(x_0) + 2(\alpha-1)R}{t^2}$ holds for each value of α. For the smallest value of $\alpha = 3$, we see a large number of oscillations in the trajectory and in the values of $\varphi(x(t))$, respectively. This behavior is typical for systems with asymptotic vanishing damping. For larger values of α, we observe fewer oscillations and see improved convergence rates, with slower movement in the beginning due to the high friction. These phenomena are consistent with the observations made in the singleobjective setting.

A nonquadratic multiobjective optimization problem

In our second example, we consider an instance of problem (MOP) with two objective functions

$$f_i : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \log\left(\sum_{j=1}^{p} \exp\left(\left(a_j^{(i)}\right)^\top x - b_j^{(i)} \right) \right), \tag{4.108}$$

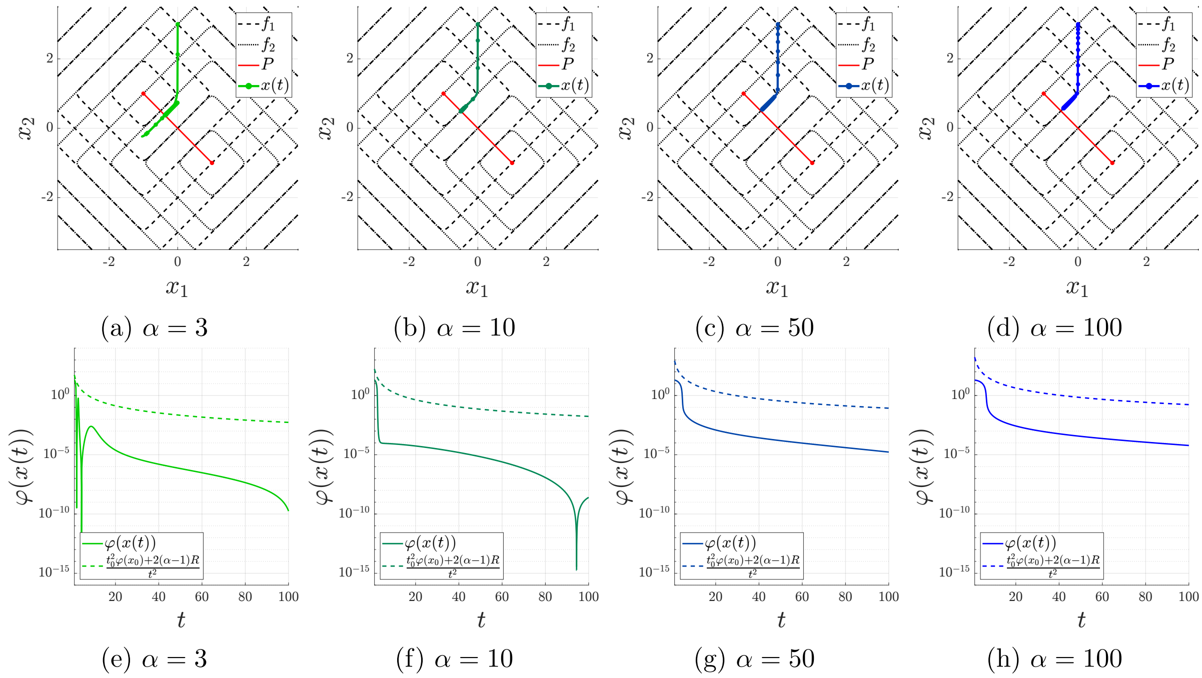
for $i = 1, 2$, $p = 4$ and given matrices and vectors

$$A^{(1)} = \begin{pmatrix} \left(a_1^{(1)}\right)^\top \\ \vdots \\ \left(a_4^{(1)}\right)^\top \end{pmatrix} = \begin{pmatrix} 10 & 10 \\ 10 & -10 \\ -10 & -10 \\ -10 & 10 \end{pmatrix}, \quad b^{(1)} = \begin{pmatrix} 0 \\ -20 \\ 0 \\ 20 \end{pmatrix},$$

$$A^{(2)} = \begin{pmatrix} \left(a_1^{(2)}\right)^\top \\ \vdots \\ \left(a_4^{(2)}\right)^\top \end{pmatrix} = \begin{pmatrix} 10 & 10 \\ 10 & -10 \\ -10 & -10 \\ -10 & 10 \end{pmatrix}, \quad b^{(2)} = \begin{pmatrix} 0 \\ 20 \\ 0 \\ -20 \end{pmatrix}.$$

The objective functions given by (4.108) are convex but not strongly convex. Taking advantage of the symmetry in the objective functions f_i, the Pareto set $\mathcal{P}$ can be explicitly computed as

$$\mathcal{P} = \left\{ x \in \mathbb{R}^2 : x = \begin{pmatrix} -1 + 2\lambda \\ 1 - 2\lambda \end{pmatrix}, \text{ for } \lambda \in [0, 1] \right\}.$$



(a) $\alpha = 3$ (b) $\alpha = 10$ (c) $\alpha = 50$ (d) $\alpha = 100$

(e) $\alpha = 3$ (f) $\alpha = 10$ (g) $\alpha = 50$ (h) $\alpha = 100$

Figure 4.6: Trajectories $x(\cdot)$ and inequalities $\varphi(x(t)) \leq \frac{t_0^2\varphi(x_0)+2(\alpha-1)R}{t^2}$ for different values of $\alpha \in \{3, 10, 50, 100\}$.

We choose the initial value $x_0 = (0, 3)^\top$ and compute an approximate solution to (MAVD) as described in the beginning of this subsection. Analogous to the last example, we present the results of the computations in Figure 4.6. Again, Subfigures 4.6a - 4.6d contain plots of the trajectories and Subfigures 4.6e - 4.6h contain the values of the merit function $\varphi(x(t))$. We observe results similar to the example in Subsection 4.5.5. Since the objective functions given in (4.108) are not strongly convex, we experience slower convergence. Once more, we see for small values of α oscillations in the trajectory $x(\cdot)$ and the merit function values $\varphi(x(t))$ introduced by the inertia in the system (MAVD). Larger values of α correspond to higher friction in the beginning and we therefore experience slower convergence for the time interval we consider. Oscillations can only be seen for $\alpha = 3$ and close to the end for $\alpha = 10$. The slower convergence in this example is expected due to the lack of strong convexity.

4.6 The multiobjective Tikhonov regularized inertial gradient system (MTRIGS)

In this section we study the *multiobjective Tikhonov regularized inertial gradient system*

$$\text{(MTRIGS)} \qquad \frac{\alpha}{t^q}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t)}(0) = 0, \quad \text{for} \quad t > t_0,$$

with α, $\beta > 0$, $q \in (0,1]$, $p \in (0,2]$ and $C(x) \coloneqq \operatorname{conv}(\{\nabla f_i(x) : i = 1,\dots,m\})$, with initial data $t_0 > 0$, $x(t_0) = x_0 \in \mathcal{H}$ and $\dot{x}(t_0) = v_0 \in \mathcal{H}$. In the case of scalar optimization ($m = 1$), the system (MTRIGS) reduces to the *Tikhonov regularized inertial gradient system*

$$\text{(TRIGS)} \qquad \ddot{x}(t) + \frac{\alpha}{t^q}\dot{x}(t) + \nabla f(x(t)) + \frac{\beta}{t^p}x(t) = 0,$$

which is extensively studied in the literature [12, 21, 146]. Assuming that $\arg\min_{x\in\mathcal{H}} f(x)$ is not empty, if, for instance, $p \in (0,2)$, $q \in (0,1)$ and $p < q+1$, then for the trajectory solution $x(\cdot)$ of (TRIGS) it holds $f(x(t)) - \min_{x\in\mathcal{H}} f(x) = \mathcal{O}(t^{-p})$ as $t \to +\infty$. Thus, a convergence rate arbitrary close to $\mathcal{O}\left(t^{-2}\right)$ can be obtained. Additionally, the trajectory solution converges strongly to the element with minimum norm in $\arg\min_{x\in\mathcal{H}} f(x)$, that is, $x(t) \to \operatorname{proj}_{\arg\min_{x\in\mathcal{H}} f(x)}(0)$ as $t \to +\infty$. On the other hand, (MTRIGS) is related to the *multiobjective gradient system with asymptotic vanishing damping*

$$\text{(MAVD)} \qquad \frac{\alpha}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0,$$

with $\alpha \geq 3$, which we discuss in Section 4.5. We have shown fast convergence of the function values along the trajectory solution, namely, $\varphi(x(t)) = \mathcal{O}(t^{-2})$ as $t \to +\infty$. In addition, for $\alpha > 3$, the trajectory solutions $x(\cdot)$ of (MAVD) converge weakly to a weakly Pareto optimal point of (MOP). In the scalar case, when $m = 1$ and $f := f_1$, the system (MAVD) reduces to the celebrated *inertial gradient system with asymptotic vanishing damping*

$$\text{(AVD)} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0,$$

which was introduced in [218] as the continuous counterpart of Nesterov's accelerated gradient method [182]. The system (AVD) has further been studied in several papers, including [13, 58, 59, 166]. It holds that $f(x(t)) - \min_{x\in\mathcal{H}} f = \mathcal{O}(t^{-2})$ as $t \to +\infty$ and, for $\alpha > 3$, the trajectory solutions converge weakly to a global minimizer of f, provided that $\arg\min_{x\in\mathcal{H}} f(x)$ is not empty. Due to its convergence properties, (MAVD) is the natural counterpart of (AVD) when considering multiobjective optimization problems. The dynamical system (TRIGS) enhances the asymptotic properties of (AVD) by ensuring, depending on the chosen parameters α, β, p and q, weak and even strong convergence of the trajectory to the minimum norm solution, while retaining the rapid convergence of function values. The dynamical system (MTRIGS) provides a similar improvement over (MAVD) in the context of multiobjective optimization. The main results regarding the asymptotic behavior of (MTRIGS) obtained in this section are summarized in Table 4.1. In principal, we obtain convergence rates for the function values which can be arbitrarily close to $\mathcal{O}(t^{-2})$ as $t \to +\infty$. Furthermore, for $p \in (0,2)$, $q \in (0,1)$ and $p < q+1$ the trajectory solution $x(\cdot)$ converges strongly to a weakly Pareto optimal point which has the minimal norm in the set $\mathcal{L}(F, F^\infty) \subseteq \mathcal{P}_w$, with $F^\infty := \lim_{t\to+\infty} F(x(t))$. For $p \in (0,2)$, $q \in (0,1)$ and $p > q+1$, we show that the trajectory converges weakly to a weakly Pareto optimal point. The case $p = q+1$ is critical, as it seems that convergence results for the trajectories cannot be obtained. In addition, we treat some boundary cases for the parameters p and q, which require additional conditions on the parameters α and β.

Conditions on p, q, α, β	$\varphi(x(t))$	$\|\dot{x}(t)\|$	$\|x(t) - z(t)\|$	$x(t)$
$p \in (0,2]$, $2q < p$	$\mathcal{O}\left(t^{-2q}\right)$	$\mathcal{O}\left(t^{-q}\right)$	$\mathcal{O}(1)$	-
$q \in (0,1)$, $p < q+1$	$\mathcal{O}\left(t^{-p}\right)$	$\mathcal{O}\left(t^{\frac{\max(q,p-q)-(p+1)}{2}}\right)$	$\mathcal{O}\left(t^{\frac{\max(q,p-q)-1}{2}}\right)$	strong convergence
$q = 1$, $\alpha \geq 3$	$\mathcal{O}\left(t^{-p}\right)$	$\mathcal{O}\left(t^{-\frac{p}{2}}\right)$	$\mathcal{O}(1)$	-
$p \in (0,2)$, $q+1 < p$	$\mathcal{O}\left(t^{-2q}\right)$	$\mathcal{O}\left(t^{-q}\right)$, $\int_{t_0}^{+\infty} s\|\dot{x}(s)\|^2 < +\infty$	$\mathcal{O}(1)$	weak convergence
$q \in (0,1)$, $p = 2$, $\beta \geq q(1-q)$	$\mathcal{O}\left(t^{-2q}\right)$	$\mathcal{O}\left(t^{-q}\right)$, $\int_{t_0}^{+\infty} s\|\dot{x}(s)\|^2 < +\infty$	$\mathcal{O}(1)$	weak convergence

Table 4.1: Summary of main asymptotic results for (MTRIGS). The function $z(\cdot)$ is the generalized regularization path, that will be introduced in Section 4.6.2. The merit function $\varphi(\cdot)$ measures the decay of the function values and gets introduced in Subsection 2.3.3. All results have to be understood asymptotically, i.e., as $t \to +\infty$.

To this end, we extend the concept of Tikhonov regularization, initially developed in order to handle ill-posed integral equations in [228, 229], to multiobjective optimization. The Tikhonov regularization of a convex optimization problem

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

reads

$$\min_{x\in\mathcal{H}} f(x) + \frac{\varepsilon}{2}\|x\|^2,$$

where $\varepsilon > 0$ is a positive constant. Denoting for all $\varepsilon > 0$ its unique minimizer by

$$x_\varepsilon := \arg\min_{x\in\mathcal{H}} \left\{ f(x) + \frac{\varepsilon}{2}\|x\|^2 \right\},$$

it holds that x_ε converges strongly to $\mathrm{proj}_{\arg\min_{x\in\mathcal{H}} f(x)}(0)$ as $\varepsilon \to 0$, given $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$. The set $\{x_\varepsilon : \varepsilon > 0\}$ forms a smooth curve called *regularization path*. This is one of the key ingredients used to prove the strong convergence of the trajectory solution of (TRIGS) to the element of minimum norm in $\arg\min_{x\in\mathcal{H}} f(x)$. To extend this approach to the multiobjective optimization setting, we need to define an appropriate generalization of the regularization path.

Although there are a few studies addressing Tikhonov regularization in multiobjective optimization (see [69, 68, 73, 74]), these works are limited to the finite dimensional case and impose stringent assumptions, such as the compactness of the weak Pareto set. Furthermore, these studies do not address whether a Pareto optimum with the minimum norm is achieved and are thus not suitable for our convergence analysis.

Therefore, given a regularization function $\varepsilon(\cdot)$ and a solution $x(\cdot)$ to (MTRIGS), we define the *generalized regularization path* for our problem as

$$z(t) \coloneqq \operatorname*{arg\,min}_{z \in \mathcal{H}} \max_{i=1,\ldots,m} f_i(z) - f_i(x(t)) + \frac{\varepsilon(t)}{2}\|z\|^2. \tag{4.109}$$

The optimization problem in (4.109) can be seen as a regularization of an adaptive Pascoletti-Serafini scalarization of (MOP) [91]. In fact, $z(\cdot)$ converges strongly to the weakly Pareto optimal point of (MOP) with minimal norm in a particular lower level set of the objective function. This result will allow us to conclude that the trajectory solutions $x(\cdot)$ of (MTRIGS) converges strongly to the same weakly Pareto optimal point of (MOP).

This section is organized as follows. In Subsection 4.6.1, we discuss the standing assumptions for this section. Subsection 4.6.2 is dedicated to Tikhonov regularization. We discuss the single objective case, provide a brief overview of existing work for the multiobjective setting, and prove the strong convergence of the generalized regularization path to the weakly Pareto optimal point of (MOP) with minimal norm in a particular lower level set of the objective function. In Subsection 4.6.3 we prove existence of solutions in finite dimensions and discuss uniqueness. Subsection 4.6.5 contains the asymptotic analysis of solutions of (MTRIGS). The main results of this section concern the fast convergence rate of the function values in terms of the merit function and the strong convergence of the trajectory solutions. To verify the theoretical results, we conclude this section with several numerical experiments presented in Subsection 4.6.6.

The content of this section was already published in the following paper:

[49] Boţ, R. I. and Sonntag, K. *Inertial dynamics with vanishing Tikhonov regularization for multiobjective optimization.* In: *Journal of Mathematical Analysis and Applications* 554 (2) (2025). DOI: 10.1016/j.jmaa.2025.129940.

4.6.1 Assumptions

We require the following standing assumptions, which we assume to hold throughout this section.

($\mathcal{A}_1$) The objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and continuously differentiable with L-Lipschitz continuous gradients $\nabla f_i : \mathcal{H} \to \mathcal{H}$ with $L > 0$ for all $i = 1, \dots, m$.

($\mathcal{A}_2$) Given initial data $t_0 > 0$ and $x_0, v_0 \in \mathcal{H}$, define $a \in \mathbb{R}^m$ with $a_i := \frac{\beta}{2t_0^p}\|x_0\|^2 + \frac{1}{2}\|v_0\|^2$ for $i = 1, \dots, m$. For all $x \in \mathcal{L}(F, F(x_0) + a)$ it holds that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ and further

$$R \coloneqq \sup_{F^* \in F(\mathcal{LP}_w(F, F(x_0)+a))} \inf_{z \in F^{-1}(\{F^*\})} \|z\| < +\infty. \tag{4.110}$$

($\mathcal{A}_3$) The set $S(q) := \arg\min_{z \in \mathcal{H}} \max_{i=1,\dots,m} f_i(z) - q_i \neq \emptyset$ is nonempty for all $q \in \mathbb{R}^m$ and the mapping $z_0 : \mathbb{R}^m \to \mathcal{H}$, $q \mapsto \mathrm{proj}_{S(q)}(0)$, is continuous.

Discussion of Assumption ($\mathcal{A}_2$)

Assumption ($\mathcal{A}_2$) is in the spirit of a hypothesis used in the literature (see [156, 216, 217, 221, 222, 224]) in the asymptotic analysis of gradient systems and first-order methods for multiobjective optimization. There, the assumption is formulated only for $a = 0$, which is recovered in our setting if we restrict the initial conditions to $x_0 = v_0 = 0$. For arbitrary initial conditions, our analysis requires the assumption to hold for $a \in \mathbb{R}^m_+$ with $a_i := \frac{\beta}{2t_0^p}\|x(t_0)\| + \frac{1}{2}\|\dot{x}(t_0)\|^2 \geq 0$ for $i = 1, \dots, m$, as for this choice of a, the solutions of (MTRIGS) can be shown to remain in $\mathcal{L}(F, F(x(t_0))+a)$. This expansion of the level set is necessary because of the additional Tikhonov regularization which can produce trajectories that leave the initial level set $\mathcal{L}(F, F(x(t_0)))$. We visualize ($\mathcal{A}_2$) in Figure 4.7, which shows the schematic image space for an (MOP) with two objective functions. Given an initial point $x_0 \in \mathcal{H}$ and $a \in \mathbb{R}^m$ from ($\mathcal{A}_2$), the set $F(\mathcal{LP}_w(F(x_0)+a))$ is shown in blue. For all function values $F^* \in F(\mathcal{LP}_w(F(x_0) + a))$ the constant R gives a uniform bound on the minimum norm element in the preimage $F^{-1}(\{F^*\})$. For the case of scalar optimization ($m = 1$) this assumption is naturally satisfied if a solution to the optimization problem exists.

Discussion of Assumption ($\mathcal{A}_3$)

We need Assumption ($\mathcal{A}_3$) to show the strong convergence of the generalized regularization path for multiobjective optimization problems. We illustrate the necessity of this assumption with an example in Subsection 4.6.2. In the following, we show that the continuity of the projection $q \mapsto z_0(q) := \mathrm{proj}_{S(q)}(0)$ is closely connected with the continuity of the set-valued map (see [27, 34, 35, 42, 176, 226] for related discussions)

$$S : \mathbb{R}^m \rightrightarrows \mathcal{H}, \quad q \mapsto S(q) := \arg\min_{z \in \mathcal{H}} \max_{i=1,\dots,m} f_i(z) - q_i.$$

To this end, we recall the notion of Mosco convergence [34].

Definition 4.6.1. *Let $\left(C^k\right)_{k\geq 0}, C^* \subseteq \mathcal{H}$ be nonempty, convex and closed sets. We say that the sequence $\left(C^k\right)_{k\geq 0}$ is Mosco convergent to C^* if*

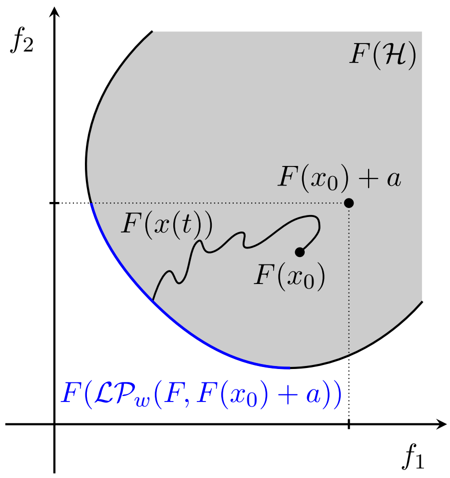


Figure 4.7: Visualization of $(\mathcal{A}_2)$ with a trajectory $x(t) \in \mathcal{LP}_w(F, F(x_0)+a)$.

i) For any $x^* \in C^*$ *there exists* $(x^k)_{k\geq 0}$ *with* $x^k \to x^*$ *such that* $x^k \in C^k$ *for all* $k \geq 0$*;*

ii) For any sequence $(k_l)_{l\geq 0} \subseteq \mathbb{N}$ *with* $x^{k_l} \in C^{k_l}$ *for all* $l \geq 0$ *such that* $x^{k_l} \rightharpoonup x^*$ *as* $l \to +\infty$*, it holds* $x^* \in C^*$*.*

The following theorem can be used to derive the continuity of $z_0(\cdot)$ from the *Mosco continuity* of $S(\cdot)$. We recall that a set-valued map $S(\cdot)$ is said to be Mosco continuous if for all $q^* \in \mathbb{R}^m$ and any sequence $\left(q^k\right)_{k\geq 0} \subseteq \mathbb{R}^m$ with $q^k \to q^*$ the sequence $\left(S(q^k)\right)_{k\geq 0}$ is Mosco convergent to $S(q^*)$.

Theorem 4.6.2. *([34, Sonntag-Attouch Theorem]) Let* $\left(C^k\right)_{k\geq 0}$, $C^* \subseteq \mathcal{H}$ *be nonempty, convex and closed sets. The following statements are equivalent:*

i) $\left(C^k\right)_{k\geq 0}$ *is Mosco convergent to* C^**;*

ii) $\left(C^k\right)_{k\geq 0}$ *is Wijsman convergent to* C^**, i.e., for all* $x \in \mathcal{H}$*, it holds* $\lim_{k\to+\infty} \operatorname{dist}(x, C^k) = \operatorname{dist}(x, C^*)$*;*

iii) For all $x \in \mathcal{H}$*, it holds* $\lim_{k\to+\infty} \operatorname{proj}_{C^k}(x) = \operatorname{proj}_{C^*}(x)$*.*

The following proposition shows that for all $q^* \in \mathbb{R}^m$ and for any sequence $\left(q^k\right)_{k\geq 0} \subseteq \mathbb{R}^m$ with $q^k \to q^*$, condition *ii)* in the definition of the Mosco convergence of $\left(S(q^k)\right)_{k\geq 0}$ to $S(q^*)$ is always fulfilled.

Proposition 4.6.3. *Let* $q^* \in \mathbb{R}^m$ *and* $\left(q^k\right)_{k\geq 0} \subseteq \mathbb{R}^m$ *be a sequence with* $q^k \to q^*$ *as* $k \to +\infty$*. Let* $(x^k)_{k\geq 0} \subseteq \mathcal{H}$ *be a sequence with* $x^k \in S(q^k)$ *for all* $k \geq 0$ *such that* $x^k \rightharpoonup x^* \in \mathcal{H}$ *as* $k \to +\infty$*. Then,* $x^* \in S(q^*)$*.*

Proof. We show that for all $z \in \mathcal{H}$

$$\max_{i=1,\ldots,m} f_i(x^*) - q_i^* \leq \max_{i=1,\ldots,m} f_i(z) - q_i^*.$$

Let $z \in \mathcal{H}$ be arbitrary. We use the weak lower semicontinuity of $\max_{i=1,\ldots,m} f_i(\cdot) - q_i^*$ to conclude

$$\begin{aligned}
\max_{i=1,\ldots,m} f_i(x^*) - q_i^* &\leq \liminf_{k\to+\infty} \max_{i=1,\ldots,m} f_i(x^k) - q_i^* \leq \liminf_{k\to+\infty} \left(\max_{i=1,\ldots,m} f_i(x^k) - q_i^k + \max_{i=1,\ldots,m} q_i^k - q_i^* \right) \\
&= \liminf_{k\to+\infty} \max_{i=1,\ldots,m} f_i(x^k) - q_i^k \leq \liminf_{k\to+\infty} \max_{i=1,\ldots,m} f_i(z) - q_i^k \\
&\leq \liminf_{k\to+\infty} \left(\max_{i=1,\ldots,m} f_i(z) - q_i^* + \max_{i=1,\ldots,m} q_i^* - q_i^k \right) = \max_{i=1,\ldots,m} f_i(z) - q_i^*.
\end{aligned}$$

Hence $x^* \in S(q^*)$, which completes the proof. □

The condition *i)* in the definition of the Mosco convergence of $\left(S(q^k)\right)_{k\geq 0}$ to $S(q^*)$ when $q^k \to q^*$ as $k \to +\infty$ does not hold in general, but can be shown to be satisfied under various circumstances. One of these is when the function $x \mapsto \max_{i=1,\ldots,m} f_i(x) - q_i$ exhibits a growth property uniformly for $q \in \mathbb{R}^m$ along approximating sequences.

Definition 4.6.4. *(Growth property uniformly along approximating sequences) Assume $S(q) \neq \emptyset$ for all $q \in \mathbb{R}^m$. We say that the function $x \mapsto \max_{i=1,\ldots,m} f_i(x) - q_i$ satisfies the growth property uniformly along approximating sequences if for all $q^* \in \mathbb{R}^m$ there exists a strictly increasing function $\psi : [0, +\infty) \to [0, +\infty)$ with $\psi(0) = 0$ such that for all sequences $\left(q^k\right)_{k\geq 0} \subseteq \mathbb{R}^m$ with $q^k \to q^*$ as $k \to +\infty$ it holds*

$$\max_{i=1,\ldots,m} f_i(x^*) - q_i^k - \inf_{z\in\mathcal{H}} \max_{i=1,\ldots,m} f_i(z) - q_i^k \geq \psi\left(\operatorname{dist}(x^*, S(q^k))\right),$$

for all $x^* \in S(q^*)$ and $k \geq 0$.

The following lemma states the Lipschitz continuity of the optimal value function arising in the definition of the set-valued map $S(\cdot)$.

Lemma 4.6.5. *Assume $S(q) \neq \emptyset$ for all $q \in \mathbb{R}^m$. Then, the optimal value function*

$$v : \mathbb{R}^m \to \mathbb{R}, \quad q \mapsto v(q) := \inf_{z\in\mathcal{H}} \max_{i=1,\ldots,m} f_i(z) - q_i,$$

is Lipschitz continuous.

Proof. Let $q^1, q^2 \in \mathbb{R}^m$ and choose $x^1 \in S(q^1)$ and $x^2 \in S(q^2)$. It holds

$$\begin{aligned}
v(q^1) &= \max_{i=1,\ldots,m} f_i(x^1) - q_i^1 \leq \max_{i=1,\ldots,m} f_i(x^2) - q_i^1 \\
&\leq \max_{i=1,\ldots,m} f_i(x^2) - q_i^2 + \max_{i=1,\ldots,m} q_i^2 - q_i^1 \leq v(q^2) + \|q^1 - q^2\|_\infty.
\end{aligned}$$

Analogously,

$$v(q^2) \leq v(q^1) + \|q^1 - q^2\|_\infty,$$

thus,

$$|v(q^1) - v(q^2)| \leq \|q^1 - q^2\|_\infty.$$

□

The next theorem shows that the uniform growth property indeed guarantees that for all $q^* \in \mathbb{R}^m$ and for any sequence $\left(q^k\right)_{k\geq 0} \subseteq \mathbb{R}^m$ with $q^k \to q^*$, the sequence $\left(S(q^k)\right)_{k\geq 0}$ is Mosco convergent to $S(q^*)$. Therefore, in the light of Theorem 4.6.2, Assumption $(\mathcal{A}_3)$ is fulfilled.

Theorem 4.6.6. *Assume $S(q) \neq \emptyset$ for all $q \in \mathbb{R}^m$ and that $x \mapsto \max_{i=1,\dots,m} f_i(x) - q_i$ satisfies the growth property uniformly along approximating sequences. Let $q^* \in \mathbb{R}^m$ and $\left(q^k\right)_{k\geq 0} \subseteq \mathbb{R}^m$ be a sequence with $q^k \to q^*$ as $k \to +\infty$. Then, $\left(S(q^k)\right)_{k\geq 0}$ is Mosco convergent to $S(q^*)$.*

Proof. Condition *ii)* in Definition 4.6.1 is satisfied according to Proposition 4.6.3. We prove by contradiction that condition *i)* is also satisfied. Let $x^* \in S(q^*)$ be such that for any sequence $\left(x^k\right)_{k\geq 0}$ with $x^k \in S(q^k)$ for all $k \geq 0$, it holds $x^k \not\to x^*$ as $k \to +\infty$. Hence, there exist $\delta > 0$ and a subsequence $(k_l)_{l\geq 0} \subseteq \mathbb{N}$ such that $\operatorname{dist}(x^*, S(q^{k_l})) > \delta$ for all $l \geq 0$. We use the growth property to conclude for all $l \geq 0$

$$\max_{i=1,\dots,m} f_i(x^*) - q_i^{k_l} - \inf_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(z) - q_i^{k_l} \geq \psi\left(\operatorname{dist}(x^*, S(q^{k_l}))\right) \geq \psi(\delta) > 0,$$

which yields for all $l \geq 0$

$$\max_{i=1,\dots,m} q_i^* - q_i^{k_l} + v(q^*) - v(q^{k_l}) \geq \psi(\delta) > 0.$$

We let $l \to +\infty$ and use $q^{k_l} \to q^*$ and the continuity of the optimal value function to derive a contradiction. $\square$

4.6.2 Tikhonov regularization for multiobjective optimization

In this subsection we extend the concept of Tikhonov regularization from scalar optimization to multiobjective optimization and study the properties of the associated regularization path. The obtained results will play a crucial role in the asymptotic analysis we perform in the following subsections for (MTRIGS).

A fundamental concept in the study of Tikhonov regularization when minimizing a convex and differentiable function $f : \mathcal{H} \to \mathbb{R}$, is the regularization path. This path, defined as $\{x_\varepsilon : \varepsilon > 0\}$ where x_ε is the unique minimizer of $f + \frac{\varepsilon}{2}\|\cdot\|^2$, is a smooth and bounded curve. As $\varepsilon \to 0$, it holds $x_\varepsilon \to \operatorname{proj}_{\arg\min_{x\in\mathcal{H}} f(x)}(0)$ converges strongly in $\mathcal{H}$ (see [32, Theorem 27.23]). The regularization path is crucial in the asymptotic analysis conducted in [12] for the system (TRIGS), where the convergence of the trajectory solution $x(\cdot)$ to the minimum norm solution gets demonstrated by showing that $\lim_{t\to+\infty}\|x(t) - x_{\varepsilon(t)}\| = 0$. We aim to extend this idea to the multiobjective setting when studying (MOP) and the dynamical system (MTRIGS).

Although the analysis presented in this section holds in a more general form for any continuously differentiable and monotonically decreasing function $\varepsilon : [t_0, +\infty) \to (0, +\infty)$ which satisfies $\lim_{t\to+\infty} \varepsilon(t) = 0$, we restrict the analysis in this section to the case $\varepsilon(t) = \frac{\beta}{t^p}$ in order to be consistent with the formulation of the system (MTRIGS). Define for all $t \geq t_0$ the Tikhonov regularized multiobjective optimization problem

$$(\text{MOP}_{\frac{\beta}{t^p}}) \qquad \min_{x\in\mathcal{H}} \begin{bmatrix} f_{t,1}(x) \\ \vdots \\ f_{t,m}(x) \end{bmatrix} := \begin{bmatrix} f_1(x) + \frac{\beta}{2t^p}\|x\|^2 \\ \vdots \\ f_m(x) + \frac{\beta}{2t^p}\|x\|^2 \end{bmatrix},$$

where we use for $i = 1, \dots, m$ the component-wise regularization

$$f_{t,i} : \mathcal{H} \to \mathbb{R}, \quad x \mapsto f_{t,i}(x) := f_i(x) + \frac{\beta}{2t^p}\|x\|^2.$$

Although the functions $f_{t,i}$ are strongly convex, one cannot expect (MOP$_{\frac{\beta}{t^p}}$) to have a unique Pareto optimal solution. This necessitates a suitable concept of a regularization path. To address this, we utilize the merit function defined in (2.23) for the regularized problem (MOP$_{\frac{\beta}{t^p}}$), that we define for all $t \geq t_0$ as

$$\begin{aligned} \varphi_t : \mathcal{H} \to \mathbb{R}, \quad x \mapsto \varphi_t(x) &:= \sup_{z\in\mathcal{H}} \min_{i=1,\dots,m} f_{t,i}(x) - f_{t,i}(z) \\ &= \sup_{z\in\mathcal{H}} \min_{i=1,\dots,m} f_i(x) - f_i(z) + \frac{\beta}{2t^p}\|x\|^2 - \frac{\beta}{2t^p}\|z\|^2. \end{aligned} \tag{4.111}$$

The optimization problem in the definition of the merit function $\varphi_t(\cdot)$ can be interpreted as the Pascoletti-Serafini scalarization of the problem (MOP$_{\frac{\beta}{t^p}}$) (see, e.g., [91, Section 2.1]). Inspired by the formulation of the merit function and by the Tikhonov regularization in scalar optimization, we consider for all $t \geq t_0$ the unique minimizer

$$z(t) \in \operatorname*{arg\,min}_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(z) - f_i(x(t)) + \frac{\beta}{2t^p}\|z\|^2, \tag{4.112}$$

as an element of the regularization path, where $x : [t_0, +\infty) \to \mathcal{H}$ is a trajectory which will be specified later. Note that in scalar optimization, namely when $m = 1$, we recover the classical regularization path independent of the trajectory $x(\cdot)$. Since the function $z \mapsto \max_{i=1,\dots,m} f_i(z) - f_i(x(t))$ depends on t, we cannot make use of the properties of the regularization path in scalar optimization to characterize the asymptotic behavior of this new path. This will be done in the following result.

Theorem 4.6.7. *Let $q : [t_0, +\infty) \to \mathbb{R}^m$ be a continuous function with $q(t) \to q^* \in \mathbb{R}^m$ as $t \to +\infty$, and*

$$\begin{aligned} z(t) &:= \operatorname*{arg\,min}_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(z) - q_i(t) + \frac{\beta}{2t^p}\|z\|^2 \quad \text{for all} \quad t \geq t_0, \\ S(q) &:= \operatorname*{arg\,min}_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(z) - q_i \quad \text{for all} \quad q \in \mathbb{R}^m, \\ z_0(q) &:= \operatorname*{proj}_{S(q)}(0) \quad \text{for all} \quad q \in \mathbb{R}^m. \end{aligned} \tag{4.113}$$

Then, $z(t) \to z_0(q^*)$ converges strongly as $t \to +\infty$.

Proof. Let $(t_k)_{k\geq 0} \subset [t_0, +\infty)$ be an arbitrary sequence with $t_k \to +\infty$ as $k \to +\infty$. For all $k \geq 0$, we denote $\varepsilon_k := \frac{\beta}{(t_k)^p}$, $q^k := q(t_k)$, $z^k := z(t_k)$, and $z_0^k := z_0(q^k)$. For all $k \geq 0$ it holds

$$\begin{aligned} \max_{i=1,\dots,m} f_i(z^k) - q_i^k + \frac{\varepsilon_k}{2}\|z^k\|^2 &\leq \max_{i=1,\dots,m} f_i(z_0^k) - q_i^k + \frac{\varepsilon_k}{2}\|z_0^k\|^2 \\ &\leq \max_{i=1,\dots,m} f_i(z^k) - q_i^k + \frac{\varepsilon_k}{2}\|z_0^k\|^2, \end{aligned} \tag{4.114}$$

hence,

$$\|z^k\| \leq \|z_0^k\|. \tag{4.115}$$

According to Assumption $(\mathcal{A}_3)$, $z_0(\cdot)$ is continuous, consequently, $\left(z_0^k\right)_{k\geq 0}$ is bounded. This implies that $\left(z^k\right)_{k\geq 0}$ is also bounded and hence possesses a weak sequential cluster point. We show that this cluster point is unique, which will imply that $\left(z^k\right)_{k\geq 0}$ is weakly convergent.

Let z^∞ be an arbitrary weak sequential cluster point of $\left(z^k\right)_{k\geq 0}$. Then, there exists a subsequence with $z^{k_l} \rightharpoonup z^\infty$ weakly in $\mathcal{H}$ as $l \to +\infty$. For all $z \in \mathcal{H}$ it holds

$$\begin{aligned}
&\max_{i=1,\ldots,m} \left(f_i(z^\infty) - q_i^*\right) \leq \liminf_{l\to+\infty} \max_{i=1,\ldots,m} \left(f_i(z^{k_l}) - q_i^*\right) + \frac{\varepsilon_{k_l}}{2}\|z^{k_l}\|^2 \\
&\leq \liminf_{l\to+\infty} \left(\max_{i=1,\ldots,m} \left(f_i(z^{k_l}) - q_i^{k_l}\right) + \frac{\varepsilon_{k_l}}{2}\|z^{k_l}\|^2 + \max_{i=1,\ldots,m} \left(q_i^{k_l} - q_i^*\right)\right) \\
&\leq \liminf_{l\to+\infty} \left(\max_{i=1,\ldots,m} \left(f_i(z) - q_i^{k_l}\right) + \frac{\varepsilon_{k_l}}{2}\|z\|^2 \right) \\
&\leq \liminf_{l\to+\infty} \left(\max_{i=1,\ldots,m} \left(f_i(z) - q_i^*\right) + \frac{\varepsilon_{k_l}}{2}\|z\|^2 + \max_{i=1,\ldots,m} \left(q_i^* - q_i^{k_l}\right)\right) \\
&= \max_{i=1,\ldots,m} \left(f_i(z) - q_i^*\right).
\end{aligned} \tag{4.116}$$

From here, $z^\infty \in S(q^*)$ follows. Next, we show that $z^\infty = z_0(q^*)$. From the continuity of $z_0(\cdot)$ we have

$$z_0^{k_l} = z_0(q^{k_l}) \to z_0(q^*) \quad \text{as} \quad l \to +\infty, \tag{4.117}$$

and the weak lower semicontinuity of the norm gives

$$\|z^\infty\| \leq \liminf_{l\to+\infty}\|z^{k_l}\| \leq \limsup_{l\to+\infty}\|z^{k_l}\| \leq \limsup_{l\to+\infty}\|z_0^{k_l}\| = \|z_0(q^*)\|. \tag{4.118}$$

Since $z^\infty \in S(q^*)$ and $z_0(q^*) = \operatorname{proj}_{S(q^*)}(0)$, we get $z^\infty = z_0(q^*)$. This proves that $\left(z^k\right)_{k\geq 0}$ converges weakly to $z_0(q^*)$. Using again (4.118), we get

$$\lim_{k\to+\infty}\|z^k\| = \|z_0(q^*)\|,$$

from which we conclude that $z^k \to z_0(q^*)$ converges strongly as $k \to +\infty$. □

Remark 4.6.8. *The continuity of $z_0(\cdot)$ formulated in Assumption $(\mathcal{A}_3)$ can be seen as a regularity condition on the objective functions f_i for $i = 1, \ldots, m$. It is satisfied for convex scalar optimization problems as long as the set of minimizers is not empty. In this setting the mapping $q \to z_0(q)$ is constant. The following example shows that the Assumption $(\mathcal{A}_3)$ is crucial for obtaining convergence of $z(t)$ as $t \to +\infty$.*

Example 4.6.9. *Define the functions*

$$\phi : \mathbb{R} \to \mathbb{R}, \quad y \mapsto \frac{1}{2} \max(y-3,0)^2 + \frac{1}{2} \max(2-y,0)^2,$$

$$g : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \begin{cases} \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2, & \textit{if } |x_1| \le 1, \quad x_2 + 1 \le \sqrt{1-x_1^2}, \\ |x_1| + \frac{1}{2}x_2^2 - \frac{1}{2}, & \textit{if } |x_1| > 1, \quad x_2 + 1 \le 0, \\ \sqrt{x_1^2 + (x_2+1)^2} - (x_2+1), & \textit{else}, \end{cases}$$

$$f_1 : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \frac{1}{2}(x_1 - 1)^2 + \phi(x_2) + g(x),$$

$$f_2 : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \frac{1}{2}(x_1 + 1)^2 + \phi(x_2) + g(x), \tag{4.119}$$

which are all convex and continuously differentiable with Lipschitz continuous gradients. In the following, we verify the differentiability of $g : \mathbb{R}^2 \to \mathbb{R}$ and show that ∇g is Lipschitz continuous. Then, the regularity of f_1 and f_2 follows.

The gradient of g is given by

$$\nabla g : \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto \begin{cases} x, & \textit{if } x \in M_1, \\ \begin{bmatrix} \frac{x_1}{|x_1|} \\ x_2 \end{bmatrix}, & \textit{if } x \in M_2, \\ \begin{bmatrix} \frac{x_1}{\sqrt{x_1^2+(x_2+1)^2}} \\ \frac{x_2+1}{\sqrt{x_1^2+(x_2+1)^2}} - 1 \end{bmatrix}, & \textit{if } x \in M_3, \end{cases}$$

with

$$\begin{aligned} M_1 &:= \left\{ x \in \mathbb{R}^2 : |x_1| \le 1, \, x_2 + 1 \le \sqrt{1 - x_1^2} \right\}, \\ M_2 &:= \left\{ x \in \mathbb{R}^2 : |x_1| > 1, x_2 + 1 \le 0 \right\}, \\ M_3 &:= \mathbb{R}^2 \setminus (M_1 \cup M_2). \end{aligned}$$

The gradient ∇g is Lipschitz continuous on $\overline{M_i}$ for $i = 1, 2, 3$. Since $\nabla g\big|_{\overline{M_i}}$ and $\nabla g\big|_{\overline{M_j}}$ coincide on $\overline{M_i} \cap \overline{M_j}$ for $i \neq j \in \{1, 2, 3\}$, the Lipschitz continuity of ∇g follows. In fact, $\nabla g(\cdot) = \mathrm{proj}_{M_1}(\cdot)$, hence the Lipschitz constant of the gradient is 1.

Now, we consider the multiobjective optimization problem

$$\min_{x \in \mathcal{H}} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}, \tag{4.120}$$

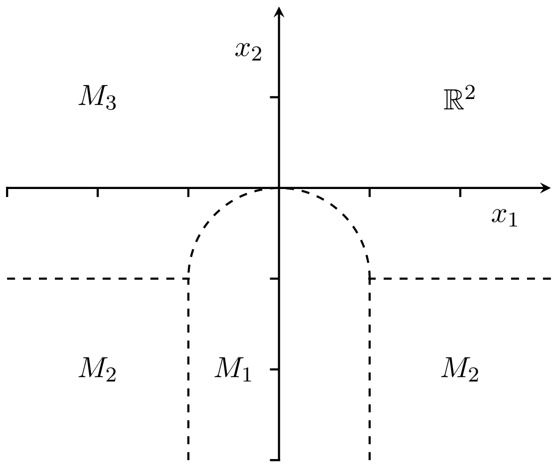


Figure 4.8: The sets $M_i \subseteq \mathbb{R}^2$ for $i = 1, 2, 3$.

and the Tikhonov regularized problem

$$\min_{x \in \mathcal{H}} \begin{bmatrix} f_1(x) + \frac{\varepsilon}{2}\|x\|^2 \\ f_2(x) + \frac{\varepsilon}{2}\|x\|^2 \end{bmatrix}. \tag{4.121}$$

Figure 4.9a illustrates the Pareto set of the problem (4.120) *(denoted by P) alongside the Pareto set of the regularized problem* (4.121) *for various values of $\varepsilon > 0$ (denoted by P_ε). As ε decreases, the Pareto set of* (4.121) *"converges" to the Pareto set of* (4.120)*. Due to the T-shape of the Pareto set, the edges of the regularized Pareto sets become sharper as ε diminishes. For this problem the map*

$$z_0 : \mathbb{R}^2 \to \mathbb{R}^2, \quad q \mapsto z_0(q) = \operatorname*{proj}_{S(q)}(0),$$

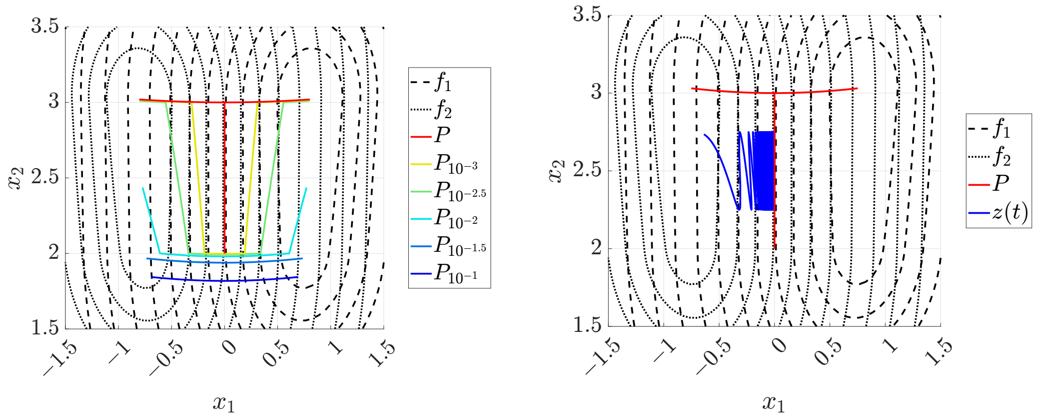
with $S(q) = \operatorname*{arg\,min}_{z \in \mathbb{R}^2} \max\left(f_1(z) - q_1, f_2(z) - q_2\right)$ *is not continuous everywhere. Indeed,*

$$z_0(q_1, 0) \to (0, 3) \neq (0, 2) = \operatorname*{proj}_{\{0\}\times[2,3]}(0) = z_0((0, 0)), \quad \textit{as} \quad q_1 \to 0.$$

We define, for $t_0 := (192\beta)^{\frac{1}{p}}$,

$$q : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} := \begin{bmatrix} 2(\omega(t) + 1)\sqrt{\left(\frac{t^p}{t^p - \beta\omega(t)}\right)^2 - 1} \\ 0 \end{bmatrix},$$

with $\omega(t) := \frac{10+\sin(\eta t)}{4}$, *where $\eta > 0$ is a positive scaling parameter. It holds $q(t) \to q^* = (0, 0)^\top$ as $t \to +\infty$.*



(a) (b)

Figure 4.9: Contour plots of the functions f_1 and f_2 defined in (4.119): (a) The Pareto sets of (4.120) and (4.121) for $\varepsilon \in \{10^{-1}, 10^{-1.5}, 10^{-2}, 10^{-2.5}, 10^{-3}\}$. (b) The Pareto set of (4.120) and the regularization path $z(\cdot)$ defined in (4.123) with parameters $p = 1$, $\beta = \frac{1}{2}$, $\eta = \frac{1}{50}$.

For this example, for all $t \geq t_0$ the regularization path

$$z(t) \in \arg\min_{z \in \mathbb{R}^2} \max\left(f_1(z) - q_1(t), f_2(z) - q_2(t)\right) + \frac{\beta}{2t^p}\|z\|^2, \tag{4.122}$$

is given by

$$z(t) = \begin{bmatrix} -(\omega(t)+1)\sqrt{\left(\frac{t^p}{t^p - \beta\omega(t)}\right)^2 - 1} \\ \omega(t) \end{bmatrix}, \tag{4.123}$$

as we show in the following. For all $t \geq t_0$, the function

$$\Phi_t : \mathbb{R}^2 \to \mathbb{R}, \quad z \mapsto \Phi_t(z) := \max\left(f_1(z) - q_1(t), f_2(z) - q_2(t)\right) + \frac{\beta}{2t^p}\|z\|^2,$$

is strongly convex and therefore has a unique minimizer. We show that

$$0 \in \partial_z \Phi_t(z(t)), \tag{4.124}$$

where $\partial_z \Phi_t(z(t))$ denotes the convex subdifferential of $\Phi_t(\cdot)$ evaluated at $z(t)$. Note that $z_2(t) \in [2.25, 2.75]$ for all $t \geq t_0$ and hence

$$\Phi_t(z) = \frac{1}{2}z_1^2 + \frac{1}{2} + g(z) + \frac{\beta}{2t^p}\|z\|^2 + \max\left(-z_1 - q_1(t), z_1\right),$$

on an open neighborhood of $z(t)$. We have

$$\partial_z \Phi_t(z(t)) = \begin{bmatrix} z_1(t) + \frac{z_1(t)}{\sqrt{z_1(t)^2 + (z_2(t)+1)^2}} + \frac{\beta}{t^p} z_1(t) \\ \frac{z_2(t)+1}{\sqrt{z_1(t)^2 + (z_2(t)+1)^2}} - 1 + \frac{\beta}{t^p} z_2(t) \end{bmatrix} + \partial_z \max\left(-z_1(t) - q_1(t), z_1(t)\right).$$

Since $z_1(t) = -\frac{1}{2}q_1(t)$ *we have* $\partial_z \max(-z_1(t) - q_1(t), z_1(t)) = [-1,1] \times \{0\}$ *and hence*

$$\partial_z \Phi_t(z(t)) = \begin{bmatrix} z_1(t) + \frac{z_1(t)}{\sqrt{z_1(t)^2+(z_2(t)+1)^2}} + \frac{\beta}{t^p} z_1(t) \\ \frac{z_2(t)+1}{\sqrt{z_1(t)^2+(z_2(t)+1)^2}} - 1 + \frac{\beta}{t^p} z_2(t) \end{bmatrix} + [-1,1] \times \{0\}. \tag{4.125}$$

For all $t \geq t_0 = (192\beta)^{\frac{1}{p}}$*, taking into account the definition of* $z_1(t)$ *and* $z_2(t) \in [2.25, 2.75]$*, it holds that*

$$z_1(t) + \frac{z_1(t)}{\sqrt{z_1(t)^2 + (z_2(t)+1)^2}} + \frac{\beta}{t^p} z_1(t) \in [-1,1].$$

On the other hand, since

$$z_1(t) = -(z_2(t)+1)\sqrt{\left(\frac{t^p}{t^p - \beta z_2(t)}\right)^2 - 1},$$

we have

$$\frac{z_2(t)+1}{\sqrt{z_1(t)^2 + (z_2(t)+1)^2}} = 1 - \frac{\beta}{t^p} z_2(t),$$

which proves that (4.124) *and therefore* (4.122) *are satisfied.*

In Figure 4.9 (b), the regularization path $z(\cdot)$ *given by* (4.123) *is depicted. One can observe that it oscillates in the* x_2*-coordinate between the values* 2.25 *and* 2.75 *as* $t \to +\infty$*. The function* $z(t)$ *does not converge as* $t \to +\infty$*, although all accumulation points are Pareto optimal and global minimizers of* $\max(f_1(z) - q_1^*, f_2(z) - q_2^*)$*. The minimal norm solution* $z_0(q^*) = (0,2)$ *is not an accumulation point of* $z(\cdot)$*. This example clearly shows that the continuity of* $z_0(\cdot)$ *is essential to derive Theorem 4.6.7.*

We conclude this section by introducing three propositions that summarize the main properties of $z(\cdot)$.

Proposition 4.6.10. *Let* $a \in \mathbb{R}^m_+$ *and assume that the function* $x : [t_0, +\infty) \to \mathcal{H}$ *fulfills* $x(t) \in \mathcal{L}(F, F(x(t_0)) + a)$ *for all* $t \geq t_0$*. Then, the regularization path,*

$$z(t) := \underset{z \in \mathcal{H}}{\arg\min} \max_{i=1,\dots,m} f_i(z) - f_i(x(t)) + \frac{\beta}{2t^p}\|z\|^2, \quad \textit{for all} \quad t \geq t_0,$$

is bounded. Specifically, $z(t) \in B_R(0)$ *for all* $t \geq t_0$*, where* R *is defined in* $(\mathcal{A}_2)$*.*

Proof. By $(\mathcal{A}_3)$, it holds $S(F(x(t))) := \underset{z \in \mathcal{H}}{\arg\min} \max_{i=1,\dots,m}(f_i(z) - f_i(x(t))) \neq \emptyset$ for all $t \geq t_0$. Fix some $t \geq t_0$. From the properties of Tikhonov regularization in scalar optimization (see [32, Theorem 27.23]), we know

$$\|z(t)\| \leq \|z\| \quad \text{for all} \quad z \in S(F(x(t))). \tag{4.126}$$

Next, we show that

$$F^{-1}(\{F^*\}) \subseteq S(x(t)) \quad \text{for all} \quad F^* \in F(S(F(x(t)))). \tag{4.127}$$

Let $F^* \in F(S(F(x(t)))$. Then, there exists $z \in S(F(x(t))$ with $F(z) = F^*$. Let $w \in F^{-1}(\{F^*\})$ then $F(w) = F(z)$ and hence

$$\max_{i=1,\dots,m} f_i(w) - f_i(x(t)) = \max_{i=1,\dots,m} f_i(z) - f_i(x(t)) = \inf_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(z) - f_i(x(t)).$$

This shows $w \in S(F(x(t))$ and hence (4.127) holds. From (4.126) and (4.127) we conclude that for all $F^* \in F(S(F(x(t))))$ we get

$$\|z(t)\| \leq \|z\| \quad \text{for all} \quad z \in F^{-1}(\{F^*\}),$$

and hence

$$\|z(t)\| \leq \inf_{z\in F^{-1}(\{F^*\})} \|z\| \quad \text{for all} \quad F^* \in F(S(F(x(t)))).$$

Since this bound holds for all $F^* \in F(S(F(x(t))))$, we get

$$\begin{aligned} \|z(t)\| &\leq \inf_{z\in F^{-1}(F(S(F(x(t)))))} \|z\| = \inf_{\{z\in\mathcal{H}:F(z)\in F(S(F(x(t))))\}} \|z\| \\ &\leq \sup_{F^*\in F(S(F(x(t)))} \inf_{z\in F^{-1}(\{F^*\})} \|z\|. \end{aligned} \tag{4.128}$$

Next, we prove that

$$S(F(x(t))) \subseteq \mathcal{LP}_w(F, F(x(t_0)) + a). \tag{4.129}$$

Let $z \in S(F(x(t)))$. Then,

$$\max_{i=1,\dots,m} f_i(z) - f_i(x(t)) \leq \max_{i=1,\dots,m} f_i(x(t)) - f_i(x(t)) = 0,$$

hence

$$f_i(z) \leq f_i(x(t)) \leq f_i(x(t_0)) + a_i \quad \text{for all} \quad i = 1, \dots, m,$$

and therefore $z \in \mathcal{L}(F, F(x(t_0)) + a)$. Assuming that $z \notin \mathcal{LP}_w(F, F(x(t_0)) + a)$, it follows that $z \notin \mathcal{P}_w$ and hence there exists some $y \in \mathcal{H}$ with

$$f_i(y) < f_i(z) \quad \text{for all} \quad i = 1, \dots, m.$$

Therefore,

$$\max_{i=1,\dots,m} f_i(x) - f_i(x(t)) < \max_{i=1,\dots,m} f_i(z) - f_i(x(t)),$$

which is a contradiction to $z \in S(F(x(t)))$. This proves inclusion (4.129). Consequently, according to (4.128) and (4.129),

$$\|z(t)\| \leq \sup_{F^*\in F(\mathcal{LP}_w(F,F(x(t_0))+a)} \inf_{z\in F^{-1}(\{F^*\})} \|z\| = R < +\infty,$$

where the upper bound R is given by $(\mathcal{A}_2)$. □

Proposition 4.6.11. *Let $q : [t_0, +\infty) \to \mathbb{R}^m$ be a continuous function and define*

$$z : [t_0, +\infty) \to \mathcal{H}, \quad t \mapsto z(t) := \operatorname*{arg\,min}_{z \in \mathcal{H}} \max_{i=1,\dots,m} f_i(z) - q_i(t) + \frac{\beta}{2t^p}\|z\|^2.$$

Then, $z(\cdot)$ is a continuous mapping.

Proof. We fix an arbitrary $\bar{t} \geq t_0$ and show that $z(\cdot)$ is continuous (continuous from the right if $\bar{t} = t_0$) in $\bar{t}$. Let $t \in \left[\bar{t} - \kappa, \bar{t} + \kappa\right] \cap [t_0, +\infty)$ for some $\kappa > 0$. Then, by strong convexity and the minimizing properties of $z(t)$ and $z(\bar{t})$, we get

$$\begin{aligned} &\max_{i=1,\dots,m} \left(f_i(z(\bar{t})) - q_i(t)\right) + \frac{\beta}{2t^p}\|z(\bar{t})\|^2 \\ &- \max_{i=1,\dots,m} \left(f_i(z(t)) - q_i(t)\right) - \frac{\beta}{2t^p}\|z(t)\|^2 \geq \frac{\beta}{2t^p}\|z(\bar{t}) - z(t)\|^2, \end{aligned} \tag{4.130}$$

and

$$\begin{aligned} &\max_{i=1,\dots,m} \left(f_i(z(t)) - q_i(\bar{t})\right) + \frac{\beta}{2\bar{t}^p}\|z(t)\|^2 \\ &- \max_{i=1,\dots,m} \left(f_i(z(\bar{t}) - q_i(\bar{t})\right) - \frac{\beta}{2\bar{t}^p}\|z(\bar{t})\|^2 \geq \frac{\beta}{2\bar{t}^p}\|z(t) - z(\bar{t})\|^2, \end{aligned} \tag{4.131}$$

respectively. Using the monotonicity of $t \mapsto \frac{\beta}{2t^p}$, (4.130) and (4.131) lead to

$$\begin{aligned} &\max_{i=1,\dots,m} \left(f_i(z(\bar{t}) - q_i(\bar{t})\right) + \max_{i=1,\dots,m} \left(q_i(\bar{t}) - q_i(t)\right) + \frac{\beta}{2t^p}\|z(\bar{t})\|^2 \\ &- \max_{i=1,\dots,m} \left(f_i(z(t) - q_i(t)\right) - \frac{\beta}{2t^p}\|z(t)\|^2 \geq \frac{\beta}{2(\bar{t} + \kappa)^p}\|z(\bar{t}) - z(t)\|^2, \end{aligned} \tag{4.132}$$

respectively,

$$\begin{aligned} &\max_{i=1,\dots,m} \left(f_i(z(t)) - q_i(t)\right) + \max_{i=1,\dots,m} \left(q_i(t) - q_i(\bar{t})\right) + \frac{\beta}{2\bar{t}^p}\|z(t)\|^2 \\ &- \max_{i=1,\dots,m} \left(f_i(z(\bar{t}) - q_i(\bar{t})\right) - \frac{\beta}{2\bar{t}^p}\|z(\bar{t})\|^2 \geq \frac{\beta}{2(\bar{t} + \kappa)^p}\|z(t) - z(\bar{t})\|^2. \end{aligned} \tag{4.133}$$

Adding (4.132) and (4.133) yields

$$2\|q(t) - q(\bar{t})\|_\infty + \frac{1}{2}\left(\frac{\beta}{\bar{t}^p} - \frac{\beta}{t^p}\right)\left(\|z(t)\|^2 - \|z(\bar{t})\|^2\right) \geq \frac{\beta}{(\bar{t} + \kappa)^p}\|z(t) - z(\bar{t})\|^2. \tag{4.134}$$

By Proposition 4.6.10, the function $z(\cdot)$ is bounded, so by the continuity of $q(\cdot)$ the left-hand-side of (4.134) vanishes as $t \to \bar{t}$. This demonstrates the continuity of $z(\cdot)$ in $\bar{t}$. □

In the next proposition, we describe the connection between the original merit function $\varphi(\cdot)$ and the merit function $\varphi_t(\cdot)$ of the regularized problem. This will allow us to derive asymptotic convergence results on $\varphi(x(t))$ for $t \to +\infty$.

Proposition 4.6.12. *Let $a \in \mathbb{R}^m_+$ be the vector introduced in Assumption $(\mathcal{A}_2)$ and assume that $x : [t_0, +\infty) \to \mathcal{H}$ fulfills $x(t) \in \mathcal{L}(F, F(x(t_0)) + a)$ for all $t \geq t_0$. We define*

$$z(t) := \underset{z \in \mathcal{H}}{\arg\min} \max_{i=1,\dots,m} f_i(z) - f_i(x(t)) + \frac{\beta}{2t^p}\|z\|^2 \quad \text{for all} \quad t \geq t_0.$$

Then, the following statements hold:

i) *For all $t \geq t_0$ and all $y \in \mathcal{H}$*

$$\min_{i=1,\dots,m} f_i(x(t)) - f_i(y) \leq \min_{i=1,\dots,n} f_{t,i}(x(t)) - f_{t,i}(z(t)) + \frac{\beta}{2t^p}\|y\|^2,$$

hence

$$\varphi(x(t)) \leq \varphi_t(x(t)) + \frac{\beta R^2}{2t^p},$$

where R is defined in $(\mathcal{A}_2)$.

ii) *For all $t \geq t_0$*

$$\|x(t) - z(t)\|^2 \leq \frac{t^p \varphi_t(x(t))}{\beta}.$$

Proof. *i)* Fix $t \geq t_0$ and $y \in \mathcal{H}$. From the definition of $z(t)$, we have

$$\max_{i=1,\dots,m} f_{t,i}(y) - f_{t,i}(x(t)) \geq \max_{i=1,\dots,m} f_{t,i}(z(t)) - f_{t,i}(x(t)),$$

hence

$$\min_{i=1,\dots,m} f_i(x(t)) - f_i(y) + \frac{\beta}{2t^p}\|x(t)\|^2 - \frac{\beta}{2t^p}\|y\|^2 \leq \min_{i=1,\dots,m} f_{t,i}(x(t)) - f_{t,i}(z(t)).$$

Using the definition of $\varphi_t(\cdot)$, we get

$$\min_{i=1,\dots,m} f_i(x(t)) - f_i(y) \leq \varphi_t(x(t)) + \frac{\beta}{2t^p}\|y\|^2. \tag{4.135}$$

By $(\mathcal{A}_2)$, it holds $\mathcal{LP}_w(F, F(x(t_0)) + a) \neq \emptyset$, therefore,

$$\begin{aligned} &\sup_{F^* \in F(\mathcal{LP}_w(F,F(x(t_0))+a))} \inf_{y \in F^{-1}(\{F^*\})} \min_{i=1,\dots,m} f_i(x(t)) - f_i(y) \\ &\leq \varphi_t(x(t)) + \frac{\beta}{2t^p} \sup_{F^* \in F(\mathcal{LP}_w(F,F(x(t_0))+a))} \inf_{y \in F^{-1}(\{F^*\})} \|y\|^2. \end{aligned} \tag{4.136}$$

Additionally, we have

$$\begin{aligned} &\sup_{y \in \mathcal{LP}_w(F,F(x(t_0))+a)} \min_{i=1,\dots,m} f_i(x(t)) - f_i(y) \\ &= \sup_{F^* \in F(\mathcal{LP}_w(F,F(x(t_0))+a))} \inf_{y \in F^{-1}(\{F^*\})} \min_{i=1,\dots,m} f_i(x(t)) - f_i(y). \end{aligned} \tag{4.137}$$

Note that (4.137) holds since for all $y \in \mathcal{LP}_w(F, F(x(t_0)) + a)$ there exists $F^* = F(y) \in F(\mathcal{LP}_w(F, F(x(t_0))+a))$ with $\min_{i=1,\dots,m} f_i(x(t)) - f_i(y) = \min_{i=1,\dots,m} f_i(x(t)) - f_i(z)$ for all $z \in F^{-1}(\{F^*\})$. On the other hand, for all $F^* \in F(\mathcal{LP}_w(F, F(x(t_0))+a))$ any $y \in \mathcal{LP}_w(F, F(x(t_0))+a)$ with $F(y) = F^*$ satisfies $\min_{i=1,\dots,m} f_i(x(t)) - f_i(y) = \inf_{z\in F^{-1}(\{F^*\})} \min_{i=1,\dots,m} f_i(x(t)) - f_i(z)$. Combining (4.136) and (4.137), and using Lemma 2.3.15 and $(\mathcal{A}_2)$ yields

$$\varphi(x(t)) \leq \varphi_t(x(t)) + \frac{\beta R^2}{2t^p}.$$

ii) From the strong convexity of $f_{t,i}$ with modulus $\frac{\beta}{t^p}$, we conclude the strong convexity of $z \mapsto \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ with modulus $\frac{\beta}{t^p}$. This gives for all $t \geq t_0$

$$\begin{aligned}\varphi_t(x(t)) &= \min_{i=1,\dots,m} f_{t,i}(x(t)) - f_{t,i}(z(t))\\ &= \max_{i=1,\dots,m} f_{t,i}(x(t)) - f_{t,i}(x(t)) - \max_{i=1,\dots,m} f_{t,i}(z(t)) - f_{t,i}(x(t))\\ &\geq \frac{\beta}{t^p}\|x(t) - z(t)\|^2,\end{aligned}$$

and the desired inequality follows. □

4.6.3 Discussion of existence and uniqueness of solutions

The discussion of existence and uniqueness of solutions to (MTRIGS) is based on Section 4.3. We begin by properly defining a solution to (MTRIGS).

Definition 4.6.13. *We call a function $x : [t_0, +\infty) \to \mathcal{H}$ a solution to* (MTRIGS) *with initial data $t_0 > 0$, $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ if it satisfies the following conditions:*

i) *$x \in C^1([t_0, +\infty), \mathcal{H})$, i.e., $x(\cdot)$ is continuously differentiable on $[t_0, +\infty)$;*

ii) *$\dot{x}(\cdot)$ is absolutely continuous on $[t_0, T]$ for all $T \geq t_0$;*

iii) *There exists a (Bochner) measurable function $\ddot{x} : [t_0, +\infty) \to \mathcal{H}$ with $\dot{x}(t) = \dot{x}(t_0) + \int_{t_0}^t \ddot{x}(s)ds$ for all $t \geq t_0$;*

iv) *$\dot{x}(\cdot)$ is differentiable almost everywhere and $\frac{d}{dt}\dot{x}(t) = \ddot{x}(t)$ holds for almost all $t \in [t_0, +\infty)$;*

v) *$\frac{\alpha}{t^q}\dot{x}(t) + \operatorname{proj}_{C(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t)}(0) = 0$ holds for almost all $t \in [t_0, +\infty)$;*

vi) *$x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ hold.*

Theorem 4.6.14. *Assume $\mathcal{H}$ is finite-dimensional and and assume the gradients of the objective function $\nabla f_i(\cdot)$ are L-Lipschitz continuous for all $i = 1, \dots, m$. Then, for all $t_0 > 0$ and $x_0, v_0 \in \mathcal{H}$, there exists a function $x(\cdot)$ which is a solution to equation* (IMOG') *in the sense of Definition 4.4.2.*

Proof. The proof follows from Theorem 4.3.9. We show (MTRIGS) is a special instance of (D) for appropriate choices of $\gamma(\cdot)$ and $d_i(\cdot, \cdot)$. Define the functions

$$\begin{aligned}&\gamma : (0, +\infty) \to [0, +\infty), \quad t \mapsto \gamma(t) := \frac{\alpha}{t^q},\\ &d_i : (0, +\infty) \times \mathcal{H} \to \mathcal{H}, \quad (t, u) \mapsto d_i(t, u) := \nabla f_i(u) + \frac{\beta}{t^p}u,\end{aligned} \tag{4.138}$$

and let $D(\cdot,\cdot)$ be as defined in Definition 4.3.1. By (4.138) the functions $\gamma(\cdot)$ and $d_i(\cdot,\cdot)$ have the following properties. The function $\gamma(\cdot)$ is continuous and monotonically decreasing. The functions $d_i(\cdot,\cdot)$ are continuous on $(0,+\infty)\times\mathcal{H}$ and uniformly $\left(L+\frac{\beta}{(t_0)^p}\right)$-Lipschitz continuous in the second component on $[t_0,+\infty)\times\mathcal{H}$ for all $i=1,\dots,m$. Further, for all $(t,u)\in(0,+\infty)\times\mathcal{H}$ it holds that

$$D(t,u)=C(u)+\frac{\beta}{t^p}u.$$

For this choice of $\gamma(\cdot)$ and $d_i(\cdot,\cdot)$ equation (D) reduces to (MTRIGS) and we conclude the existence of a solution to (D) in the sense of Definition 4.3.6 by Theorem 4.3.9. □

Remark 4.6.15. *The uniqueness of the trajectory solutions of* (MTRIGS) *remains an open problem. There are two major difficulties in deriving uniqueness, as for the dynamical system* (MAVD)*. First, the multiobjective steepest descent direction is not Lipschitz continuous, but only Hölder continuous. So even for simpler multiobjective gradient systems like* $\dot{x}(t)=-\operatorname{proj}_{C(x(t))}(0)$ *it is not trivial to show uniqueness of trajectories in the general setting. The second problem is the implicit structure of the equation* (MTRIGS)*. Therefore, we cannot use standard arguments like the Cauchy–Lipschitz Theorem to derive the uniqueness of solutions. Note that the asymptotic analysis presented in this section applies to any trajectory solution* $x(\cdot)$ *of* (MTRIGS)*, which reduces the importance of the uniqueness statement.*

4.6.4 Preparatory results

In this subsection, we derive some properties that all trajectory solution $x(\cdot)$ of the system (MTRIGS) share.

Proposition 4.6.16. *Let* $x(\cdot)$ *be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, for all* $i=1,\dots,m$ *and almost all* $t\ge t_0$ *it holds*

$$\left\langle \nabla f_i(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t)+\frac{\alpha}{t^q}\dot{x}(t),\dot{x}(t)\right\rangle\le 0,$$

and therefore

$$\left\langle \nabla f_i(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t),\dot{x}(t)\right\rangle\le -\frac{\alpha}{t^q}\|\dot{x}(t)\|^2.$$

Proof. According to Definition 4.6.13, each solution $x(\cdot)$ satisfies

$$-\frac{\alpha}{t^q}\dot{x}(t)=\operatorname*{proj}_{C(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t)}(0),$$

for almost all $t\ge t_0$. From the variational characterization of the projection, it follows that

$$\left\langle \nabla f_i(x(t))+\frac{\beta}{t^p}x(t)+\ddot{x}(t)+\frac{\alpha}{t^q}\dot{x}(t),\frac{\alpha}{t^q}\dot{x}(t)\right\rangle\le 0,$$

for almost all $t\ge t_0$ and all $i=1,\dots,m$, which leads to the desired inequality. □

In the next proposition, we define component-wise energy functions for the system (MTRIGS) and show they fulfill a general decay property.

Proposition 4.6.17. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. For all $i = 1, \ldots, m$, define the energy function*

$$\mathcal{W}_i : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{W}_i(t) := f_i(x(t)) + \frac{\beta}{2t^p}\|x(t)\|^2 + \frac{1}{2}\|\dot{x}(t)\|^2. \tag{4.139}$$

Then, for all $i = 1, \ldots, m$ and almost all $t \geq t_0$ it holds

$$\frac{d}{dt}\mathcal{W}_i(t) \leq -\frac{p\beta}{2t^{p+1}}\|x(t)\|^2 - \frac{\alpha}{t^q}\|\dot{x}(t)\|^2 \leq 0.$$

Further, for $a \in \mathbb{R}^m_+$ defined as $a_i := \frac{\beta}{2t_0^p}\|x(t_0)\|^2 + \frac{1}{2}\|\dot{x}(t_0)\|^2$ for $i = 1, \ldots, m$, it holds

$$x(t) \in \mathcal{L}(F, F(x(t_0)) + a) \quad \textit{for all} \quad t \geq t_0.$$

Proof. According to Definition 4.6.13, the velocity $\dot{x}(\cdot)$ of a trajectory solution is differentiable almost everywhere. For all $i = 1, \ldots, m$ and almost all $t \geq t_0$ it holds

$$\begin{aligned}
\frac{d}{dt}\mathcal{W}_i(t) &= \langle \nabla f_i(x(t)), \dot{x}(t)\rangle - \frac{p\beta}{2t^{p+1}}\|x(t)\|^2 + \frac{\beta}{t^p}\langle x(t), \dot{x}(t)\rangle + \langle \dot{x}(t), \ddot{x}(t)\rangle \\
&= -\frac{p\beta}{2t^{p+1}}\|x(t)\|^2 + \left\langle \nabla f_i(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t), \dot{x}(t)\right\rangle \\
&\leq -\frac{p\beta}{2t^{p+1}}\|x(t)\|^2 - \frac{\alpha}{t^q}\|\dot{x}(t)\|^2 \leq 0,
\end{aligned}$$

where the last inequality follows from Proposition 4.6.16. The last statement of the proposition follows using the monotonicity of $\mathcal{W}_i(\cdot)$ for $i = 1, \ldots, m$, on $[t_0, +\infty)$. ☐

For all $t \geq t_0$ we have $\operatorname{proj}_{C(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t)}(0) \in C(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t)$ and hence there exists $\theta(t) \in \Delta^m$ with

$$-\frac{\alpha}{t^q}\dot{x}(t) = \operatorname*{proj}_{C(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t)}(0) = \sum_{i=1}^m \theta_i(t)\nabla f_i(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t). \tag{4.140}$$

In the following proposition, we show that there exists a measurable function $\theta(\cdot)$ satisfying (4.140).

Proposition 4.6.18. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, there exists a measurable function*

$$\theta : [t_0, +\infty) \to \Delta^m, \quad t \mapsto \theta(t),$$

with

$$-\frac{\alpha}{t^q}\dot{x}(t) = \operatorname*{proj}_{C(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t)}(0) = \sum_{i=1}^m \theta_i(t)\nabla f_i(x(t)) + \frac{\beta}{t^p}x(t) + \ddot{x}(t), \tag{4.141}$$

for all $t \in [t_0, +\infty)$.

Proof. The proof follows the lines of the proof of Lemma 4.5.8, where a similar result is shown for the system (MAVD). For almost all $t \geq t_0$, there exists $\theta(t) \in \Delta^m$ such that

$$\theta(t) \in \underset{\theta \in \Delta^m}{\arg\min}\, j(t,\theta), \text{ where } j(t,\theta) := \left\| \sum_{i=1}^m \theta_i \nabla f_i(x(t)) + \frac{\beta}{t^p} x(t) + \ddot{x}(t) \right\|^2 . \tag{4.142}$$

The existence of a measurable selection $\theta : [t_0, +\infty) \to \Delta^m$, $t \mapsto \theta(t) \in \arg\min_{\theta \in \Delta^m} j(t,\theta)$ can be verified using [204, Theorem 14.37]. To this end, we have to show that $j(\cdot,\cdot)$ is a Carathéodory integrand, i.e., $j(\cdot,\theta)$ is measurable for all $\theta \in \Delta^m$ and $j(t,\cdot)$ is continuous for all $t \geq t_0$. The second condition is obviously satisfied. Since $x(\cdot)$ is a trajectory solution of (MTRIGS) in the sense of Definition 4.6.13, $\ddot{x}(\cdot)$ is (Bochner) measurable. Hence, for all $\theta \in \Delta^m$, $j(\theta,\cdot)$ is measurable as a composition of a measurable and a continuous function. This demonstrates that the first condition is also satisfied. □

By using the weight function $\theta(\cdot)$ we can give a further variational characterization of a trajectory solution of (MTRIGS).

Proposition 4.6.19. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13 and $\theta(\cdot)$ the corresponding measurable weight function given by Proposition 4.6.18. Then, for all $i = 1, \ldots, m$ and almost all $t \geq t_0$ it holds that*

$$\langle \nabla f_i(x(t)), \dot{x}(t) \rangle \leq \left\langle \sum_{i=1}^m \theta_i(t) \nabla f_i(x(t)), \dot{x}(t) \right\rangle .$$

Proof. By Proposition 4.6.16, we have for all $i = 1, \ldots, m$ and almost all $t \geq t_0$

$$\left\langle \nabla f_i(x(t)) + \frac{\beta}{t^p} x(t) + \ddot{x}(t) + \frac{\alpha}{t^q} \dot{x}(t), \dot{x}(t) \right\rangle \leq 0, \tag{4.143}$$

which, combined with (4.141), yields

$$\langle \nabla f_i(x(t)), \dot{x}(t) \rangle \leq \left\langle \sum_{i=1}^m \theta_i(t) \nabla f_i(x(t)), \dot{x}(t) \right\rangle .$$

□

We conclude this part on the preparatory results with the following proposition.

Proposition 4.6.20. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, the following statements are true:*

i) $\dot{x}(\cdot)$ is bounded;

ii) If $x(\cdot)$ is bounded, then $\ddot{x}(\cdot)$ is essentially bounded.

Proof. *i)* According to Proposition 4.6.17, we have for all $i = 1, \ldots, m$ and all $t \geq t_0$

$$\frac{1}{2} \|\dot{x}(t)\|^2 \leq \mathcal{W}_i(t) \leq \mathcal{W}_i(t_0),$$

which proves the first statement.
ii) If $x(\cdot)$ is bounded, then $\nabla f_i(x(\cdot))$ is also bounded for all $i = 1, \dots, m$, as a consequence of the Lipschitz continuity of the gradients. According to (MTRIGS), we have for almost all $t \geq t_0$

$$\ddot{x}(t) + \frac{\alpha}{t^q}\dot{x}(t) = \operatorname*{proj}_{C(x(t))+\frac{\beta}{t^p}x(t)}(-\ddot{x}(t)),$$

hence,

$$\|\ddot{x}(t)\| \leq \frac{\alpha}{t^q}\|\dot{x}(t)\| + \left\|\operatorname*{proj}_{C(x(t))+\frac{\beta}{t^p}x(t)}(-\ddot{x}(t))\right\|. \tag{4.144}$$

Since all expressions on the right hand side of (4.144) are bounded on $[t_0, +\infty)$, $\ddot{x}(\cdot)$ is essentially bounded. □

A general energy function

The following energy functions are the key to the asymptotic analysis of (MTRIGS).

Definition 4.6.21. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, $r \in [q, 1]$ and $z \in \mathcal{H}$. Let $\gamma : [t_0, +\infty) \to [0, +\infty)$ and $\xi : [t_0, +\infty) \to \mathbb{R}$ be continuously differentiable functions. Define for $i = 1, \dots, m$*

$$\mathcal{G}^r_{i,\gamma,\xi,z}(t) := t^{2r}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + \frac{1}{2}\|\gamma(t)(x(t) - z) + t^r\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z\|^2$$

and

$$\mathcal{G}^r_{\gamma,\xi,z}(t) := t^{2r}\min_{i=1,\dots,m}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + \frac{1}{2}\|\gamma(t)(x(t) - z) + t^r\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z\|^2.$$

For $z(t) := \arg\min_{z\in\mathcal{H}}\max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ for $t \geq t_0$, we define

$$\begin{aligned}
\mathcal{G}^r_{\gamma,\xi} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{G}^r_{\gamma,\xi,z(t)}(t) =\ & t^{2r}\min_{i=1,\dots,m}\left(f_{t,i}(x(t)) - f_{t,i}(z(t))\right)\\
&+ \frac{1}{2}\|\gamma(t)(x(t) - z(t)) + t^r\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z(t)\|^2.\\
=\ & t^{2r}\varphi_t(x(t))\\
&+ \frac{1}{2}\|\gamma(t)(x(t) - z(t)) + t^r\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z(t)\|^2.
\end{aligned}$$

The functions $\gamma(\cdot)$ and $\xi(\cdot)$ will be specified at a later point in the analysis. In the next proposition, we derive estimates for the derivatives of the energy functions introduced above.

Proposition 4.6.22. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, $r \in [q, 1]$ and $z \in \mathcal{H}$. Let $\gamma : [t_0, +\infty) \to [0, +\infty)$ and $\xi : [t_0, +\infty) \to \mathbb{R}$ be continuously differentiable functions.*

i) For all $i = 1, \ldots, m$, the function $\mathcal{G}^r_{i,\gamma,\xi,z}(\cdot)$ is absolutely continuous on every interval $[t_0, T]$ for $T \geq t_0$, differentiable almost everywhere on $[t_0, +\infty)$, and its derivative satisfies for almost all $t \in [t_0, +\infty)$

$$\begin{aligned}\frac{d}{dt}\mathcal{G}^r_{i,\gamma,\xi,z}(t) \leq\ & 2rt^{2r-1}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\ & - t^r\gamma(t) \min_{i=1,\ldots,m}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) \\ & + \big(\gamma(t)(\gamma(t) + rt^{r-1} - \alpha t^{r-q}) + t^r\gamma'(t) + \xi(t)\big)\left\langle x(t) - z, \dot{x}(t)\right\rangle \\ & + \left(\gamma(t)\gamma'(t) + \frac{\xi'(t)}{2} - \gamma(t)t^r\frac{\beta}{2t^p}\right)\|x(t) - z\|^2 \\ & + t^r(\gamma(t) + rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2.\end{aligned} \tag{4.145}$$

ii) The function $\mathcal{G}^r_{\gamma,\xi,z}(\cdot)$ is absolutely continuous on every interval $[t_0, T]$ for $T \geq t_0$, differentiable almost everywhere on $[t_0, +\infty)$, and its derivative satisfies for almost all $t \in [t_0, +\infty)$

$$\begin{aligned}\frac{d}{dt}\mathcal{G}^r_{\gamma,\xi,z}(t) \leq\ & \big(2rt^{2r-1} - t^r\gamma(t)\big)\min_{i=1,\ldots,m}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\ & + \big(\gamma(t)(\gamma(t) + rt^{r-1} - \alpha t^{r-q}) + t^r\gamma'(t) + \xi(t)\big)\left\langle x(t) - z, \dot{x}(t)\right\rangle \\ & + \left(\gamma(t)\gamma'(t) + \frac{\xi'(t)}{2} - \gamma(t)t^r\frac{\beta}{2t^p}\right)\|x(t) - z\|^2 \\ & + t^r(\gamma(t) + rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2.\end{aligned} \tag{4.146}$$

Proof. Fix an arbitrary $i \in \{1, \ldots, m\}$. It is obvious that $\mathcal{G}^r_{i,\gamma,\xi,z}(\cdot)$ is absolutely continuous on every interval $[t_0, T]$ for $T \geq t_0$ and therefore differentiable almost everywhere on $[t_0, +\infty)$. Let $t \geq t_0$ be a point at which $\mathcal{G}^r_{i,\gamma,z}(\cdot)$ is differentiable. By the chain rule, it holds that

$$\begin{aligned}\frac{d}{dt}\mathcal{G}^r_{i,\gamma,\xi,z}(t) =\ & 2rt^{2r-1}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + t^{2r}\langle \nabla f_{t,i}(x(t)), \dot{x}(t)\rangle - \frac{p\beta t^{2r}}{2t^{p+1}}\|x(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\ & + \big\langle \gamma(t)(x(t) - z) + t^r\dot{x}(t), (\gamma(t) + rt^{r-1})\dot{x}(t) + \gamma'(t)(x(t) - z) + t^r\ddot{x}(t)\big\rangle \\ & + \xi(t)\langle x(t) - z, \dot{x}(t)\rangle + \frac{\xi'(t)}{2}\|x(t) - z\|^2.\end{aligned}$$

Let $\theta(\cdot)$ be the measurable weight function given by Proposition 4.6.18. By Proposition 4.6.19, we have

$$\begin{aligned}&\frac{d}{dt}\mathcal{G}^r_{i,\gamma,\xi,z}(t) \leq \\ & 2rt^{2r-1}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + t^{2r}\left\langle \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t)), \dot{x}(t)\right\rangle + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\ & + \big\langle \gamma(t)(x(t) - z) + t^r\dot{x}(t), (\gamma(t) + rt^{r-1})\dot{x}(t) + \gamma'(t)(x(t) - z) + t^r\ddot{x}(t)\big\rangle \\ & + \xi(t)\langle x(t) - z, \dot{x}(t)\rangle + \frac{\xi'(t)}{2}\|x(t) - z\|^2.\end{aligned} \tag{4.147}$$

Using (4.141), we write

$$t^r \ddot{x}(t) = -\alpha t^{r-q}\dot{x}(t) - t^r \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t)),$$

which we use to evaluate

$$\begin{aligned}
&\big\langle \gamma(t)(x(t)-z) + t^r\dot{x}(t), (\gamma(t)+rt^{r-1})\dot{x}(t) + \gamma'(t)(x(t)-z) + t^r\ddot{x}(t)\big\rangle \\
&= \gamma(t)(\gamma(t)+rt^{r-1}-\alpha t^{r-q})\left\langle x(t)-z,\dot{x}(t)\right\rangle + \gamma(t)\gamma'(t)\|x(t)-z\|^2 \\
&\quad - t^r\gamma(t)\left\langle x(t)-z, \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t))\right\rangle - t^{2r}\left\langle \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t)), \dot{x}(t)\right\rangle \\
&\quad + t^r(\gamma(t)+rt^{r-1}-\alpha t^{r-q})\|\dot{x}(t)\|^2 + t^r\gamma'(t)\langle \dot{x}(t), x(t)-z\rangle \\
&= \left[\gamma(t)(\gamma(t)+rt^{r-1}-\alpha t^{r-q}) + t^r\gamma'(t)\right]\langle x(t)-z,\dot{x}(t)\rangle + \gamma(t)\gamma'(t)\|x(t)-z\|^2 \\
&\quad - t^r\gamma(t)\left\langle x(t)-z, \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t))\right\rangle + t^r(\gamma(t)+rt^{r-1}-\alpha t^{r-q})\|\dot{x}(t)\|^2 \\
&\quad - t^{2r}\left\langle \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t)), \dot{x}(t)\right\rangle .
\end{aligned} \tag{4.148}$$

We combine (4.147) and (4.148) to derive

$$\begin{aligned}
&\frac{d}{dt}\mathcal{G}^r_{i,\gamma,\xi,z}(t) \leq \\
&2rt^{2r-1}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + t^{2r}\left\langle \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t)), \dot{x}(t)\right\rangle + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\
&+ \big(\gamma(t)(\gamma(t)+rt^{r-1}-\alpha t^{r-q}) + t^r\gamma'(t)\big)\langle x(t)-z,\dot{x}(t)\rangle + \gamma(t)\gamma'(t)\|x(t)-z\|^2 \\
&- t^r\gamma(t)\left\langle x(t)-z, \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t))\right\rangle + t^r(\gamma(t)+rt^{r-1}-\alpha t^{r-q})\|\dot{x}(t)\|^2 \\
&- t^{2r}\left\langle \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t)), \dot{x}(t)\right\rangle + \xi(t)\langle x(t)-z,\dot{x}(t)\rangle + \frac{\xi'(t)}{2}\|x(t)-z\|^2 \\
&= 2rt^{2r-1}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + \left(\gamma(t)\gamma'(t) + \frac{\xi'(t)}{2}\right)\|x(t)-z\|^2 \\
&+ \big(\gamma(t)(\gamma(t)+rt^{r-1}-\alpha t^{r-q}) + t^r\gamma'(t) + \xi(t)\big)\langle x(t)-z,\dot{x}(t)\rangle \\
&+ t^r\gamma(t)\left\langle z-x(t), \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t))\right\rangle + t^r(\gamma(t)+rt^{r-1}-\alpha t^{r-q})\|\dot{x}(t)\|^2.
\end{aligned} \tag{4.149}$$

By strong convexity of $x \mapsto \sum_{i=1}^m \theta_i(t)(f_{t,i}(x) - f_{t,i}(z))$ we have

$$\begin{aligned}
\left\langle z-x(t), \sum_{i=1}^m \theta_i(t)\nabla f_{t,i}(x(t))\right\rangle &\leq \sum_{i=1}^m \theta_i(t)\left(f_{t,i}(z) - f_{t,i}(x(t))\right) - \frac{\beta}{2t^p}\|x(t)-z\|^2 \\
&\leq -\min_{i=1,\dots,m} f_{t,i}(x(t)) - f_{t,i}(z) - \frac{\beta}{2t^p}\|x(t)-z\|^2.
\end{aligned} \tag{4.150}$$

Plugging (4.149) into (4.150) gives

$$\begin{aligned}
&\frac{d}{dt}\mathcal{G}^r_{i,\gamma,\xi,z}(t) \leq \\
&2rt^{2r-1}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) - t^r\gamma(t)\min_{i=1,\dots,m}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) - \gamma(t)t^r\frac{\beta}{2t^p}\|x(t)-z\|^2 \\
&+ \big(\gamma(t)(\gamma(t) + rt^{r-1} - \alpha t^{r-q}) + t^r\gamma'(t) + \xi(t)\big)\langle x(t)-z, \dot{x}(t)\rangle \\
&+ \left(\gamma(t)\gamma'(t) + \frac{\xi'(t)}{2}\right)\|x(t)-z\|^2 + t^r(\gamma(t) + rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2,
\end{aligned}$$

concluding part *i*). Statement *ii*) follows immediately from *i*) and Lemma 2.2.14. □

In the asymptotical analysis, we do also use the following special instance of $\mathcal{G}^r_{i,\gamma,\xi,z}(\cdot)$.

Definition 4.6.23. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Let $\lambda > 0$ and $r \in [q, 1]$ and define*

$$\begin{aligned}
&\gamma : [t_0, +\infty) \to [0, +\infty), \quad t \mapsto \gamma(t) := \lambda, \quad \textit{and} \\
&\xi : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \xi(t) := \lambda\left(rt^{r-1} + \alpha t^{r-q} - 2\lambda\right).
\end{aligned}$$

For this choice of the two parameter functions, we rename the energy functions as follows:

$$\begin{aligned}
\mathcal{E}^r_{i,\lambda,z} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}^r_{i,\lambda,z}(t) &:= \mathcal{G}^r_{i,\gamma,\xi,z}(t) \\
&= t^{2r}(f_{t,i}(x(t)) - f_{t,i}(z)) + \frac{1}{2}\|\lambda(x(t)-z) + t^r\dot{x}(t)\|^2 \\
&+ \frac{\lambda}{2}\left(rt^{r-1} + \alpha t^{r-q} - 2\lambda\right)\|x(t)-z\|^2,
\end{aligned}$$

for $i = 1, \dots, m$,

$$\begin{aligned}
\mathcal{E}^r_{\lambda,z} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}^r_{\lambda,z}(t) &:= \mathcal{G}^r_{\gamma,\xi,z}(t) \\
&= t^{2r}\min_{i=1,\dots,m}(f_{t,i}(x(t)) - f_{t,i}(z)) + \frac{1}{2}\|\lambda(x(t)-z) + t^r\dot{x}(t)\|^2 \\
&+ \frac{\lambda}{2}\left(rt^{r-1} + \alpha t^{r-q} - 2\lambda\right)\|x(t)-z\|^2,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{E}^r_{\lambda} : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \mathcal{E}^r_{\lambda}(t) &:= \mathcal{G}^r_{\gamma,\xi}(t) \\
&= t^{2r}\min_{i=1,\dots,m}(f_{t,i}(x(t)) - f_{t,i}(z(t))) + \frac{1}{2}\|\lambda(x(t)-z(t)) + t^r\dot{x}(t)\|^2 \\
&+ \frac{\lambda}{2}\left(rt^{r-1} + \alpha t^{r-q} - 2\lambda\right)\|x(t)-z(t)\|^2 \\
&= t^{2r}\varphi_t(x(t)) + \frac{1}{2}\|\lambda(x(t)-z(t)) + t^r\dot{x}(t)\|^2 \\
&+ \frac{\lambda}{2}\left(rt^{r-1} + \alpha t^{r-q} - 2\lambda\right)\|x(t)-z(t)\|^2,
\end{aligned}$$

where $z(t) := \arg\min_{z\in\mathcal{H}}\max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ for $t \geq t_0$.

In the following, we formulate a proposition on $\mathcal{E}^r_{i,\lambda,z}(\cdot)$ and $\mathcal{E}^r_{\lambda,z}(\cdot)$ similar to Proposition 4.6.22.

Proposition 4.6.24. *Let $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, $\lambda > 0$, $r \in [q, 1]$ and $z \in \mathcal{H}$.*

i) For all $i = 1, \dots, m$, the function $\mathcal{E}^r_{i,\lambda,z}(\cdot)$ is absolutely continuous on every interval $[t_0, T]$ for $T \geq t_0$, differentiable almost everywhere on $[t_0, +\infty)$, and its derivative satisfies for almost all $t \in [t_0, +\infty)$

$$\begin{aligned} &\frac{d}{dt}\mathcal{E}^r_{i,\lambda,z}(t) \leq \\ &2rt^{2r-1}(f_{t,i}(x(t)) - f_{t,i}(z) - \lambda t^r \min_{i=1,\dots,m}(f_{t,i}(x(t)) - f_{t,i}(z)) + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\ &+ \lambda\left(2rt^{r-1} - \lambda\right)\langle x(t) - z, \dot{x}(t)\rangle + t^r\left(\lambda + rt^{r-1} - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 \\ &+ \frac{\lambda}{2}\left(r(r-1)t^{r-2} + \alpha(r-q)t^{r-q-1} - \beta t^{r-p}\right)\|x(t) - z\|^2. \end{aligned} \tag{4.151}$$

ii) The functions $\mathcal{E}^r_{\lambda,z}(\cdot)$ is absolutely continuous on every interval $[t_0, T]$ for $T \geq t_0$, differentiable almost everywhere on $[t_0, +\infty)$, and its derivative satisfies for almost all $t \in [t_0, +\infty)$

$$\begin{aligned} \frac{d}{dt}\mathcal{E}^r_{\lambda,z}(t) \leq &\left(2rt^{2r-1} - \lambda t^r\right)\min_{i=1,\dots,m}(f_{t,i}(x(t)) - f_{t,i}(z)) + + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\ &+ \lambda\left(2rt^{r-1} - \lambda\right)\langle x(t) - z, \dot{x}(t)\rangle + t^r\left(\lambda + rt^{r-1} - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 \\ &+ \frac{\lambda}{2}\left(r(r-1)t^{r-2} + \alpha(r-q)t^{r-q-1} - \beta t^{r-p}\right)\|x(t) - z\|^2. \end{aligned} \tag{4.152}$$

Proof. The proof follows immediately by Proposition 4.6.22 using $\gamma'(t) = 0$ and $\xi'(t) = \lambda(r(r-1)t^{r-2} + \alpha(r-q)t^{r-q-1})$ for $t \geq t_0$. □

Lemma 4.6.25. *Let $q \in (0, 1)$, $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, $\lambda > 0$, $r \in [q, 1)$, and $z \in \mathcal{H}$. Define $\mu_r : [t_0, +\infty) \to \mathbb{R}$, $\mu_r(t) := \frac{\lambda}{t^r} - \frac{2r}{t}$. Then, for almost all $t \geq t_1 := \max\left(\left(\frac{2r}{\lambda}\right)^{\frac{1}{1-r}}, t_0\right)$, it holds that*

$$\begin{aligned} &\frac{d}{dt}\mathcal{E}^r_{\lambda,z}(t) + \mu_r(t)\mathcal{E}^r_{\lambda,z}(t) \\ \leq &t^r\left(\frac{3}{2}\lambda - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + \frac{\lambda}{2}\left[\frac{3\lambda r}{t} - \frac{\lambda^2}{t^r} + \frac{\lambda\alpha}{t^q} - \frac{\beta}{t^{p-r}}\right]\|x(t) - z\|^2. \end{aligned} \tag{4.153}$$

Proof. For all $t \geq t_0$ it holds

$$\begin{aligned} \mathcal{E}^r_{\lambda,z}(t) = &t^{2r}\min_{i=1,\dots,m}(f_{t,i}(x(t)) - f_{t,i}(z)) + \frac{\lambda^2}{2}\|x(t) - z\|^2 + \lambda t^r\langle x(t) - z, \dot{x}(t)\rangle \\ &+ \frac{t^{2r}}{2}\|\dot{x}(t)\|^2 + \frac{\lambda}{2}\left(rt^{r-1} + \alpha t^{r-q} - 2\lambda\right)\|x(t) - z\|^2 \\ = &t^{2r}\min_{i=1,\dots,m}(f_{t,i}(x(t)) - f_{t,i}(z)) + \frac{\lambda}{2}\left(rt^{r-1} + \alpha t^{r-q} - \lambda\right)\|x(t) - z\|^2 \\ &+ \lambda t^r\langle x(t) - z, \dot{x}(t)\rangle + \frac{t^{2r}}{2}\|\dot{x}(t)\|^2. \end{aligned} \tag{4.154}$$

Note that $\mu_r(t) \geq 0$ for all $t \geq \left(\frac{2r}{\lambda}\right)^{\frac{1}{1-r}}$. Then, combining (4.152) and (4.154), it yields for almost all $t \geq t_1$

$$
\begin{aligned}
&\frac{d}{dt}\mathcal{E}^r_{\lambda,z}(t) + \mu_r(t)\mathcal{E}^r_{\lambda,z}(t) \leq \\
&\quad \left(2rt^{2r-1} - \lambda t^r\right) \min_{i=1,\dots,m} \left(f_{t,i}(x(t)) - f_{t,i}(z)\right) + t^r\left(\lambda + rt^{r-1} - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 \\
&\quad + \frac{\lambda}{2}\left(r(r-1)t^{r-2} + \alpha(r-q)t^{r-q-1} - \beta t^{r-p}\right)\|x(t)-z\|^2 \\
&\quad + \lambda\left(2rt^{r-1} - \lambda\right)\langle x(t) - z, \dot{x}(t)\rangle + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\
&\quad + \left(\lambda t^r - 2rt^{2r-1}\right) \min_{i=1,\dots,m}\left(f_{t,i}(x(t)) - f_{t,i}(z)\right) \\
&\quad + \frac{\lambda}{2}\left[\frac{3\lambda r}{t} + \frac{\lambda\alpha}{t^q} - \frac{\lambda^2}{t^r} - \frac{2r^2}{t^{2-r}} - \frac{2r\alpha}{t^{1-r+q}}\right]\|x(t)-z\|^2 \\
&\quad + \lambda\left(\lambda - 2rt^{r-1}\right)\langle x(t)-z, \dot{x}(t)\rangle + \frac{1}{2}\left(\lambda t^r - 2rt^{2r-1}\right)\|\dot{x}(t)\|^2 \\
&= t^r\left(\frac{3}{2}\lambda - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 \\
&\quad + \frac{\lambda}{2}\left[-\frac{r(r+1)}{t^{2-r}} - \frac{\alpha(r+q)}{t^{1-r+q}} + \frac{3\lambda r}{t} + \frac{\lambda\alpha}{t^q} - \frac{\lambda^2}{t^r} - \beta t^{r-p}\right]\|x(t)-z\|^2 \\
&\leq t^r\left(\frac{3}{2}\lambda - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + \frac{\lambda}{2}\left[\frac{3\lambda r}{t} - \frac{\lambda^2}{t^r} + \frac{\lambda\alpha}{t^q} - \frac{\beta}{t^{p-r}}\right]\|x(t)-z\|^2.
\end{aligned}
$$

□

The result above can be extended to the case $q \in (0, 1]$ and $r = 1$ for $\lambda \geq 2$ as we state in the following lemma.

Lemma 4.6.26. *Let $q \in (0,1]$, $x(\cdot)$ be a trajectory solution of* (MTRIGS), *$\lambda \geq 2$, $r = 1$ and $z \in \mathcal{H}$. Define $\mu_1 : [t_0, +\infty) \to \mathbb{R}$, $t \mapsto \mu_1(t) := \frac{\lambda-2}{t}$. Then, for almost all $t \geq t_0$, it holds that*

$$
\begin{aligned}
\frac{d}{dt}\mathcal{E}^1_{\lambda,z}(t) + \mu_1(t)\mathcal{E}^1_{\lambda,z}(t) \leq t\left(\frac{3}{2}\lambda - \alpha t^{1-q}\right)\|\dot{x}(t)\|^2 + \frac{p\beta}{2t^{p-1}}\|z\|^2 \\
+ \frac{\lambda}{2}\left[\frac{(1-\lambda)(\lambda-2)}{t} + \frac{\alpha(\lambda-(1+q))}{t^q} - \frac{\beta}{t^{p-1}}\right]\|x(t)-z\|^2.
\end{aligned}
\tag{4.155}
$$

Proof. The proof is analogous to the one of Lemma 4.6.25. □

4.6.5 Asymptotic analysis

In this subsection, we study the asymptotic behavior of the trajectory solutions to (MTRIGS). The convergence rates for the merit function values and the convergence of the trajectory depend heavily on the parameters $p \in (0, 2], q \in (0, 1]$ and $\alpha, \beta > 0$.

The case $p \in (0,2]$ and $q < \frac{p}{2}$: convergence rates

In Theorem 4.6.27, we derive convergence rates for the merit function along trajectory solutions of (MTRIGS) when the parameters p and q satisfy $p \in (0,2]$ and $q < \frac{p}{2}$.

Theorem 4.6.27. *Let $p \in (0,2]$ with $q < \frac{p}{2}$, $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, and $z(t) := \arg\min_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ for $t \geq t_0$. Then, we have the following convergence rates as $t \to +\infty$:*

i) $\mathcal{E}_{\lambda}^{q}(t) = \mathcal{O}(1)$ *for* $0 < \lambda < \frac{\alpha}{2}$*;*

ii) $\varphi_t(x(t)) = \mathcal{O}\left(t^{-2q}\right);$

iii) $\varphi(x(t)) = \mathcal{O}\left(t^{-2q}\right);$

iv) $\|x(t) - z(t)\| = \mathcal{O}(1);$

v) $\|\dot{x}(t)\| = \mathcal{O}\left(t^{-q}\right).$

Proof. i) Let $0 < \lambda < \frac{\alpha}{2}$ and $z \in \mathcal{H}$ fixed. We derive a bound for the energy function $\mathcal{E}_{\lambda,z}^{q}(\cdot)$ by considering inequality (4.153) with $r = q$, i.e., for almost all $t \geq \max\left(\left(\frac{2q}{\lambda}\right)^{\frac{1}{1-q}}, t_0\right)$

$$\begin{aligned}&\frac{d}{dt}\mathcal{E}_{\lambda,z}^{q}(t) + \mu_q(t)\mathcal{E}_{\lambda,z}^{q}(t) \leq \\ &t^q\left(\frac{3}{2}\lambda - \alpha\right)\|\dot{x}(t)\|^2 + \frac{p\beta}{2}t^{2q-p-1}\|z\|^2 + \frac{\lambda}{2}\left[\frac{3\lambda q}{t} - \frac{\lambda^2}{t^q} + \frac{\lambda\alpha}{t^q} - \frac{\beta}{t^{p-q}}\right]\|x(t) - z\|^2.\end{aligned} \tag{4.156}$$

From here, we derive for almost all $t \geq \max\left(\left(\frac{2q}{\lambda}\right)^{\frac{1}{1-q}}, t_0, 1\right)$

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{\lambda,z}^{q}(t) + \mu_q(t)\mathcal{E}_{\lambda,z}^{q}(t) &\leq \frac{p\beta}{2}t^{2q-p-1}\|z\|^2 + \frac{\lambda^2(3+\alpha-\lambda)}{2t^q}\|x(t) - z\|^2 \\ &\leq \frac{p\beta}{2}t^{2q-p-1}\|z\|^2 + \lambda^2(3+\alpha-\lambda)t^{-q}\left(\|z\|^2 + \|x(t)\|^2\right).\end{aligned}$$

Since $x(\cdot)$ is bounded and $q < \frac{p}{2} \leq 1$, there exist $t_2 \geq \max\left(\left(\frac{2q}{\lambda}\right)^{\frac{1}{1-q}}, t_0, 1\right)$ and $c, M > 0$ such that for almost all $t \geq t_2$

$$\frac{d}{dt}\mathcal{E}_{\lambda,z}^{q}(t) + \mu_q(t)\mathcal{E}_{\lambda,z}^{q}(t) \leq c\left(M + \|z\|^2\right)t^{-q}. \tag{4.157}$$

We define the function

$$\begin{aligned}\mathfrak{M}_q : [t_2, +\infty) \to \mathbb{R}, \quad t \mapsto \mathfrak{M}_q(t) :=& \exp\left(\int_{t_2}^{t}\mu_q(s)ds\right) = \exp\left(\int_{t_2}^{t}\frac{\lambda}{s^q} - \frac{2q}{s}ds\right) \\ =& C_{\mathfrak{M}_q}\frac{\exp\left(\frac{\lambda}{1-q}t^{1-q}\right)}{t^{2q}},\end{aligned} \tag{4.158}$$

with $C_{\mathfrak{M}_q} = \frac{t_2^{2q}}{\exp\left(\frac{\lambda}{1-q}t_2^{1-q}\right)} > 0$. The function $\mathfrak{M}_q(\cdot)$ is constructed such that $\frac{d}{dt}\mathfrak{M}_q(t) = \mathfrak{M}_q(t)\mu_q(t)$ and hence

$$\frac{d}{dt}\left(\mathfrak{M}_q(t)\mathcal{E}^q_{\lambda,z}(t)\right) = \mathfrak{M}_q(t)\left(\frac{d}{dt}\mathcal{E}^q_{\lambda,z}(t) + \mu_q(t)\mathcal{E}^q_{\lambda,z}(t)\right) \quad \text{for almost all} \quad t \geq t_2. \tag{4.159}$$

The relations (4.159) and (4.157) give for almost all $t \geq t_2$

$$\frac{d}{dt}\left(\mathfrak{M}_q(t)\mathcal{E}^q_{\lambda,z}\right) \leq c\,\mathfrak{M}_q(t)\left(M + \|z\|^2\right)t^{-q}. \tag{4.160}$$

We integrate (4.160) from t_2 to $t \geq t_2$ to get

$$\mathfrak{M}_q(t)\mathcal{E}^q_{\lambda,z}(t) - \mathfrak{M}_q(t_2)\mathcal{E}^q_{\lambda,z}(t_2) \leq c\left(M + \|z\|^2\right)\int_{t_2}^t \mathfrak{M}_q(s)s^{-q}ds,$$

thus, for all $t \geq t_2$ it holds

$$\mathcal{E}^q_{\lambda,z}(t) \leq \frac{\mathfrak{M}_q(t_2)\mathcal{E}^q_{\lambda,z}(t_2)}{\mathfrak{M}_q(t)} + c\left(M + \|z\|^2\right)\frac{C_{\mathfrak{M}_q}}{\mathfrak{M}_q(t)}\int_{t_2}^t \exp\left(\frac{\lambda}{1-q}s^{1-q}\right)s^{-3q}ds. \tag{4.161}$$

The inequality above holds for all $z \in \mathcal{H}$ and all $t \geq t_2$. For all $t \geq t_2$, we choose

$$z := z(t) \in \operatorname*{arg\,min}_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t)),$$

which, since $\mathcal{E}^q_\lambda(t) = \mathcal{E}^q_{\lambda,z(t)}(t)$, yields

$$\mathcal{E}^q_\lambda(t) \leq \frac{\mathfrak{M}_q(t_2)\mathcal{E}^q_{\lambda,z(t)}(t_2)}{\mathfrak{M}_q(t)} + c\left(M + \|z(t)\|^2\right)\frac{C_{\mathfrak{M}_q}}{\mathfrak{M}_q(t)}\int_{t_2}^t \exp\left(\frac{\lambda}{1-q}s^{1-q}\right)s^{-3q}ds.$$

By Proposition 4.6.10, $z(\cdot)$ is bounded, and hence there exist constants $C_1, C_2 > 0$ such that for all $t \geq t_2$

$$\mathcal{E}^q_\lambda(t) \leq \frac{C_1}{\mathfrak{M}_q(t)} + \frac{C_2}{\mathfrak{M}_q(t)}\int_{t_2}^t \exp\left(\frac{\lambda}{1-q}s^{1-q}\right)s^{-3q}ds. \tag{4.162}$$

We apply Lemma 2.2.15 to the integral in (4.162) to derive the asymptotic bound

$$\int_{t_2}^t \exp\left(\frac{\lambda}{1-q}s^{1-q}\right)s^{-3q}ds = \mathcal{O}\left(t^{-2q}\exp\left(\frac{\lambda}{1-q}t^{1-q}\right)\right) \quad \text{as} \quad t \to +\infty,$$

hence

$$\frac{C_2}{\mathfrak{M}_q(t)}\int_{t_2}^t \exp\left(\frac{\lambda}{1-q}s^{1-q}\right)s^{-2q}ds = \mathcal{O}(1) \quad \text{as} \quad t \to +\infty. \tag{4.163}$$

We conclude from (4.162) and (4.163) that

$$\mathcal{E}^q_\lambda(t) = \mathcal{O}(1) \quad \text{as} \quad t \to +\infty, \tag{4.164}$$

proving statement *i)*. From here, we can prove the remaining four statements of the theorem.

ii) By the choice of $0 < \lambda < \frac{\alpha}{2}$, we have for all $t \geq t_0$

$$qt^{q-1} + \alpha - 2\lambda \geq 0.$$

Then, by the definition of $\mathcal{E}_\lambda^q(\cdot)$ we have for all $t \geq t_0$

$$t^{2q}\varphi_t(x(t)) \leq \mathcal{E}_\lambda^q(t),$$

which, according to (4.164), gives

$$\varphi_t(x(t)) = \mathcal{O}\left(t^{-2q}\right) \quad \text{as} \quad t \to +\infty.$$

iii) Using Proposition 4.6.12 and *ii)* yields

$$\varphi(x(t)) \leq \varphi_t(x(t)) + \frac{\beta R^2}{2t^p} = \mathcal{O}\left(t^{-2q}\right) \quad \text{as} \quad t \to +\infty.$$

iv) Since for all $t \geq t_0$

$$qt^{q-1} + \alpha - 2\lambda \geq \alpha - 2\lambda > 0,$$

it holds

$$\frac{\lambda}{2}(\alpha - 2\lambda)\|x(t) - z(t)\|^2 \leq \mathcal{E}_\lambda^q(t).$$

This estimate together with (4.164) implies that

$$\|x(t) - z(t)\| = \mathcal{O}(1) \quad \text{as} \quad t \to +\infty. \tag{4.165}$$

v) From *i)* and *iv)*, we have

$$\begin{aligned}\frac{t^{2q}}{2}\|\dot{x}(t)\|^2 &\leq \|\lambda(x(t) - z(t)) + t^q\dot{x}(t)\|^2 + \lambda^2\|x(t) - z(t)\|^2 \\ &\leq 2\mathcal{E}_\lambda^q(t) + \lambda^2\|x(t) - z(t)\|^2 = \mathcal{O}(1) \quad \text{as} \quad t \to +\infty.\end{aligned}$$

From here, we conclude

$$\|\dot{x}(t)\| = \mathcal{O}\left(t^{-q}\right) \quad \text{as} \quad t \to +\infty.$$

□

The case $q \in (0, 1)$ and $p < q + 1$: convergence rates and strong convergence of the trajectories

In this part, we perform the asymptotic analysis for (MTRIGS) for the case $p < q + 1$.

Theorem 4.6.28. *Let $q \in (0, 1)$ and $p < q + 1$, $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, and $z(t) := \arg\min_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ for $t \geq t_0$. Then, for $r \in [q, 1) \cap [p - q, 1)$, we have the following convergence rates as $t \to +\infty$:*

i) $\mathcal{E}^r_\lambda(t) = \mathcal{O}\left(t^{3r-(p+1)}\right)$ *for* $\lambda \in \left(0, \frac{2\alpha}{3}\right] \cap \left(0, \frac{\beta}{\alpha}\right]$;

ii) $\varphi_t(x(t)) = \mathcal{O}\left(t^{r-(p+1)}\right)$;

iii) $\varphi(x(t)) = \mathcal{O}\left(t^{-p}\right)$;

iv) $\|x(t) - z(t)\| = \mathcal{O}\left(t^{\frac{r-1}{2}}\right)$;

v) $\|\dot{x}(t)\| = \mathcal{O}\left(t^{\frac{r-(p+1)}{2}}\right)$.

Proof. i) Let $r \in [q, 1) \cap [p-q, 1)$ and $z \in \mathcal{H}$ fixed. From (4.153), we have for almost all $t \geq \max\left(\left(\frac{2r}{\lambda}\right)^{\frac{1}{1-r}}, t_0\right)$

$$\begin{aligned}&\frac{d}{dt}\mathcal{E}^r_{\lambda,z}(t) + \mu_r(t)\mathcal{E}^r_{\lambda,z}(t) \leq \\ &t^r\left(\frac{3}{2}\lambda - \alpha t^{r-q}\right)\|\dot{x}(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + \frac{\lambda}{2}\left[\frac{3\lambda r}{t} - \frac{\lambda^2}{t^r} + \frac{\lambda\alpha}{t^q} - \frac{\beta}{t^{p-r}}\right]\|x(t) - z\|^2.\end{aligned} \tag{4.166}$$

Since $r < 1$, and $p - r \leq q$, $\lambda \leq \frac{\beta}{\alpha}$, and $r - q \geq 0$, $\lambda \leq \frac{2\alpha}{3}$ there exists $t_2 \geq \max\left(\left(\frac{2r}{\lambda}\right)^{\frac{1}{1-r}}, t_0\right)$ such that for almost all $t \geq t_2$

$$\frac{d}{dt}\mathcal{E}^r_{\lambda,z}(t) + \mu_q(t)\mathcal{E}^r_{\lambda,z}(t) \leq \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2. \tag{4.167}$$

As before, we define the function

$$\begin{aligned}\mathfrak{M}_r : [t_2, +\infty) \to \mathbb{R}, \quad t \mapsto \mathfrak{M}_r(t) &:= \exp\left(\int_{t_2}^t \mu_r(s)ds\right) = \exp\left(\int_{t_1}^t \frac{\lambda}{s^r} - \frac{2r}{s}ds\right) \\ &= C_{\mathfrak{M}_r}\frac{\exp\left(\frac{\lambda}{1-r}t^{1-r}\right)}{t^{2r}},\end{aligned} \tag{4.168}$$

with $C_{\mathfrak{M}_r} = \frac{t_2^{2r}}{\exp\left(\frac{\lambda}{1-r}t_2^{1-r}\right)} > 0$. The function $\mathfrak{M}_r(\cdot)$ is constructed such that $\frac{d}{dt}\mathfrak{M}_r(t) = \mathfrak{M}_r(t)\mu_r(t)$ and hence

$$\frac{d}{dt}\left(\mathfrak{M}_r(t)\mathcal{E}^r_{\lambda,z}(t)\right) = \mathfrak{M}_r(t)\left(\frac{d}{dt}\mathcal{E}^r_{\lambda,z}(t) + \mu_r(t)\mathcal{E}^r_{\lambda,z}(t)\right) \quad \text{for almost all} \quad t \geq t_2. \tag{4.169}$$

The relations (4.169) and (4.167) give for almost all $t \geq t_2$

$$\frac{d}{dt}\left(\mathfrak{M}_r(t)\mathcal{E}^r_{\lambda,z}(t)\right) \leq \frac{p\beta}{2}\|z\|^2\mathfrak{M}_r(t)t^{2r-(p+1)}. \tag{4.170}$$

We integrate (4.170) from t_2 to $t \geq t_2$ to get

$$\mathfrak{M}_r(t)\mathcal{E}^r_{\lambda,z}(t) - \mathfrak{M}_r(t_2)\mathcal{E}^r_{\lambda,z}(t_2) \leq \frac{p\beta}{2}\|z\|^2\int_{t_2}^t \mathfrak{M}_r(s)s^{2r-(p+1)}ds,$$

thus, for all $t \geq t_2$ it holds

$$\mathcal{E}^r_{\lambda,z}(t) \leq \frac{\mathfrak{M}_r(t_2)\mathcal{E}^r_{\lambda,z}(t_2)}{\mathfrak{M}_r(t)} + \frac{p\beta}{2}\|z\|^2 \frac{C_{\mathfrak{M}_r}}{\mathfrak{M}_r(t)} \int_{t_2}^t \exp\left(\frac{\lambda}{1-r}s^{1-r}\right) s^{-(p+1)} ds. \tag{4.171}$$

The inequality above holds for all $z \in \mathcal{H}$ and all $t \geq t_2$. For all $t \geq t_2$, we choose

$$z := z(t) \in \arg\min_{z \in \mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t)),$$

which, since $\mathcal{E}^r_\lambda(t) = \mathcal{E}^r_{\lambda,z(t)}(t)$, yields

$$\mathcal{E}^r_\lambda(t) \leq \frac{\mathfrak{M}_r(t_2)\mathcal{E}^r_{\lambda,z(t)}(t_2)}{\mathfrak{M}_r(t)} + \frac{p\beta}{2}\|z(t)\|^2 \frac{C_{\mathfrak{M}_r}}{\mathfrak{M}_r(t)} \int_{t_2}^t \exp\left(\frac{\lambda}{1-r}s^{1-r}\right) s^{-(p+1)} ds.$$

By Proposition 4.6.10, $z(\cdot)$ is bounded, hence there exist constants $C_1, C_2 > 0$ such that for all $t \geq t_2$

$$\mathcal{E}^r_\lambda(t) \leq \frac{C_1}{\mathfrak{M}_r(t)} + \frac{C_2}{\mathfrak{M}_r(t)} \int_{t_2}^t \exp\left(\frac{\lambda}{1-r}s^{1-r}\right) s^{-(p+1)} ds. \tag{4.172}$$

We apply Lemma 2.2.15 to the integral in (4.172) to derive the asymptotic bound

$$\int_{t_2}^t \exp\left(\frac{\lambda}{1-r}s^{1-r}\right) s^{-(p+1)} ds = \mathcal{O}\left(t^{r-(p+1)} \exp\left(\frac{\lambda}{1-r}t^{1-r}\right)\right) \quad \text{as} \quad t \to +\infty,$$

hence

$$\frac{C_2}{\mathfrak{M}_r(t)} \int_{t_2}^t \exp\left(\frac{\lambda}{1-r}s^{1-r}\right) s^{-(p+1)} ds = \mathcal{O}\left(t^{3r-(p+1)}\right) \quad \text{as} \quad t \to +\infty. \tag{4.173}$$

We conclude from (4.172) and (4.173) that

$$\mathcal{E}^r_\lambda(t) = \mathcal{O}\left(t^{3r-(p+1)}\right) \quad \text{as} \quad t \to +\infty, \tag{4.174}$$

proving statement *i)*. From here, we can prove the other four statements of the theorem.

ii) If $r > q$, for $t \geq \left(\frac{2\lambda}{\alpha}\right)^{\frac{1}{r-q}}$ we have $rt^{r-1} + \alpha t^{r-q} - 2\lambda \geq 0$ and hence

$$t^{2r}\varphi_t(x(t)) \leq \mathcal{E}^r_\lambda(t). \tag{4.175}$$

For the case $r = q$ the argument follows in a similar manner. We apply part *i)* for $\lambda \in \left(0, \frac{\alpha}{2}\right) \cap \left(0, \frac{\beta}{\alpha}\right] \subseteq \left(0, \frac{2\alpha}{3}\right] \cap \left(0, \frac{\beta}{\alpha}\right]$. Then, $qt^{q-1} + \alpha - 2\lambda \geq 0$ for all $t \geq t_0$ and hence

$$t^{2q}\varphi_t(x(t)) \leq \mathcal{E}^q_\lambda(t). \tag{4.176}$$

Both cases, together with (4.174), imply that for all $r \in [q, 1) \cap [p-q, 1)$

$$\varphi_t(x(t)) = \mathcal{O}\left(t^{r-(p+1)}\right) \quad \text{as} \quad t \to +\infty.$$

iii) Using Proposition 4.6.12 and *ii)* yields

$$\varphi(x(t)) \leq \varphi_t(x(t)) + \frac{\beta R^2}{2t^p} = \mathcal{O}\left(t^{-p}\right) \quad \text{as} \quad t \to +\infty.$$

iv) By Proposition 4.6.12, we have for all $t \geq t_0$

$$\|x(t) - z(t)\|^2 \leq \frac{2t^p}{\beta}\varphi_t(x(t)),$$

and hence by *ii)* we get

$$\|x(t) - z(t)\| = \mathcal{O}\left(t^{\frac{r-1}{2}}\right) \quad \text{as} \quad t \to +\infty. \tag{4.177}$$

v) From the above considerations, we have

$$\begin{aligned}\frac{t^{2r}}{2}\|\dot{x}(t)\|^2 &\leq \|\lambda(x(t) - z(t)) + t^r\dot{x}(t)\|^2 + \lambda^2\|x(t) - z(t)\|^2 \\ &\leq 2\mathcal{E}_\lambda^r(t) + \lambda^2\|x(t) - z(t)\|^2 = \mathcal{O}\left(t^{3r-(p+1)}\right) \quad \text{as} \quad t \to +\infty.\end{aligned}$$

From here, we conclude

$$\|\dot{x}(t)\| = \mathcal{O}\left(t^{\frac{r-(p+1)}{2}}\right) \quad \text{as} \quad t \to +\infty.$$

□

For this parameter settings, alongside establishing convergence rates, we demonstrate that the bounded trajectory solutions of (MTRIGS) converge strongly to a weakly Pareto optimal point of (MOP). Notably, this point is also the element of minimum norm within the lower level set of the objective function with respect to its value at the weakly Pareto optimal point.

Theorem 4.6.29. *Let $q \in (0, 1)$, $p < q + 1$, and $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, $x(t)$ converges strongly to a weakly Pareto optimal point x^* of* (MOP) *as $t \to +\infty$, which is the element of minimum norm in $\mathcal{L}(F, F(x^*))$.*

Proof. To prove the strong convergence of the trajectory solution $x(\cdot)$ we use Theorem 4.6.7, which states that $z(\cdot)$ converges strongly, in combination with Theorem 4.6.28 *iv)*, which states that $\|x(t) - z(t)\| \to 0$ as $t \to +\infty$. Since $x(\cdot)$ is bounded, it holds $\inf_{t \geq t_0} f_i(x(t)) > -\infty$ for $i = 1, \dots, m$, and therefore

$$\inf_{t \geq t_0} \mathcal{W}_i(t) = \inf_{t \geq t_0} \left(f_i(x(t)) + \frac{\beta}{2t^p}\|x(t)\|^2 + \frac{1}{2}\|\dot{x}(t)\|^2 \right) \geq \inf_{t \geq t_0} f_i(x(t)) > -\infty,$$

where $\mathcal{W}_i(\cdot)$ is the function introduced in (4.139). By Proposition 4.6.17, the function $\mathcal{W}_i(\cdot)$ is monotonically decreasing and therefore, $\lim_{t \to +\infty} \mathcal{W}_i(t)$ exists for $i = 1, \dots, m$. According to Theorem 4.6.28, $\dot{x}(t) \to 0$, hence $\frac{\beta}{2t^p}\|x(t)\|^2 + \frac{1}{2}\|\dot{x}(t)\|^2 \to 0$ as $t \to +\infty$. Thus, for $i = 1, \dots, m$,

$$\lim_{t \to +\infty} f_i(x(t)) = \lim_{t \to +\infty} \mathcal{W}_i(t) = \inf_{t \geq t_0} \mathcal{W}_i(t) > -\infty.$$

We denote by $F^* \coloneqq \lim_{t\to+\infty} F(x(t)) = \lim_{t\to+\infty} (f_1(x(t)), \dots, f_m(x(t))) \in \mathbb{R}^m$. We use Theorem 4.6.7 with $q(t) \coloneqq F(x(t))$ to conclude

$$z(t) \to x^* \coloneqq \operatorname*{proj}_{S(F^*)}(0) \quad \text{as} \quad t \to +\infty,$$

where $z(t) \coloneqq \arg\min_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ and $S(F^*) \coloneqq \arg\min_{z\in\mathcal{H}} \max_{i=1,\dots,m} (f_i(z) - f_i^*)$. According to Theorem 4.6.28, we have $\|x(t) - z(t)\| \to 0$, hence

$$x(t) \to x^* \quad \text{as} \quad t \to +\infty.$$

Since $\varphi(x(t)) \to 0$ as $t \to +\infty$, it yields $\varphi(x^*) = 0$, thus x^* is a weakly Pareto optimal point of (MOP). By continuity, $F^* = F(x^*)$ and, since x^* is a weakly Pareto optimal solution of (MOP), it holds $S(F^*) = \mathcal{L}(F, F(x^*))$. □

The case $p \in (0, 2]$ and $q = 1$

In this part, we consider the boundary case $q = 1$, allowing for $p \in (0, 2]$. The assumption we make for α is consistent with that made in the setting of inertial dynamics with vanishing damping in the scalar case, see [13, 218].

Theorem 4.6.30. *Let $p \in (0, 2]$, $q = 1$ and $\alpha \geq 3$, $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, and $z(t) \coloneqq \arg\min_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t))$ for $t \geq t_0$. Then, we have the following convergence rates as $t \to +\infty$:*

i) $\mathcal{E}^1_\lambda(t) = \mathcal{O}\left(t^{2-p}\right)$ *for* $\lambda \in \left[2, \frac{2\alpha}{3}\right]$;

ii) $\varphi_t(x(t)) = \mathcal{O}\left(t^{-p}\right)$;

iii) $\varphi(x(t)) = \mathcal{O}\left(t^{-p}\right)$;

iv) $\|x(t) - z(t)\| = \mathcal{O}(1)$;

v) $\|\dot{x}(t)\| = \mathcal{O}\left(t^{-\frac{p}{2}}\right)$.

Proof. i) Let $r = q = 1$ and $z \in \mathcal{H}$ fixed. We consider the energy function $\mathcal{E}^r_{\lambda,z}(\cdot)$. From inequality (4.155) we get for almost all $t \geq t_0$

$$\begin{aligned} &\frac{d}{dt}\mathcal{E}^1_{\lambda,z}(t) + \mu_1(t)\mathcal{E}^1_{\lambda,z}(t) \leq \\ &t\left(\frac{3}{2}\lambda - \alpha\right)\|\dot{x}(t)\|^2 + \frac{p\beta}{2t^{p-1}}\|z\|^2 + \frac{\lambda}{2}\left[\frac{\alpha(\lambda-2)}{t} - \frac{\beta}{t^{p-1}}\right]\|x(t) - z\|^2. \end{aligned} \tag{4.178}$$

Since $p - 1 \leq 1$, $\lambda \leq \frac{2\alpha}{3}$ and $x(\cdot)$ is bounded, there exist $t_1 \geq t_0$ and $M, c > 0$ such that for almost all $t \geq t_1$

$$\frac{d}{dt}\mathcal{E}^1_{\lambda,z}(t) + \mu_1(t)\mathcal{E}^1_{\lambda,z}(t) \leq \frac{c}{2t^{p-1}}\left(M + \|z\|^2\right). \tag{4.179}$$

As before, we define the function

$$\mathfrak{M}_1 : [t_1, +\infty) \to \mathbb{R}, \quad t \mapsto \mathfrak{M}_1(t) := \exp\left(\int_{t_1}^{t} \mu_1(s) ds\right) = \exp\left(\int_{t_1}^{t} \frac{\lambda - 2}{s} ds\right) \\ = C_{\mathfrak{M}_1} t^{\lambda-2}, \tag{4.180}$$

with $C_{\mathfrak{M}_1} = t_1^{2-\lambda}$. The function $\mathfrak{M}_1(\cdot)$ is constructed such that $\frac{d}{dt}\mathfrak{M}_1(t) = \mathfrak{M}_1(t)\mu_1(t)$, hence

$$\frac{d}{dt}\left(\mathfrak{M}_1(t)\mathcal{E}^1_{\lambda,z}(t)\right) = \mathfrak{M}_1(t)\left(\frac{d}{dt}\mathcal{E}^1_{\lambda,z}(t) + \mu_1(t)\mathcal{E}^1_{\lambda,z}(t)\right) \quad \text{for almost all} \quad t \geq t_1. \tag{4.181}$$

The relations (4.181) and (4.179) give for almost all $t \geq t_1$

$$\frac{d}{dt}\left(\mathfrak{M}_1(t)\mathcal{E}^1_{\lambda,z}(t)\right) \leq \frac{c}{2}\left(M + \|z\|^2\right)\mathfrak{M}_1(t)t^{1-p}. \tag{4.182}$$

We integrate (4.182) from t_1 to $t \geq t_1$ to get

$$\mathfrak{M}_1(t)\mathcal{E}^1_{\lambda,z}(t) - \mathfrak{M}_1(t_1)\mathcal{E}^1_{\lambda,z}(t_1) \leq \frac{c}{2}\left(M + \|z\|^2\right)\int_{t_1}^{t} \mathfrak{M}_1(s)s^{1-p}ds.$$

Thus, for all $t \geq t_1$ it holds

$$\mathcal{E}^1_{\lambda,z}(t) \leq \frac{\mathfrak{M}_1(t_1)\mathcal{E}^1_{\lambda,z}(t_1)}{\mathfrak{M}_1(t)} + \frac{c}{2}\left(M + \|z\|^2\right)\frac{C_{\mathfrak{M}_1}}{\mathfrak{M}_1(t)}\int_{t_1}^{t} s^{\lambda-(p+1)}ds. \tag{4.183}$$

The inequality above holds for all $z \in \mathcal{H}$ and all $t \geq t_1$. For all $t \geq t_1$, we choose

$$z := z(t) \in \operatorname*{arg\,min}_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_{t,i}(z) - f_{t,i}(x(t)),$$

which, since $\mathcal{E}^1_{\lambda}(t) = \mathcal{E}^1_{\lambda,z(t)}(t)$, yields

$$\mathcal{E}^1_{\lambda}(t) \leq \frac{\mathfrak{M}_1(t_1)\mathcal{E}^1_{\lambda,z(t)}(t_1)}{C_{\mathfrak{M}_1}t^{\lambda-2}} + \frac{c}{2t^{\lambda-2}}\left(M + \|z(t)\|^2\right)\left[\frac{t^{\lambda-p}}{\lambda-p} - \frac{t_1^{\lambda-p}}{\lambda-p}\right].$$

By Proposition 4.6.10, $z(\cdot)$ is bounded, which means that there exist constants $C_1, C_2 > 0$ such that for all $t \geq t_1$

$$\mathcal{E}^1_{\lambda}(t) \leq C_1 + C_2 t^{2-p}, \tag{4.184}$$

and hence

$$\mathcal{E}^1_{\lambda}(t) = \mathcal{O}\left(t^{2-p}\right) \quad \text{as} \quad t \to +\infty, \tag{4.185}$$

proving statement *i)*. From here, the remaining four statements of the theorem follow as in the proof of Theorem 4.6.28. □

Remark 4.6.31. *If we choose $\lambda = 2$ in the proof of Theorem 4.6.30, we do not need to assume the boundedness of $x(\cdot)$ to conclude* (4.179) *from* (4.178)*. This implies that in the case $q = 1$ and $\alpha \geq 3$ the bound $\|x(t) - z(t)\| = \mathcal{O}(1)$ as $t \to +\infty$ follows without the boundedness assumption on $x(\cdot)$.*

The case $p \in (0, 2]$ and $q + 1 < p$: weak convergence of the trajectories

In this part, we show that in the case $p \in (0, 2]$ and $q + 1 < p$ the bounded trajectory solutions of (MTRIGS) converge weakly to a weakly Pareto optimal point of (MOP). To this end, we make use of Opial's Lemma and the energy function from Definition 4.6.21 with $\gamma(\cdot)$ and $\xi(\cdot)$ to be specified later. The convergence rates derived in Subsection 4.6.5 are valid in this setting.

Theorem 4.6.32. *Let $p \in (0, 2)$, $q + 1 < p$, and $x(\cdot)$ be a trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, for $r \in \left[q, \frac{q+1}{2}\right]$, we have*

$$\int_{t_0}^{+\infty} s^{2r-q}\|\dot{x}(s)\|^2 ds < +\infty.$$

Proof. Let $z \in \mathcal{H}$ fixed. Define

$$\gamma : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \gamma(t) = 2rt^{r-1}.$$

With this choice, inequality (4.146) reads for almost all $t \geq t_0$

$$\begin{aligned}
&\frac{d}{dt}\mathcal{G}^r_{\gamma,\xi,z}(t) \leq \\
&\frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + t^r(2rt^{r-1} + rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2 \\
&+ \left(2rt^{r-1}(2rt^{r-1} + rt^{r-1} - \alpha t^{r-q}) + 2r(r-1)t^{2r-2} + \xi(t)\right) \langle x(t) - z, \dot{x}(t)\rangle \\
&+ \left(4r^2(r-1)t^{2r-3} + \frac{\xi'(t)}{2} - \beta r t^{2r-1-p}\right) \|x(t) - z\|^2 \\
= &\frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + \left(2rt^{r-1}(3rt^{r-1} - \alpha t^{r-q}) + 2r(r-1)t^{2r-2} + \xi(t)\right) \langle x(t) - z, \dot{x}(t)\rangle \\
&+ \left(4r^2(r-1)t^{2r-3} + \frac{\xi'(t)}{2} - \beta r t^{2r-1-p}\right) \|x(t) - z\|^2 + t^r(3rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2.
\end{aligned} \tag{4.186}$$

Now we choose

$$\begin{aligned}
\xi : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto \xi(t) &:= 2rt^{r-1}(\alpha t^{r-q} - 3rt^{r-1}) + 2r(1-r)t^{2(r-1)} \\
&= 2\alpha r t^{2r-q-1} + 2r(1-4r)t^{2(r-1)},
\end{aligned}$$

and notice that $\xi'(t) = 2\alpha r(2r - q - 1)t^{2r-q-2} + 4r(r-1)(1-4r)t^{2r-3}$ for all $t \geq t_0$. With this choice, inequality (4.186) simplifies for almost all $t \geq t_0$ to

$$\begin{aligned}
\frac{d}{dt}\mathcal{G}^r_{\gamma,\xi,z}(t) \leq &\left(2r(r-1)(1-2r)t^{2r-3} + \alpha r(2r-q-1)t^{2r-q-2} - \beta r t^{2r-1-p}\right) \|x(t) - z\|^2 \\
&+ \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2 + t^r(3rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2.
\end{aligned} \tag{4.187}$$

Since $r \leq \frac{q+1}{2}$, we conclude from (4.187) that for almost all $t \geq \max\left(\left(\frac{\max(2(r-1)(1-2r),0)}{\beta}\right)^{\frac{1}{2-p}}, t_0\right)$

$$\frac{d}{dt}\mathcal{G}^r_{\gamma,\xi,z}(t) \leq t^r(3rt^{r-1} - \alpha t^{r-q})\|\dot{x}(t)\|^2 + \frac{p\beta t^{2r}}{2t^{p+1}}\|z\|^2. \tag{4.188}$$

Hence, there exist $t_1 \geq \max\left(\left(\frac{\max(2(r-1)(1-2r),0)}{\beta}\right)^{\frac{1}{2-p}}, t_0\right)$ and $a, b > 0$ such that for almost all $t \geq t_1$

$$\frac{d}{dt}\mathcal{G}^r_{\gamma,\xi,z}(t) \leq -\, at^{2r-q}\|\dot{x}(t)\|^2 + bt^{2r-p-1}\|z\|^2,$$

therefore for all $t \geq t_1$

$$\mathcal{G}^r_{\gamma,\xi,z}(t) - \mathcal{G}^r_{\gamma,\xi,z}(t_1) \leq -\, a\int_{t_1}^{t} s^{2r-q}\|\dot{x}(s)\|^2 ds + b\|z\|^2 \int_{t_1}^{t} s^{2r-p-1} ds.$$

Since this holds for all $z \in \mathcal{H}$, we conclude for all $t \geq t_1$

$$\mathcal{G}^r_{\lambda,\xi}(t) - \mathcal{G}^r_{\lambda,\xi,z(t)}(t_1) \leq -\, a\int_{t_1}^{t} s^{2r-q}\|\dot{x}(s)\|^2 ds + b\|z(t)\|^2 \int_{t_1}^{t} s^{2r-p-1} ds.$$

For $t \geq \left(\frac{\max(1-4r,0)}{\alpha}\right)^{\frac{1}{1-q}}$, it holds that $\xi(t) \geq 0$ and hence $\mathcal{G}^r_{\lambda,\xi}(t) \geq 0$. Then, for all $t \geq \max\left(\frac{\max(1-4r,0)}{\alpha}, t_1\right)$

$$a\int_{t_1}^{t} s^{2r-q}\|\dot{x}(s)\|^2 ds \leq \mathcal{G}^r_{\lambda,\xi,z(t)}(t_1) + b\|z(t)\|^2 \int_{t_1}^{t} s^{2r-p-1} ds.$$

Since $z(\cdot)$ is bounded by Proposition 4.6.10 and $2r - p - 1 < -1$, the right hand side of the previous inequality is uniformly bounded for all $t \geq \max\left(\left(\frac{1-4r}{\alpha}\right)^{\frac{1}{1-q}}, t_1\right)$, hence

$$\int_{t_0}^{+\infty} s^{2r-q}\|\dot{x}(s)\|^2 ds < +\infty.$$

□

Next, we discuss the boundary case $p = 2$. To derive weak convergence, we need an additional condition on the parameter $\beta > 0$.

Theorem 4.6.33. *Let $p = 2$, $q \in (0,1)$, $\beta \geq q(1-q)$, and $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, for $r \in \left[q, \frac{1+q}{2}\right]$, we have*

$$\int_{t_0}^{+\infty} s^{2r-q}\,\|\dot{x}(s)\|^2 ds < +\infty. \tag{4.189}$$

Proof. The proof follows analogously to the proof of Theorem 4.6.32, with the difference that in order to conclude (4.188) from (4.187) the additional inequality

$$2(r-1)(1-2r) \leq \beta, \tag{4.190}$$

is necessary. Since $r := \frac{q+1}{2}$ satisfies (4.190), it holds

$$\int_{t_0}^{+\infty} s\,\|\dot{x}(s)\|^2 ds < +\infty,$$

which implies that (4.189) holds for all $r \in \left[q, \frac{q+1}{2}\right]$. □

Remark 4.6.34. *In both regimes, namely, for $p \in (0,2)$ with $q+1 < p$, and for $p = 2$, $q \in (0,1)$ with $\beta \geq q(1-q)$, choosing $r := \frac{1+q}{2}$ we obtain the following integral estimate, which describes the convergence behavior of the velocity of the trajectory*

$$\int_{t_0}^{+\infty} s \, \|\dot{x}(s)\|^2 ds < +\infty.$$

We use the integral estimates given in Theorem 4.6.32 and in Theorem 4.6.33 to prove the weak convergence of the trajectory solution using Opial's Lemma (see Lemma 2.1.6). The following two results prove that the first condition in Opial's Lemma is satisfied, while the final weak convergence statement is shown in Theorem 4.6.37.

Lemma 4.6.35. *Let $p \in (0,2]$ and let $q \in (0,1)$, or $q = 1$ with $\alpha \geq 3$, and let $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Let $\mathcal{W}_i(\cdot), i = 1, ..., m$, be the energy function defined in Proposition 4.6.17. Then, for all $i = 1, \dots, m$, the limit*

$$f_i^\infty := \lim_{t \to +\infty} f_i(x(t)) = \lim_{t \to +\infty} \mathcal{W}_i(t) = \inf_{t \geq t_0} \mathcal{W}_i(t) \in \mathbb{R}$$

exists.

Proof. Let $i \in \{1, \dots, m\}$ be fixed. Since $x(\cdot)$ is bounded, $\inf_{t \geq t_0} f_i(x(t)) \in \mathbb{R}$ holds, therefore

$$\inf_{t \geq t_0} \mathcal{W}_i(t) = \inf_{t \geq t_0} \left(f_i(x(t)) + \frac{\beta}{2t^p} \|x(t)\|^2 + \frac{1}{2} \|\dot{x}(t)\|^2 \right) \geq \inf_{t \geq t_0} f_i(x(t)) \in \mathbb{R}. \tag{4.191}$$

By Proposition 4.6.17, $\mathcal{W}_i(\cdot)$ is monotonically decreasing, thus

$$\lim_{t \to +\infty} \mathcal{W}_i(t) = \inf_{t \geq t_0} \mathcal{W}_i(t) > -\infty. \tag{4.192}$$

By Theorem 4.6.27, Theorem 4.6.28 and Theorem 4.6.30, it holds $\dot{x}(t) \to 0$ as $t \to +\infty$. Hence, $\frac{\beta}{2t^p} \|x(t)\|^2 + \frac{1}{2} \|\dot{x}(t)\|^2 \to 0$ as $t \to +\infty$. Thus

$$\lim_{t \to +\infty} f_i(x(t)) = \lim_{t \to +\infty} \mathcal{W}_i(t), \tag{4.193}$$

which leads to the desired result. □

Lemma 4.6.36. *Let $p \in (0,2)$, $q \in (0,1)$ with $q+1 < p$, or $p = 2$, $q \in (0,1)$ and $\beta \geq q(1-q)$, $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13, and assume that*

$$S := \{ z \in \mathcal{H} \, : \, f_i(z) \leq f_i^\infty \text{ for } i = 1, \dots, m \} \neq \emptyset,$$

with $f_i^\infty = \lim_{t \to \infty} f_i(x(t)) \in \mathbb{R}$. Then, for all $z \in S$, the limit $\lim_{t \to +\infty} \|x(t) - z\|$ exists.

Proof. Let $z \in S$, and define the function

$$h_z : [t_0, +\infty) \to \mathbb{R}, \quad t \mapsto h_z(t) := \frac{1}{2} \|x(t) - z\|^2.$$

For almost all $t \geq t_0$ it holds that

$$\dot{h}_z(t) = \langle x(t) - z, \dot{x}(t) \rangle \quad \text{and} \quad \ddot{h}_z(t) = \langle x(t) - z, \ddot{x}(t) \rangle + \|\dot{x}(t)\|^2. \tag{4.194}$$

From (4.194) and (4.141), we have for almost all $t \geq t_0$

$$\begin{aligned}\ddot{h}_z(t) + \frac{\alpha}{t^q}\dot{h}_z(t) &= \left\langle \ddot{x}(t) + \frac{\alpha}{t^q}\dot{x}(t), x(t) - z \right\rangle + \|\dot{x}(t)\|^2, \\ &= \left\langle -\sum_{i=1}^m \theta_i(t)\nabla f_i(x(t)) - \frac{\beta}{t^p}x(t), x(t) - z \right\rangle + \|\dot{x}(t)\|^2,\end{aligned} \tag{4.195}$$

where $\theta(\cdot)$ be the measurable weight function given by Proposition 4.6.18. Since $z \in S$, we have for all $i = 1, \ldots, m$, and almost all $t \geq t_0$

$$\begin{aligned}f_i(x(t)) + \frac{\beta}{2t^p}\|x(t)\|^2 + \frac{1}{2}\|\dot{x}(t)\|^2 \geq\ & f_i(z) = f_i(z) + \frac{\beta}{2t^p}\|z\|^2 - \frac{\beta}{2t^p}\|z\|^2 \\ \geq f_i(x(t)) + \frac{\beta}{2t^p}\|x(t)\|^2 + &\left\langle \nabla f_i(x(t)) + \frac{\beta}{t^p}x(t), z - x(t) \right\rangle - \frac{\beta}{2t^p}\|z\|^2,\end{aligned}$$

hence

$$\left\langle \nabla f_i(x(t)) + \frac{\beta}{t^p}x(t), z - x(t) \right\rangle \leq \frac{\beta}{2t^p}\|z\|^2 + \frac{1}{2}\|\dot{x}(t)\|^2. \tag{4.196}$$

We define the function

$$k : [t_0, +\infty) \to [0, +\infty), \quad t \mapsto k(t) \coloneqq \frac{\beta}{2t^p}\|z\|^2 + \frac{3}{2}\|\dot{x}(t)\|^2.$$

By Theorem 4.6.32 and Theorem 4.6.33, we have $\left(t \mapsto t^q\|\dot{x}(t)\|^2\right) \in L^1([t_0, +\infty), \mathbb{R})$. On the other hand, since $q + 1 < p$, we get $\left(t \mapsto \frac{\beta t^q}{2t^p}\|z\|^2\right) \in L^1([t_0, +\infty), \mathbb{R})$, consequently, $(t \mapsto t^q k(t)) \in L^1([t_0, +\infty), \mathbb{R})$. Combining (4.195) and (4.196) gives for almost all $t \geq t_0$

$$\ddot{h}_z(t) + \frac{\alpha}{t^q}\dot{h}_z(t) \leq k(t).$$

Now, we can use Lemma 2.2.16 to conclude that the limit

$$\lim_{t\to+\infty} \|x(t) - z\| \text{ exists.}$$

□

Theorem 4.6.37. *Let $p \in (0, 2)$ with $q + 1 < p$, or $p = 2$, $q \in (0, 1)$ with $\beta \geq q(1 - q)$, and let $x(\cdot)$ be a bounded trajectory solution of* (MTRIGS) *in the sense of Definition 4.6.13. Then, $x(t)$ converges weakly to a weakly Pareto optimal point of* (MOP) *as $t \to +\infty$, which belongs to $\mathcal{L}(F, F^\infty)$, where $F^\infty = \lim_{t\to+\infty} F(x(t))$ for $i = 1, \ldots, m$.*

Proof. We define the set $S \coloneqq \{z \in \mathcal{H} : f_i(z) \leq f_i^\infty \text{ for } i = 1, \ldots, m\}$ as in Lemma 4.6.36. Since $x(\cdot)$ is bounded, it possesses a weak sequential cluster point $x^\infty \in \mathcal{H}$. This means that there exists a sequence $(t_k)_{k\geq 0}$ which converges to $+\infty$ with the property that $x(t_k)$ converges weakly to x^∞ as $k \to +\infty$. The functions f_i being weakly lower semicontinuous fulfill for all $i = 1, \ldots, m$

$$f_i(x^\infty) \leq \liminf_{k\to+\infty} f_i(x(t_k)) = \lim_{k\to+\infty} f_i(x(t_k)) = f_i^\infty,$$

therefore $x^\infty \in S$. We conclude that S is nonempty and all weak sequential cluster points of $x(\cdot)$ belong to S. On the other hand, according to Lemma 4.6.36 we have that $\lim_{t\to+\infty}\|x(t) - z\|$ exists for all $z \in S$. We can use Opial's Lemma (Lemma 2.1.6) to conclude that $x(t)$ converges weakly to an element in S as $t \to +\infty$. By Theorem 4.6.27, $\varphi(x(t)) \to 0$ as $t \to +\infty$, therefore, since $\varphi(\cdot)$ is weakly lower semicontinuous (see Theorem 2.3.14), $\varphi(x^\infty) \leq \liminf_{k\to+\infty} \varphi(x(t_k)) = 0$. By Theorem 2.3.13, x^∞ is a weakly Pareto optimal point of (MOP). □

4.6.6 Numerical experiments

In this section, we illustrate the typical behavior of the trajectory solution $x(\cdot)$ of (MTRIGS) using two example problems. In the first example, we show that trajectory solutions $x(\cdot)$ of (MTRIGS) converge to a weakly Pareto optimal point x^*, which is the element of minimum norm in $\bigcap_{i=1}^m \mathcal{L}(f_i, f_i(x^*))$, whereas those of (MAVD) may fail to exhibit this behavior. In the second example, we analyze the sensitivity of trajectory solutions of (MTRIGS) with respect to $q \in (0, 1]$ and $p \in (0, 2]$. We highlight how different parameter choices affect the decay of the merit function values $\varphi(x(t))$ and the asymptotic behavior of the distance $\|x(t) - z(t)\|$ to the generalized regularization path as $t \to +\infty$.

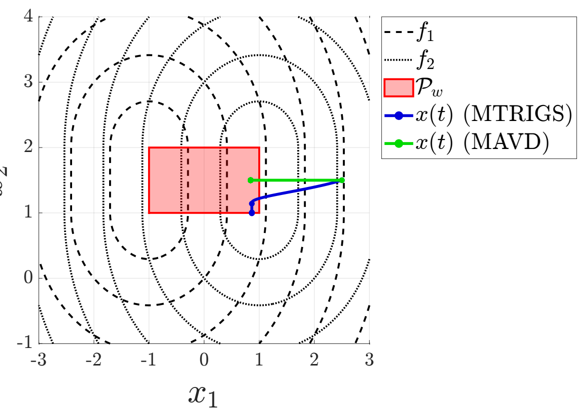


Figure 4.10: Contour plots of f_1 and f_2 defined in (4.197), the weak Pareto set $\mathcal{P}_w$ of the problem (MOP-Ex$_1$) and the trajectory solutions $x(\cdot)$ of (MTRIGS) and (MAVD) with identical initial conditions, respectively.

Comparison of (MTRIGS) **with** (MAVD)

In the first example, we consider the following instance of (MOP). Define the sets

$$S_1 := \{-1\} \times [1, 2] \subseteq \mathbb{R}^2 \quad \text{and} \quad S_2 := \{1\} \times [1, 2] \subseteq \mathbb{R}^2,$$

and the functions

$$f_i : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto f_i(x) := \frac{1}{2}\operatorname{dist}(x, S_i)^2, \quad \text{for} \quad i = 1, 2, \tag{4.197}$$

which are both convex and continuously differentiable, and have Lipschitz continuous gradients. The weak Pareto set of the multiobjective optimization problem

$$\min_{x \in \mathbb{R}^2} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \tag{MOP-Ex$_1$}$$

is given by

$$\mathcal{P}_w = \operatorname{conv}(S_1 \cup S_2) = [-1, 1] \times [1, 2].$$

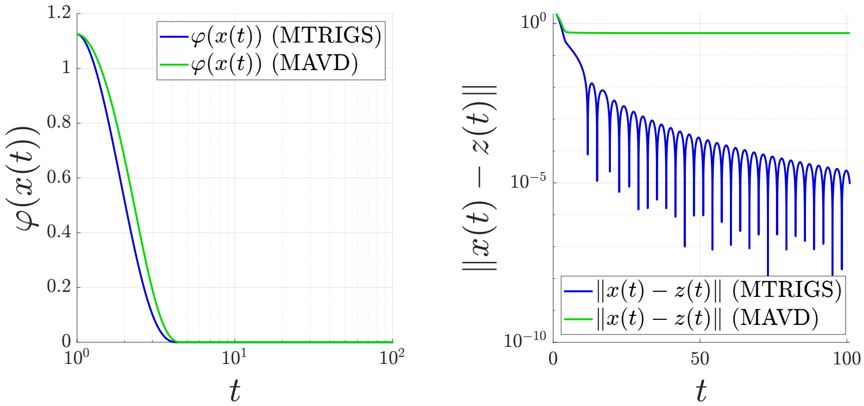
Let $z = (z_1, z_2)^\top \in \mathcal{P}_w$. Then, the element of minimum norm in $\bigcap_{i=1}^2 \mathcal{L}(f_i, f_i(z))$ is given by

$$\operatorname*{proj}_{\bigcap_{i=1}^2 \mathcal{L}(f_i, f_i(z))}(0) = (z_1, 1). \tag{4.198}$$

We approximate a trajectory solution for (MTRIGS) and (MAVD), respectively, in the following context:

- For (MTRIGS), we set $\alpha := 4$, $\beta := \frac{1}{2}$, $q := \frac{7}{8}$ and $p := \frac{7}{4}$;
- For (MAVD), we set $\alpha := 4$;
- For both systems, we use as initial conditions $x(t_0) = (2.5,\, 0.5)$ and $\dot{x}(t_0) = (0, 0)$, where $t_0 = 1$;
- For both systems, we use an equidistant discretization in time, i.e., time steps $t_k := t_0 + kh$ with step size $h = 1\text{e}{-2}$;
- For both systems, we approximate the first and second derivatives by $\dot{x}(t_k) = \frac{x(t_{k+1}) - x(t_k)}{h}$ and $\ddot{x}(t_k) = \frac{x(t_{k+1}) - 2x(t_k) + x(t_{k-1})}{h^2}$, respectively;
- For both systems, we consider the trajectory solutions for $t \in [1, 100]$.

Note that for (MTRIGS) it holds that $p < q + 1$. According to Theorem 4.6.28 and Theorem 4.6.29, we have convergence of the merit function values $\varphi(x(t)) \to 0$, convergence of the distance of the trajectory to the regularization path $\|x(t) - z(t)\| \to 0$ and strong convergence of the trajectory $x(t)$ to a weakly Pareto optimal point as $t \to +\infty$.
Figure 4.10 shows the contour plots of the objective function f_1 and f_2 defined in (4.197), along with the weak Pareto set $\mathcal{P}_w$ highlighted in red in the decision space. The figure also displays the trajectory solutions of (MTRIGS) and (MAVD) with identical initial conditions, respectively, which both converge to points in the weak Pareto set. Notably, the solution of (MAVD) evolves solely in the x_1-direction, whereas the Tikhonov regularization ensures that the solution of (MTRIGS) converges to an element as specified by (4.198).
Figure 4.11 visualizes the behavior of the trajectory solutions of (MTRIGS) and (MAVD) by showing, in two subfigures, the evolution of the merit function values and the distance of the trajectories to the generalized regularization paths. As already shown in Figure 4.10, the trajectories enter the weak Pareto set $\mathcal{P}_w$ after some time, implying that the merit function values $\varphi(x(t))$ vanish accordingly. This is illustrated in Subfigure 4.11a. Subfigure 4.11b depicts the distance between the trajectory and the generalized regularization path, i.e., $\|x(t) - z(t)\|$ for $t \in [1, 100]$. For the solution of (MAVD), this distance converges to a positive limit as $t \to +\infty$. In contrast, for the solution of (MTRIGS), the distance decays to zero at a sublinear rate, as predicted by Theorem 4.6.28.



(a) (b)

Figure 4.11: The merit function values $\varphi(x(t))$ and the distance $\|x(t) - z(t)\|$ of the trajectory solutions to the generalized regularization path for (MTRIGS) and (MAVD) for the problem (MOP-Ex$_1$).

The convergence behaviour of (MTRIGS) **for different values of** $q \in (0, 1]$ **and** $p \in (0, 2]$

The numerical experiments in this subsection demonstrate a similar influence of the parameters q and p in on the asymptotic behaviour of (MTRIGS) as was observed in [146] for the system (TRIGS) in the context of single objective optimization. Consider

$$f_1 : \mathbb{R}^4 \to \mathbb{R}, \quad x \mapsto f_1(x) := \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}(x_2 - 1)^2, \quad \text{and}$$
$$f_2 : \mathbb{R}^4 \to \mathbb{R}, \quad x \mapsto f_1(x) := \frac{1}{2}(x_1 + 1)^2 + \frac{1}{2}(x_2 - 1)^2,$$

which are both convex and continuously differentiable functions, and have Lipschitz continuous gradients. The weak Pareto set of the multiobjective optimization problem

$$\min_{x \in \mathbb{R}^4} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \qquad \text{(MOP-Ex}_2\text{)}$$

is given by

$$\mathcal{P}_w := [-1, 1] \times \{1\} \times \mathbb{R} \times \mathbb{R} \subseteq \mathbb{R}^4.$$

We approximate a trajectory solution for (MTRIGS) in the following context:

- We set $\alpha := 4$, $\beta := \frac{1}{2}$, and consider different values for $q \in (0, 1]$ and $p \in (0, 2]$;
- We use as initial conditions $x(t_0) = x_0$ and $\dot{x}(t_0) = 0$ with $t_0 = 1$ and $x_0 = (2, 3, 4, 5)^\top$;
- We use an equidistant discretization in time, i.e., time steps $t_k := t_0 + kh$ with step size $h = 1\text{e}{-3}$;

- We approximate the first and second derivative of $x(\cdot)$ in time by $\dot{x}(t_k) = \frac{x(t_{k+1})-x(t_k)}{h}$ and $\ddot{x}(t_k) = \frac{x(t_{k+1})-2x(t_k)+x(t_{k-1})}{h^2}$ respectively;
- We consider the trajectory solutions for $t \in [1, 100]$.

We first fix $q = 0.8$ and vary the parameter p over the set $\{0.25,\ 0.75,\ 1.25,\ 1.75\}$. Afterwards, we fix $p = 1.1$ and vary q over the set $\{0.3,\ 0.6,\ 0.8,\ 0.99\}$.

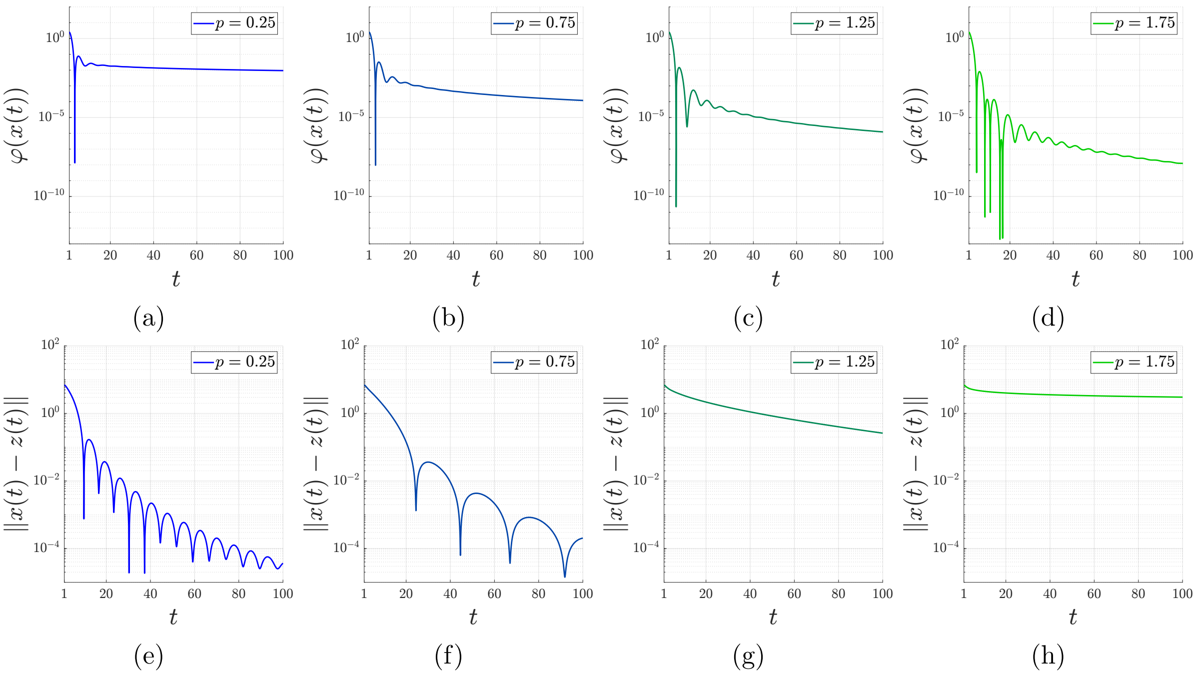


Figure 4.12: The merit function values $\varphi(x(t))$ and the distance $\|x(t) - z(t)\|$ of the trajectory to the generalized regularization path for $q = 0.8$ and $p \in \{0.25,\ 0.75,\ 1.25,\ 1.75\}$.

Figure 4.12 shows the evolution of the merit function values $\varphi(x(t))$ and of the distance $\|x(t) - z(t)\|$ of the trajectory to the generalized regularization path for $q = 0.8$ and $p \in \{0.25,\ 0.75,\ 1.25,\ 1.75\}$. The merit function values exhibit the fastest decay for the largest value of $p = 1.75$. This behavior is expected, as higher values of p cause the Tikhonov regularization parameter to decay more rapidly, thus exerting less influence and allowing the function values to converge more quickly. Conversely, the distance $\|x(t) - z(t)\|$ decays most rapidly for smaller values of p, where the regularization parameter vanishes more slowly and effectively guides the trajectory towards the regularization path.
Figure 4.13 shows the evolution of the merit function values $\varphi(x(t))$ and the distance $\|x(t)-z(t)\|$ of the trajectory to the generalized regularization path for $p = 1.1$ and $q \in \{0.3,\ 0.6,\ 0.8,\ 0.99\}$. The decay of the merit function values $\varphi(x(t))$ is generally insensitive to the choice of q; for all considered values of q, the convergence rate remains essentially the same. However, for larger values of q, the merit function exhibits more pronounced oscillations. This behavior is expected, as a larger value of q implies a faster decay of the friction term $\frac{\alpha}{t^q}$, thereby reducing damping.

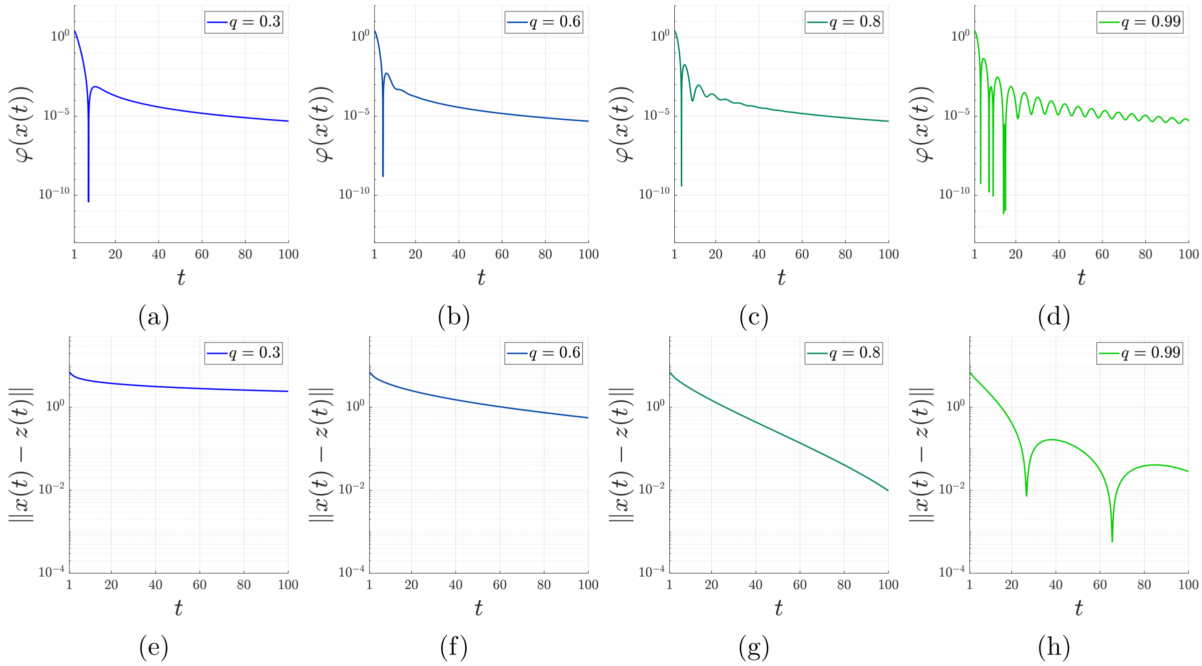


Figure 4.13: The merit function values $\varphi(x(t))$ and the distance $\|x(t) - z(t)\|$ of the trajectory to the generalized regularization path for $p = 1.1$ and $q \in \{0.3, 0.6, 0.8, 0.99\}$.

In contrast, the decay of the distance $\|x(t) - z(t)\|$ is strongly influenced by q, particularly for $q = 0.99$, where convergence is significantly faster. For the smallest value $q = 0.3$, the distance decreases only slowly, at a sublinear rate. These observations align with expectations: higher values of q correspond to weaker friction, which allows the trajectory to approach the regularization path more rapidly in the early phase.

Chapter 5

An accelerated gradient method for convex multiobjective optimization

In optimization, first-order methods – i.e., methods that use only objective function and gradient information – are very popular, as they are straightforward to implement and can be applied to a variety of problems. Additionally, they are backed by a mature theory and admit convergence guarantees in many settings. On the downside, they can suffer from slow convergence, especially if the considered optimization problems are ill-conditioned. A general idea to overcome this problem is to accelerate an iterative method by incorporating inertia or momentum, using information from past iterates in the update scheme [196]. While accelerated first-order methods for smooth optimization [180, 182], nonsmooth optimization [118], problems with separable structure [24, 33, 63], min-max problems [46, 64, 121] and related problems like monotone inclusions [4, 5, 60, 238] and variational inequalities [45, 209], are growing in popularity because of application in areas like image processing [33, 64], signal processing [179, 194] or machine learning and statistics [150], these methods are not considered extensively in multiobjective optimization. Inspired by the active research and successful application in the listed areas, we want to develop fast first-order methods for multiobjective optimization. Building on the observations in Chapter 4 on gradient dynamical systems related to multiobjective optimization problems, we develop an accelerated first-order method to solve the problem

$$\text{(MOP)} \qquad \min_{x\in\mathcal{H}} F(x) \coloneqq \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

where $\mathcal{H}$ is a real Hilbert space and the objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and sufficiently smooth for all $i = 1, \dots, m$. The accelerated gradient method we propose in this chapter is an improvement of the multiobjective steepest descent method (MGD) we present in Subsection 2.3.4 and uses the idea of Nesterov acceleration [182]. The method can be seen as a discretization of the multiobjective gradient system with asymptotic vanishing damping (MAVD), which we discuss in Section 4.5. Similar to the improvement of the system (MAVD) over the system (MSD) in terms of convergence rates we achieve an improvement in the convergence behavior for the presented method.

We use the abbreviation *multiobjective Nesterov accelerated gradient method* for the following scheme given $\alpha > 0$, step size $h > 0$ and initial iterates $x^0 = x^{-1} \in \mathcal{H}$ to define the sequences $(x^k)_{k\geq 0}$, $(y^k)_{k\geq 0} \subset \mathcal{H}$ and $(\theta^k)_{k\geq 0} \subset \Delta^m$ by

$$\text{(MNAG)} \qquad \left.\begin{aligned} y^k &= x^k + \frac{k-1}{k+\alpha-1}\left(x^k - x^{k-1}\right), \\ \theta^k &\in \arg\min_{\theta\in\Delta^m} \frac{1}{2}\left\| h\sum_{i=1}^m \theta_i \nabla f_i(y^k) + x^k - y^k \right\|^2, \\ x^{k+1} &= y^k - h\sum_{i=1}^m \theta_i^k \nabla f_i(y^k), \end{aligned}\right\} \quad \text{for} \quad k \geq 0.$$

Analogously to the observations on the multiobjective steepest descent method (MGD) and (MGD') we can write the scheme (MNAG) more concisely as

$$\text{(MNAG')} \qquad \left.\begin{aligned} y^k &= x^k + \tfrac{k-1}{k+\alpha-1}\left(x^k - x^{k-1}\right), \\ x^{k+1} &= y^k - h\,\mathrm{proj}_{C(y^k)}\left(\tfrac{1}{h}(y^k - x^k)\right), \end{aligned}\right\} \quad \text{for} \quad k \geq 0,$$

where $C(x) := \operatorname{conv}\left(\{\nabla f_i(x) : i = 1, \ldots, m\}\right)$ denotes the convex hull of the gradients in $x \in \mathcal{H}$. This chapter is outlined as follows. In Section 5.1, we present a derivation of the multiobjective Nesterov accelerated gradient method, as a discretization of (MAVD). We review the case of scalar optimization to motivate this derivation. Additionally, we discuss related methods in multiobjective optimization. The introduction of the method is followed by the asymptotic analysis in Section 5.2. After deriving some preliminary results on the sequence $(x^k)_{k\geq 0}$ defined by (MNAG), we prove an asymptotic result on the convergence of function values measured with the merit function $\varphi(\cdot)$ which is defined in (2.23). For $\alpha \geq 3$ the function values of the iterates converge at a rate of order $\varphi(x^k) = \mathcal{O}(k^{-2})$ as $k \to +\infty$. Additionally, for $\alpha > 3$, we prove weak convergence of the iterates to a weakly Pareto optimal point. We close this section with some numerical experiments laid out in Section 5.3 to verify the obtained convergence rates.

The content of this chapter is based on the following publication:

[217] Sonntag, K. and Peitz, S. *Fast Multiobjective Gradient Methods with Nesterov Acceleration via Inertial Gradient-Like Systems.* In: *Journal of Optimization Theory and Applications* 201 (2024), pp. 539–582. doi: 10.1007/s10957-024-02389-3.

In [217] only the case $\alpha = 3$ is included, while this thesis contains a generalization of the results to the case $\alpha \geq 3$. We recover the convergence rate of order $\varphi(x^k) = \mathcal{O}(k^{-2})$. Additionally, we prove weak convergence of the iterates to weakly Pareto optimal points in the Hilbert space setting given $\alpha > 3$.

5.1 Derivation of the accelerated multiobjective gradient method (MNAG)

5.1.1 Nesterov's accelerated gradient method for scalar optimization

In this subsection, we summarize Nesterov's accelerated gradient method for scalar convex optimization which was first published in 1983 in the seminal paper [182] by Nesterov. In [182] the author proposes a first-order method to solve a smooth and convex optimization problem with convergence rate $\mathcal{O}(k^{-2})$. This is an improvement to the complexity rate of $\mathcal{O}(k^{-1})$ which is for example given by the steepest descent method (see [181, Theorem 2.1.14]). The problem of interest reads as

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

where $f : \mathcal{H} \to \mathbb{R}$ is convex and smooth. In the following we recite a version of Nesterov's accelerated gradient method which uses a different acceleration parameter then the original method. The acceleration parameter we use can be found, e.g., in [63]. For $\alpha > 0$, $h > 0$ and $x^0 = x^{-1} \in \mathcal{H}$, define the sequences $(x^k)_{k\geq 0}, (y^k)_{k\geq 0} \subset \mathcal{H}$ by

$$\text{(NAG)} \qquad \left.\begin{array}{rcl} y^k &=& x^k + \frac{k-1}{k+\alpha-1}(x^k - x^{k-1}), \\ x^{k+1} &=& y^k - h\nabla f(y^k), \end{array}\right\} \quad \text{for} \quad k \geq 0.$$

The method (NAG) is straight-forward to implement and does not need more gradient evaluations per iteration, than, e.g., the steepest descent method. We recite a theorem summarizing the asymptotic properties of the sequence $(x^k)_{k\geq 0}$ given by (NAG) from [15, Theorem 2.1].

Theorem 5.1.1. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient ∇f and assume $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$. Let $\alpha \geq 3$ and $0 < h \leq \frac{1}{L}$. Let $(x^k)_{k\geq 0}$ be the sequence given by (NAG). Then, as $k \to +\infty$:*

i) $f(x^k) - \min_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(\frac{1}{k^2}\right)$;

ii) $\|x^{k+1} - x^k\| = \mathcal{O}\left(\frac{1}{k}\right)$.

If $\alpha > 3$, the following improved rates hold as $k \to +\infty$:

i) $f(x^k) - \min_{x\in\mathcal{H}} f(x) = o\left(\frac{1}{k^2}\right)$;

ii) $\|x^{k+1} - x^k\| = o\left(\frac{1}{k}\right)$;

iii) $x^k \rightharpoonup x^\infty \in \arg\min_{x\in\mathcal{H}} f(x)$.

In [218], it is shown that Nesterov's accelerated gradient method is connected to the following gradient system with asymptotically vanishing damping

$$\text{(AVD)} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0,$$

which we described earlier in the introduction of Section 4.5 in relation to the system (MAVD). If the step size $h \to 0$ tends to zero, the iterates $(x^k)_{k\geq 0}$ converge to a solution $x(\cdot)$ of the differential equation (AVD). Further, the scheme (NAG) can be recovered from an implicit discretization of (AVD). Additionally, the trajectory solutions to (AVD) share similar asymptotical features to the iterates $(x^k)_{k\geq 0}$ as we summarize in the following Theorem (see [13, 166]).

Theorem 5.1.2. *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable and assume that* $\arg\min_{x\in\mathcal{H}} f(x) \neq \emptyset$. *Let $\alpha \geq 3$ and let $x(\cdot)$ be a global solution to* (AVD). *Then, as $t \to +\infty$:*

i) $f(x(t)) - \min_{x\in\mathcal{H}} f(x) = \mathcal{O}\left(\frac{1}{t^2}\right)$;

ii) $\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right)$;

If $\alpha > 3$, the following improved rates hold as $t \to +\infty$:

i) $f(x(t)) - \min_{x\in\mathcal{H}} f(x) = o\left(\frac{1}{t^2}\right)$;

ii) $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$;

iii) $x(t) \rightharpoonup x^\infty \in \arg\min_{x\in\mathcal{H}} f(x)$.

In Section 4.5, we obtain analogous results for the system (MAVD). Motivated by this observation, we develop an accelerated gradient method from the multiobjective gradient dynamical system with asymptotic vanishing damping (MAVD) by using a discretization similar to the one used for (AVD) to obtain (NAG) (see [13, 15, 218]).

5.1.2 Discretization of the system (MAVD)

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. The starting point of the derivation is the multiobjective gradient system with asymptotic vanishing damping

$$\text{(MAVD)} \qquad \frac{\alpha}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0,$$

which gets discussed extensively in Section 4.5. Using an Ansatz similar to Section 2 of [218] (see also [15, Subsection 3.2]), we write out the following implicit discretization of the differential equation (MAVD) with step size $\sqrt{h} > 0$.

$$\frac{\alpha}{k\sqrt{h}} \frac{x^{k+1} - x^k}{\sqrt{h}} + \operatorname*{proj}_{C(y^k)+\frac{x^{k+1}-2x^k+x^{k-1}}{h}}(0) = 0, \tag{5.1}$$

with $y^k = x^k + \frac{k-1}{k+\alpha-1}(x^k - x^{k-1})$. By using $C(y^k) = \operatorname{conv}\left(\left\{\nabla f_i(y^k) : i = 1, \dots, m\right\}\right)$, we do not evaluate the gradients ∇f_i at x^k but at an extrapolated point. Before we develop a gradient method from this scheme, we show in an informal manner that we can recover from this scheme the differential equation (MAVD). We want to emphasize that the derivation of this method is not straightforward as the discretization of a dynamical system is by no means unique. The computations laid out in this subsection for the equation (MAVD) are inspired by similar considerations for the case of scalar optimization as mentioned previously when discussing (AVD) and (NAG). We multiply (5.1) by $\sqrt{h}$ and get

$$\frac{\alpha}{k} \frac{x^{k+1} - x^k}{\sqrt{h}} + \operatorname*{proj}_{\sqrt{h}C(y^k)+\frac{x^{k+1}-2x^k+x^{k-1}}{\sqrt{h}}}(0) = 0. \tag{5.2}$$

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

$$\frac{x^{k+1} - x^k}{\sqrt{h}} = \dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{h} + o(\sqrt{h}) \quad \text{and} \quad \frac{x^k - x^{k-1}}{\sqrt{h}} = \dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{h} + o(\sqrt{h}), \tag{5.3}$$

and hence

$$\frac{x^{k+1} - 2x^k + x^{k-1}}{\sqrt{h}} = \ddot{x}(t)\sqrt{h} + o(\sqrt{h}). \tag{5.4}$$

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

$$\operatorname*{proj}_{\sqrt{h}C(y^k)+\frac{x^{k+1}-2x^k+x^{k-1}}{\sqrt{h}}}(0) = \sqrt{h} \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) + o(\sqrt{h}). \tag{5.5}$$

Combining (5.3), (5.4) and (5.5), we get from (5.2)

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

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

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

Therefore, we have shown that the differential equation (MAVD) can be derived from the scheme (5.1). Next, we derive a method from this scheme. Using Lemma 2.1.20 on (5.1), we get that x^{k+1} is uniquely defined as

$$\begin{aligned} x^{k+1} &= -\left(\frac{k}{k+\alpha} \operatorname*{proj}_{hC(y^k)-2x^k+x^{k-1}}(-x^k) - \frac{\alpha}{k+\alpha}x^k\right) \\ &= x^k - \frac{k}{k+\alpha} \operatorname*{proj}_{hC(y^k)-x^k-x^{k-1}}(0). \end{aligned}$$

The last term can be written as

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

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

$$\min_{\theta \in \Delta^m} \left\| h\left(\sum_{i=1}^{m} \theta_i \nabla f_i(y^k)\right) - (x^k - x^{k-1})\right\|^2. \tag{5.6}$$

We want to drop the factor $\frac{k}{k+\alpha}$ 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 (NAG) (see [15, Remark 3.1]). In addition, we perform a shift of the index k to transform $\frac{k}{k+\alpha}$ into $\frac{k-1}{k+\alpha-1}$. The final method we obtain can be written as follows. For $x^0 = x^{-1} \in \mathcal{H}$ and $h > 0$ and $\alpha > 0$ define the scheme

$$\left.\begin{aligned} y^k &= x^k + \frac{k-1}{k+\alpha-1}\left(x^k - x^{k-1}\right), \\ x^{k+1} &= y^k - h\sum_{i=1}^{m} \theta_i^k \nabla f_i(y^k), \end{aligned}\right\} \quad \text{for} \quad k \geq 0, \tag{5.7}$$

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

$$\min_{\theta \in \Delta^m} \left\| h\left(\sum_{i=1}^{m} \theta_i \nabla f_i(y^k)\right) - \frac{k-1}{k+\alpha-1}(x^k - x^{k-1}) \right\|^2. \tag{5.8}$$

The fact that we have to transform the quadratic optimization problem from (5.6) into (5.8) is an observation from the proof of Proposition 5.2.1. Using $\frac{k-1}{k+\alpha-1}(x^k - x^{k-1}) = y^k - x^k$ we combine (5.7) and (5.8) to (MNAG) which can equivalently be written as (MNAG').

5.1.3 Relation to other existing methods

A first accelerated gradient method for multiobjective optimization is proposed in [94]. The authors are able to prove improved convergence rate of order $\mathcal{O}(k^{-2})$ but only under the very restrictive assumptions that the weights θ^k remain constant from some point on. In this case the method is equivalent to an accelerated gradient method applied to the weighted sum scalarization (see [91]) of the multiobjective optimization problem and convergence follows. In [222] the authors define an accelerated proximal gradient method for finite dimensional multiobjective optimization problems with objective functions that have a separable structure of the form $f_i(x) = g_i(x) + h_i(x)$, 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 $i = 1, \ldots, m$. This method is an improvement of the multiobjective proximal gradient method [223, 225]. Since we only treat the case of smooth objective functions f_i, we set in the comparison $h_i \equiv 0$ for all $i = 1, \ldots, m$. Then, the accelerated gradient method by Tanabe, Fukuda and Yamashita can be written as follows. For $x^0 = x^{-1} \in \mathbb{R}^n$ and $h > 0$ and $t_0 = 1$ define

$$\text{(TFY)} \quad \left.\begin{aligned} t_{k+1} &= \sqrt{t_k^2 + \tfrac{1}{4}} + \tfrac{1}{2}, \\ y^k &= x^k + \tfrac{t_k - 1}{t_{k+1}}\left(x^k - x^{k-1}\right), \\ x^{k+1} &= \operatorname*{arg\,min}_{z \in \mathbb{R}^n} \max_{i=1,\ldots,m} \langle \nabla f_i(y^k), z - y^k \rangle + f_i(y^k) - f_i(x^k) + \tfrac{1}{2h}\|z - y^k\|^2, \end{aligned}\right\} \text{for } k \geq 0.$$

In comparison to (MNAG) where the parameter $\frac{k-1}{k+\alpha-1}$ with $\alpha \geq 3$ is used in (TFY) the acceleration parameter is $\frac{t_k-1}{t_{k+1}}$ with $t_{k+1} = \sqrt{t_k^2 + \frac{1}{4}} + \frac{1}{2}$ and $t_0 = 1$. The acceleration parameter used in (TFY) is the one used originally in [182] and which is also adapted in [33] for the fast iterative shrinkage-thresholding algorithm while the parameter $\frac{k-1}{k+\alpha-1}$ is obtained from a generalization of $(t_k)_{k \geq 0}$ [63]. Another difference is the way the iterate x^{k+1} gets updated. The auxiliary problem in (TFY) is similar to the primal formulation of the multiobjective steepest descent direction (see (2.31)) while the computation of the descent direction in (MNAG), is closer to the dual formulation of the multiobjective steepest descent direction (see (2.32)). We rewrite (TFY) to point out its similarity to (MNAG). Using the dual formulation of the optimization problem which is solved in every iteration to compute x^{k+1} in (TFY) (similar to Proposition 2.3.18), we obtain

$$\text{(TFY')} \qquad \left.\begin{array}{rcl} t_{k+1} &=& \sqrt{t_k^2 + \frac{1}{4}} + \frac{1}{2}, \\ y^k &=& x^k + \dfrac{t_k - 1}{t_{k+1}}\left(x^k - x^{k-1}\right), \\ \theta^k &\in& \arg\min_{\theta \in \Delta^m} \dfrac{h}{2}\left\|\sum_{i=1}^m \theta_i \nabla f_i(y^k)\right\|^2 + \sum_{i=1}^m \theta_i \left(f_i(x^k) - f_i(y^k)\right), \\ x^{k+1} &=& y^k - h\sum_{i=1}^m \theta_i^k \nabla f_i(y^k), \end{array}\right\} \text{ for } k \geq 0.$$

The connection between (MNAG) and (TFY) becomes more evident through (TFY'). If we compare the respective computations of the weights θ^k for the gradient update we see, that they use similar objective functions. We investigate the quadratic optimization problems that have to be solved in each iteration of the methods. In (TFY'), the step direction is computed solving an optimization problem with linear constraints and the quadratic objective function

$$\Psi : \mathbb{R}^m \to \mathbb{R}, \quad \theta \mapsto \Psi(\theta) := \frac{h}{2}\left\|\sum_{i=1}^m \theta_i \nabla f_i(y^k)\right\|^2 + \sum_{i=1}^m \theta_i \left(f_i(x^k) - f_i(y^k)\right).$$

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

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

Minimizing $\Psi(\cdot)$ is equivalent to minimizing the function

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

We note that $\Phi(\cdot)$ is in fact the objective function of the quadratic optimization problem that is used to compute θ^k in (MNAG). The connection between the methods (MNAG) and (TFY) and the system (MAVD) is even stronger. In [156] it is shown that (MAVD) can be derived as the continuous model to (TFY). Building on these observations in [156] the authors introduce a novel accelerated proximal gradient method for multiobjective optimization. Further generalizations of (TFY) can be found in [221].

5.2 Asymptotic analysis

The analysis carried out in the remainder of this chapter is based on the classical results from [182]. Additionally, the proofs are influenced by the finding in [24], where improved convergence rates for (NAG) are proven, and the results in [221], where a generalization of (TFY) gets discussed.

5.2.1 Assumptions

($\mathcal{A}_1$) The objective functions $f_i : \mathcal{H} \to \mathbb{R}$ are convex and continuously differentiable with L-Lipschitz continuous gradients $\nabla f_i : \mathcal{H} \to \mathcal{H}$ for all $i = 1, \dots, m$.

($\mathcal{A}_2$) For all $x_0 \in \mathcal{H}$ and for all $x \in \mathcal{L}_F(F(x_0)) = \mathcal{L}(F, F(x_0))$ it holds that $\mathcal{LP}_w(F, F(x)) \neq \emptyset$ and further

$$R := \sup_{F^* \in F(\mathcal{LP}_w(F, F(x_0)))} \inf_{z \in F^{-1}(\{F^*\})} \frac{1}{2}\|z - x^0\|^2 < +\infty. \tag{5.9}$$

The assumptions used in this section are identical with the ones introduced in Subsection 4.5.1 for the analysis of the system (MAVD). We want to emphasize that assumption ($\mathcal{A}_2$), while looking technical, is a common assumption used in the literature [216, 217, 221, 222, 223, 224] and is natural in the sense that it reduces in the case of scalar optimization to $\arg\min_{x \in \mathcal{H}} f(x) \neq \emptyset$. In scalar optimization this condition is necessary to obtain fast convergence rates (see Theorem 5.1.1 and Theorem 5.1.2). Additionally, Subsection 4.5.1 includes a discussion of ($\mathcal{A}_2$) by means of example multiobjective optimization problems. In particular, ($\mathcal{A}_2$) is important in the context of Lemma 2.3.15 to bound the merit function value $\varphi(x^k) = \sup_{z \in \mathcal{H}} \min_{i=1,\dots,m} f_i(x^k) - f_i(z)$ using the distance of x^k to the set $\mathcal{LP}_w(F, F(x^0))$.

5.2.2 Preparatory results

We start the investigations of (MNAG) with the following energy estimate.

Proposition 5.2.1. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 0$ and $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k \geq 0}$, $(y^k)_{k \geq 0}$ and $(\theta^k)_{k \geq 0}$ be the sequences given by* (MNAG) *with initial iterates $x^0 = x^{-1} \in \mathcal{H}$. Define for all $i = 1, \dots, m$ the energy sequence $(\mathcal{W}_{i,k})_{k \geq 0}$ by*

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

For all $k \geq 0$, it holds that

$$\mathcal{W}_{i,k+1} - \mathcal{W}_{i,k} \leq -\frac{1}{2h}\frac{\alpha}{k + \alpha - 1}\|x^k - x^{k-1}\|^2.$$

Proof. From the definition of $(x^k)_{k \geq 0}$ and $(y^k)_{k \geq 0}$ in (MNAG), we get for all $k \geq 0$

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

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

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

from which we follow

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

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

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

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

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

which completes the proof.

□

Corollary 5.2.2. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 0$ and $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ be given by* (MNAG)*. Then, for all $k \geq 0$ and all $i = 1, \dots, m$*

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

Definition 5.2.3. *For an arbitrary $z \in \mathcal{H}$ we define the sequence $(\varphi_{z,k})_{k\geq 0}$ which is associated to the multiobjectve optimization problem* (MOP) *and the sequence $(x^k)_{k\geq 0}$ given by* (MNAG)*. For $z \in \mathcal{H}$ and $k \geq 0$ define*

$$\varphi_{z,k} \coloneqq \min_{i=1,\dots,m} f_i(x^k) - f_i(z). \tag{5.10}$$

The sequence $(\varphi_{z,k})_{k\geq 0}$ is naturally linked to the merit function $\varphi(\cdot)$ defined in (2.23) *by the relation*

$$\varphi(x^k) = \sup_{z\in\mathcal{H}} \varphi_{z,k},$$

for all $k \geq 0$.

We use the notation $(\varphi_{z,k})_{k\geq 0}$ to simplify the proofs carried out in the asymptotic analysis. Before we obtain results on the convergence rates of $\varphi(x^k)$ we present intermediate results on the sequence $(\varphi_{z,k})_{k\geq 0}$ for an arbitrary $z \in \mathcal{H}$.

Lemma 5.2.4. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 0$ and $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ and $(y^k)_{k\geq 0}$ be given by* (MNAG)*. Then, for all $z \in \mathcal{H}$ and all $k \geq 0$, it holds that*

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

Proof. The objective functions f_i are convex with L-Lipschitz continuous gradients. Then, by the Descent Lemma (Lemma 2.1.12), for all $i = 1, \dots, m$ it holds that

$$f_i(x^{k+1}) - f_i(z) \leq \langle \nabla f_i(y^k), x^{k+1} - z\rangle + \frac{L}{2}\|x^{k+1} - y^k\|^2. \tag{5.11}$$

By the definition of $\varphi_{z,k+1}$, we follow

$$\varphi_{z,k+1} = \min_{i=1,\dots,m} f_i(x^{k+1}) - f_i(z) \leq \sum_{i=1}^m \theta_i^k \left(f_i(x^{k+1}) - f_i(z)\right). \tag{5.12}$$

Combining (5.11) and (5.12) and using $\sum_{i=1}^m \theta_i^k \nabla f_i(y^k) = \frac{1}{h}(y^k - x^{k+1})$ and $L \leq \frac{1}{h}$, we get

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

which completes the proof. □

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 5.2.5. *Let $\alpha, h > 0$ and let $(x^k)_{k\geq 0}$, $(y^k)_{k\geq 0}$ and $(\theta^k)_{k\geq 0}$ be given by* (MNAG)*. Define for $k \geq 0$ the optimization problem*

$$(\mathrm{P}_k) \quad \begin{aligned}
&\min_{(d,\beta)\in\mathcal{H}\times\mathbb{R}} \Phi(d,\beta) := \frac{1}{2}\left\|hd + y^k - x^k\right\|^2 + \beta, \\
&\text{s.t.} \quad g_i(d,\beta) := \left\langle h\nabla f_i(y^k) - (y^k - x^k), hd + (y^k - x^k)\right\rangle - \beta \leq 0, \ \textit{for } i = 1, \dots, m.
\end{aligned}$$

Then, for all $k \geq 0$ the dual problem to (P_k) *is the quadratic problem*

$$(\mathrm{D}_k) \quad \begin{aligned}
&\min_{\theta\in\mathbb{R}^m} \frac{1}{2}\left\| h\sum_{i=1}^m \theta_i \nabla f_i(y^k) + x^k - y^k \right\|^2, \\
&\text{s.t.} \quad \sum_{i=1}^m \theta_i = 1, \\
&\qquad \theta_i \geq 0, \quad \textit{for} \quad i = 1, \dots, m,
\end{aligned}$$

and the optimal solution θ^k *to* (D_k) *satisfies*

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

Proof. Since $\mathcal{H}$ can be 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 [155]. The optimization problem (P_k) has a fairly simple structure and therefore the duality between (P_k) and (D_k) follows from a straightforward computation. Since the objective function $\Phi(\cdot,\cdot)$ of (P_k) is convex and all constraints $g_i(\cdot,\cdot)$ are linear, strong duality holds. Hence a KKT point $((d^k, \beta_k), \theta^k) \in (\mathcal{H} \times \mathbb{R}) \times \mathbb{R}^m$ of problem (P_k) yields a solution to (D_k). From the KKT conditions for (P_k) we get that

$$d^k = -h \sum_{i=1}^{m} \theta_i^k \nabla f_i(y^k).$$

By primal feasibility, $g_i(d^k, \beta_k) \leq 0$ holds for all $i = 1, \dots, m$ and hence

$$\left\langle h \nabla f_i(y^k) - (y^k - x^k), h d^k + (y^k - x^k) \right\rangle \leq \beta_k.$$

By complementarity of θ_i^k and $g_i(d^k, \beta_k)$, we get

$$\begin{aligned} &\left\langle h \sum_{i=1}^{m} \theta_i^k \nabla f_i(y^k) - (y^k - x^k), h d^k + (y^k - x^k) \right\rangle = \beta_k \\ &= \max_{i=1,\dots,m} \left\langle h \nabla f_i(y^k) - (y^k - x^k), h d^k + (y^k - x^k) \right\rangle. \end{aligned} \tag{5.13}$$

The second equality in (5.13) follows from the fact that $\theta_j^k > 0$ holds for at least one $j \in \{1, \dots, m\}$ as a consequence of dual feasibility. Using $d^k = -\sum_{i=1}^{m} \theta_i^k \nabla f_i(y^k)$, we get $h d^k = x^{k+1} - y^k$ and therefore

$$\begin{aligned} &\left\langle h \sum_{i=1}^{m} \theta_i^k \nabla f_i(y^k) - (y^k - x^k), x^{k+1} - x^k \right\rangle \\ &= \max_{i=1,\dots,m} \left\langle h \nabla f_i(y^k) - (y^k - x^k), x^{k+1} - x^k \right\rangle. \end{aligned}$$

□

Lemma 5.2.6. *Assume the objective functions* f_i *are convex and continuously differentiable with L-Lipschitz continuous gradients* ∇f_i *for all* $i = 1, \dots, m$. *Let* $\alpha > 0$ *and* $0 < h \leq \frac{1}{L}$ *and let* $(x^k)_{k \geq 0}$ *and* $(y^k)_{k \geq 0}$ *be given by* (MNAG). *Then, for all* $z \in \mathcal{H}$ *and all* $k \geq 0$, *it holds that*

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

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

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

and therefore for all $z \in \mathcal{H}$ and all $k \geq 0$

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

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

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

Now, we use Lemma 5.2.5 and get the equality

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

From here, we continue by using the definitions of $(x^k)_{k\geq 0}$, $(y^k)_{k\geq 0}$ and $(\theta^k)_{k\geq 0}$ from (MNAG) to get

$$\begin{aligned} &= \frac{1}{h}\langle y^k - x^{k+1}, x^{k+1} - x^k \rangle + \frac{1}{2h}\|x^{k+1} - y^k\|^2 \\ &= -\frac{1}{h}\langle x^{k+1} - y^k, y^k - x^k \rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2. \end{aligned}$$

□

Corollary 5.2.7. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 0$ and $0 < h \leq \frac{1}{L}$ and let $(x^k)_{k\geq 0}$ and $(y^k)_{k\geq 0}$ be given by* (MNAG)*. Then, for all $z \in \mathcal{H}$ and all $1 \leq k_1 \leq k_2$*

$$\begin{aligned} \varphi_{z,k_2} - \varphi_{z,k_1} \leq\ & \frac{1}{2h}\left[\left\|x^{k_1} - x^{k_1-1}\right\|^2 - \left\|x^{k_2} - x^{k_2-1}\right\|^2\right] \\ & + \frac{1}{2h}\sum_{k=k_1}^{k_2-1}\left(\left(\frac{k-1}{k+\alpha-1}\right)^2 - 1\right)\|x^k - x^{k-1}\|^2. \end{aligned}$$

Proof. We start from the inequality presented in Lemma 5.2.6 and perform some straight forward manipulations. For all $k \geq 1$ we have

$$\begin{aligned} \varphi_{z,k+1} - \varphi_{z,k} \leq\ & -\frac{1}{h}\langle x^{k+1} - y^k, y^k - x^k \rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2 \\ =\ & -\frac{1}{2h}\left[\|x^{k+1} - x^k\|^2 - \|x^{k+1} - y^k\|^2 - \|y^k - x^k\|^2\right] - \frac{1}{2h}\|x^{k+1} - y^k\|^2 \quad (5.14) \\ =\ & \frac{1}{2h}\left[\|y^k - x^k\|^2 - \|x^{k+1} - x^k\|^2\right]. \end{aligned}$$

Now, we use the relation $y^k = x^k + \frac{k-1}{k+\alpha-1}(x^k - x^{k-1})$ to follow

$$
\begin{aligned}
&= \frac{1}{2h}\left[\left(\frac{k-1}{k+\alpha-1}\right)^2 \|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right] \\
&= \frac{1}{2h}\left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right] + \frac{1}{2h}\left(\left(\frac{k-1}{k+\alpha-1}\right)^2 - 1\right)\|x^k - x^{k-1}\|^2.
\end{aligned}
\tag{5.15}
$$

Together, (5.14) and (5.15) give

$$
\begin{aligned}
\varphi_{z,k+1} - \varphi_{z,k} \le{}& \frac{1}{2h}\left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right] \\
&+ \frac{1}{2h}\left(\left(\frac{k-1}{k+\alpha-1}\right)^2 - 1\right)\|x^k - x^{k-1}\|^2.
\end{aligned}
\tag{5.16}
$$

Summing this inequality from $k = k_1$ to $k_2 - 1$ gives the desired result. ☐

5.2.3 Convergence of function values with rate $\mathcal{O}(k^{-2})$

In this subsection, we combine the partial results presented in the preceding subsection to conclude asymptotic convergence rates of the function values of the sequence $(x^k)_{k\ge 0}$ defined by (MNAG) measured with the merit function $\varphi(\cdot)$ defined in (2.23). Before further investigating the rate of $\varphi(x^k)$, in the following theorem, we present a weaker result on the sequence $(\varphi_{z,k})_{k\ge 0}$ for an arbitrary $z \in \mathcal{H}$.

Theorem 5.2.8. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha \ge 3$ and $0 < h \le \frac{1}{L}$. Let $z \in \mathcal{H}$ be arbitrary and let the sequence $(x^k)_{k\ge 0}$ be given by* (MNAG) *with initial iterates $x^0 = x^{-1} \in \mathcal{H}$. Then, the sequence $(\varphi_{z,k})_{k\ge 0}$ defined in* (5.10) *satisfies for all $k \ge 1$*

$$
\left[(k+\alpha-2)^2 + \frac{(k-1)k(\alpha-3)}{2} + (k-1)(\alpha-2)^2\right]\varphi_{z,k} \le \frac{(\alpha-1)^2}{2h}\|x^0 - z\|^2.
$$

Proof. Lemma 5.2.4 and Lemma 5.2.6 state for all $k \ge 0$

$$
\begin{aligned}
\varphi_{z,k+1} &\le -\frac{1}{h}\langle x^{k+1} - y^k, y^k - z\rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2 \quad \text{and} \\
\varphi_{z,k+1} - \varphi_{z,k} &\le -\frac{1}{h}\langle x^{k+1} - y^k, y^k - x^k\rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2.
\end{aligned}
$$

Taking a convex combination of these inequalities with weights $\frac{\alpha-1}{k+\alpha-1}$ and $\frac{k}{k+\alpha-1}$ yields

$$
\begin{aligned}
&\varphi_{z,k+1} - \frac{k}{k+\alpha-1}\varphi_{z,k} \\
&\le -\frac{1}{h}\left\langle x^{k+1} - y^k, y^k - \frac{k}{k+\alpha-1}x^k - \frac{\alpha-1}{k+\alpha-1}z\right\rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2 \\
&= \frac{1}{h}\left\langle x^{k+1} - y^k, \frac{k}{k+\alpha-1}(x^k - y^k) + \frac{\alpha-1}{k+\alpha-1}(z - y^k)\right\rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2.
\end{aligned}
\tag{5.17}
$$

Define

$$z^k := \frac{k+\alpha-1}{\alpha-1}y^k - \frac{k}{\alpha-1}x^k = x^k + \frac{k-1}{\alpha-1}(x^k - x^{k-1}), \tag{5.18}$$

and notice that

$$\frac{k}{k+\alpha-1}(y^k - x^k) + \frac{\alpha-1}{k+\alpha-1}(y^k - z) = \frac{\alpha-1}{k+\alpha-1}(z^k - z). \tag{5.19}$$

Using identity (5.19) in (5.17), we get

$$\varphi_{z,k+1} - \frac{k}{k+\alpha-1}\varphi_{z,k} \leq -\frac{\alpha-1}{h(k+\alpha-1)}\langle x^{k+1} - y^k, z^k - z\rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2. \tag{5.20}$$

From the definition of z^k in (5.18), one can see that

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

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

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

Rearranging this identity and multiplying with $\frac{(\alpha-1)^2}{h(k+\alpha-1)^2}$ yields

$$\begin{aligned} &\frac{(\alpha-1)^2}{2h(k+\alpha-1)^2}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right) \\ &= -\frac{\alpha-1}{h(k+\alpha-1)}\langle z^k - z, x^{k+1} - y^k\rangle - \frac{1}{2h}\|x^{k+1} - y^k\|^2. \end{aligned} \tag{5.21}$$

Combining (5.20) and (5.21), in total we have

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

Then, multiplying both sides with $(k+\alpha-1)^2$ yields

$$(k+\alpha-1)^2\varphi_{z,k+1} - k(k+\alpha-1)\varphi_{z,k} \leq \frac{(\alpha-1)^2}{2h}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right). \tag{5.22}$$

We use $k(k+\alpha-1) = (k+\alpha-2)^2 - k(\alpha-3) - (\alpha-2)^2$ and write

$$\begin{aligned} &(k+\alpha-1)^2\varphi_{z,k+1} - (k+\alpha-2)^2\varphi_{z,k} \\ &\leq -\left(k(\alpha-3) + (\alpha-2)^2\right)\varphi_{z,k} + \frac{(\alpha-1)^2}{2h}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right). \end{aligned} \tag{5.23}$$

Let $k \geq 0$ and $k-1 \leq n$. We use Corollary 5.2.7 with $k_1 = k$ and $k_2 = n+1$ to conclude

$$\begin{aligned} -\varphi_{z,k} \leq &-\varphi_{z,n+1} + \frac{1}{2h}\left[\left\|x^k - x^{k-1}\right\|^2 - \left\|x^{n+1} - x^n\right\|^2\right] \\ &+ \frac{1}{2h}\sum_{l=k}^{n}\left(\left(\frac{l-1}{l+\alpha-1}\right)^2 - 1\right)\|x^l - x^{l-1}\|^2. \end{aligned} \tag{5.24}$$

We use (5.24) on the right-hand side of (5.23), to derive

$$
\begin{aligned}
&(k+\alpha-1)^2\varphi_{z,k+1} - (k+\alpha-2)^2\varphi_{z,k} \leq \\
&-\left(k(\alpha-3)+(\alpha-2)^2\right)\varphi_{z,n+1} + \frac{1}{2h}\left(k(\alpha-3)+(\alpha-2)^2\right)\|x^k - x^{k-1}\|^2 \\
&+ \frac{1}{2h}\left(k(\alpha-3)+(\alpha-2)^2\right)\sum_{l=k}^{n}\left(\left(\frac{l-1}{l+\alpha-1}\right)^2 - 1\right)\|x^l - x^{l-1}\|^2 \\
&+ \frac{(\alpha-1)^2}{2h}\left[\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right].
\end{aligned}
$$

Summing this inequality from $k=1$ to $k=n$, we follow

$$
\begin{aligned}
&(n+\alpha-1)^2\varphi_{z,n+1} - (\alpha-1)^2\varphi_{z,1} \\
&\leq -\sum_{k=1}^{n}\left(k(\alpha-3)+(\alpha-2)^2\right)\varphi_{z,n+1} + \frac{1}{2h}\sum_{k=1}^{n}\left(k(\alpha-3)+(\alpha-2)^2\right)\|x^k - x^{k-1}\|^2 \\
&\quad + \frac{1}{2h}\sum_{k=1}^{n}\sum_{l=k}^{n}\left(k(\alpha-3)+(\alpha-2)^2\right)\left(\left(\frac{l-1}{l+\alpha-1}\right)^2 - 1\right)\|x^l - x^{l-1}\|^2 \\
&\quad + \frac{(\alpha-1)^2}{2h}\left[\|z^1 - z\|^2 - \|z^{n+1} - z\|^2\right].
\end{aligned}
\tag{5.25}
$$

For all sequences $(a_{k,l})_{k,l\geq 1} \subseteq \mathbb{R}$ and all $n \geq 1$ it holds that

$$
\sum_{k=1}^{n}\sum_{l=k}^{n} a_{k,l} = \sum_{l=1}^{n}\sum_{k=1}^{l} a_{k,l}. \tag{5.26}
$$

We apply the identity (5.26) to the double sum in (5.25) to obtain

$$
\begin{aligned}
&\sum_{k=1}^{n}\sum_{l=k}^{n}\left(k(\alpha-3)+(\alpha-2)^2\right)\left(\left(\frac{l-1}{l+\alpha-1}\right)^2 - 1\right)\|x^l - x^{l-1}\|^2 \\
&= \sum_{k=1}^{n}\left(\frac{k(k+1)(\alpha-3)}{2} + k(\alpha-2)^2\right)\left(\left(\frac{k-1}{k+\alpha-1}\right)^2 - 1\right)\|x^k - x^{k-1}\|^2.
\end{aligned}
\tag{5.27}
$$

We use this to rewrite the sums

$$
\begin{aligned}
&\sum_{k=1}^{n} \left(k(\alpha-3)+(\alpha-2)^2\right)\|x^k-x^{k-1}\|^2 \\
&+\sum_{k=1}^{n}\sum_{l=k}^{n} \left(k(\alpha-3)+(\alpha-2)^2\right)\left(\left(\frac{l-1}{l+\alpha-1}\right)^2-1\right)\|x^l-x^{l-1}\|^2 \\
&=\sum_{k=1}^{n} \left(k(\alpha-3)+(\alpha-2)^2\right)\|x^k-x^{k-1}\|^2 \\
&+\sum_{k=1}^{n}\left(\frac{k(k+1)(\alpha-3)}{2}+k(\alpha-2)^2\right)\left(\left(\frac{k-1}{k+\alpha-1}\right)^2-1\right)\|x^k-x^{k-1}\|^2 \\
&=\sum_{k=1}^{n} k(\alpha-3)\left[1+\frac{k+1}{2}\left(\left(\frac{k-1}{k+\alpha-1}\right)^2-1\right)\right]\|x^k-x^{k-1}\|^2 \\
&+\sum_{k=1}^{n}(\alpha-2)^2\left[1+k\left(\left(\frac{k-1}{k+\alpha-1}\right)^2-1\right)\right]\|x^k-x^{k-1}\|^2.
\end{aligned}
\tag{5.28}
$$

For all $k \geq 1$, we have

$$
1+k\left(\left(\frac{k-1}{k+\alpha-1}\right)^2-1\right) \leq 1+\frac{k+1}{2}\left(\left(\frac{k-1}{k+\alpha-1}\right)^2-1\right) \leq 0,
$$

and this combined with (5.28) gives

$$
\begin{aligned}
&\sum_{k=1}^{n} \left(k(\alpha-3)+(\alpha-2)^2\right)\|x^k-x^{k-1}\|^2 \\
&+\sum_{k=1}^{n}\sum_{l=k}^{n} \left(k(\alpha-3)+(\alpha-2)^2\right)\left(\left(\frac{l-1}{l+\alpha-1}\right)^2-1\right)\|x^l-x^{l-1}\|^2 \leq 0.
\end{aligned}
\tag{5.29}
$$

Combining (5.25) and (5.29) we get

$$
\begin{aligned}
(n+\alpha-1)^2\varphi_{z,n+1}-(\alpha-1)^2\varphi_{z,1} \leq &-\sum_{k=1}^{n}\left(k(\alpha-3)+(\alpha-2)^2\right)\varphi_{z,n+1} \\
&+\frac{(\alpha-1)^2}{2h}\left[\|z^1-z\|^2-\|z^{n+1}-z\|^2\right].
\end{aligned}
\tag{5.30}
$$

Using $z^1 = x^1$, inequality (5.30) simplifies to

$$
\left[(n+\alpha-1)^2+\frac{n(n+1)(\alpha-3)}{2}+n(\alpha-2)^2\right]\varphi_{z,n+1} \leq (\alpha-1)^2\varphi_{z,1}+\frac{(\alpha-1)^2}{2h}\|x^1-z\|^2.
$$

From Lemma 5.2.4, we derive

$$
\varphi_{z,1} \leq \frac{1}{2h}\|x^0-z\|^2-\frac{1}{2h}\|x^1-z\|^2,
$$

and we obtain for all $k \geq 1$

$$\left[(k+\alpha-2)^2 + \frac{(k-1)k(\alpha-3)}{2} + (k-1)(\alpha-2)^2\right]\varphi_{z,k} \leq \frac{(\alpha-1)^2}{2h}\|x^0 - z\|^2.$$

□

The theorem above is not straight forward to interpret since the sequence $(\varphi_{z,k})_{k\geq 0}$ does not remain positive for all $k \geq 0$ and we only get an upper bound of order $\mathcal{O}(k^{-2})$. This on its own does not imply the convergence of $f(x^k) = \big(f_1(x^k), \dots, f_m(x^k)\big)$ to an element of the Pareto front. However we can refine the statement of Theorem 5.2.8 in the following way to get a stronger convergence statement.

Theorem 5.2.9. *Assume that Assumption* $(\mathcal{A}_2)$ *holds and assume the objective functions* f_i *are convex and continuously differentiable with L-Lipschitz continuous gradients* ∇f_i *for all* $i = 1, \dots, m$. *Let* $\alpha \geq 3$ *and* $0 < h \leq \frac{1}{L}$. *Let the sequence* $(x^k)_{k\geq 0}$ *be given by* (MNAG) *with initial iterates* $x^0 = x^{-1} \in \mathcal{H}$. *Then, for all* $k \geq 1$

$$\varphi(x^k) \leq \frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2},$$

and hence

$$\varphi(x^k) = \mathcal{O}\left(\frac{1}{k^2}\right) \quad \text{as} \quad k \to +\infty.$$

Proof. Theorem 5.2.8 gives for all $z \in \mathcal{H}$

$$\left[(k+\alpha-2)^2 + \frac{(k-1)k(\alpha-3)}{2} + (k-1)(\alpha-2)^2\right]\varphi_{z,k} \leq \frac{(\alpha-1)^2}{2h}\|x^0 - z\|^2. \tag{5.31}$$

By Assumption $(\mathcal{A}_2)$ and Lemma 2.3.15, we have

$$\sup_{F^* \in F(\mathcal{LP}_w(F, F(x^0)))} \inf_{z \in F^{-1}(\{F^*\})} \varphi_{z,k} = \varphi(x^k).$$

Then, by the definition of $R > 0$ in Assumption $(\mathcal{A}_2)$ we conclude from (5.31)

$$\left[(k+\alpha-2)^2 + \frac{(k-1)k(\alpha-3)}{2} + (k-1)(\alpha-2)^2\right]\varphi(x^k) \leq \frac{(\alpha-1)^2 R}{h}.$$

Since $\varphi(x^k) \geq 0$ and $(k+\alpha-2)^2 + \frac{(k-1)k(\alpha-3)}{2} + (k-1)(\alpha-2)^2 \geq (k+\alpha-2)^2$ we get

$$\varphi(x^k) \leq \frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}.$$

□

The following corollary can be derived from Theorem 5.2.9.

Corollary 5.2.10. *Assume that Assumption* $(\mathcal{A}_2)$ *holds and assume the objective functions* f_i *are convex and continuously differentiable with L-Lipschitz continuous gradients* ∇f_i *for all* $i = 1, \dots, m$. *Let* $\alpha \geq 3$ *and* $0 < h \leq \frac{1}{L}$. *Let the sequence* $(x^k)_{k\geq 0}$ *be given by* (MNAG) *with initial iterates* $x^0 = x^{-1} \in \mathcal{H}$. *Then every weak sequential cluster point of* $(x^k)_{k\geq 0}$ *is weakly Pareto optimal*

Proof. The statement follows by the last Theorem and the fact that $\varphi(\cdot)$ is weakly lower semi-continuous by Theorem 2.3.14. $\square$

5.2.4 Weak convergence of iterates

In this subsection we prove the weak convergence of the iterates $(x^k)_{k\geq 0}$ to a weakly Pareto optimal point of (MOP) using Opial's Lemma (Lemma 2.1.6). Before we can apply Opial's Lemma we have to derive asymptotic bounds on $\|x^k - x^{k-1}\|$.

Lemma 5.2.11. *Assume the objective functions* f_i *are convex and continuously differentiable with L-Lipschitz continuous gradients* ∇f_i *for all* $i = 1, \dots, m$. *Let* $\alpha > 3$ *and* $0 < h \leq \frac{1}{L}$. *Let* $z \in \mathcal{H}$ *be arbitrary and let the sequence* $(x^k)_{k\geq 0}$ *be given by* (MNAG) *and* $(\varphi_{z,k})_{k\geq 0}$ *be given by* (5.10). *Then for all* $K \geq 1$ *it holds that*

$$\sum_{k=1}^{K} \left[k + \frac{(\alpha-2)^2}{\alpha-3} \right] \varphi_{z,k} \leq (\alpha-1)^2 \varphi_{z,1} - (K+\alpha-1)^2 \varphi_{z,K+1} + \frac{(\alpha-1)^2}{2h} \|x^1 - z\|^2.$$

Proof. The proof follows immediately by (5.23). $\square$

Lemma 5.2.12. *Assume the objective functions* f_i *are convex and continuously differentiable with L-Lipschitz continuous gradients* ∇f_i *for all* $i = 1, \dots, m$. *Let* $\alpha > 3$ *and* $0 < h \leq \frac{1}{L}$. *Let* $z \in \mathcal{H}$ *be arbitrary and let the sequence* $(x^k)_{k\geq 0}$ *be given by* (MNAG) *and* $(\varphi_{z,k})_{k\geq 0}$ *be given by* (5.10) *and define* $\beta := \frac{(\alpha-2)^2}{\alpha-3} + \frac{1}{2}$. *Then for all* $K \geq 1$ *it holds that*

$$\begin{aligned}
&\frac{1}{2h}(K+1)^2 \|x^{K+1} - x^K\|^2 + \frac{1}{h} \sum_{k=1}^{K} (k+1) \|x^{k+1} - x^k\|^2 \\
&\leq (\beta^2 + 2(\alpha-1)^2)\varphi_{z,1} - \left((K+\beta)^2 + 2(K+\alpha-1)^2 \right) \varphi_{z,K+1} \\
&\quad + \frac{(\alpha-1)^2}{h} \|x^1 - z\|^2 + \frac{(\beta-1)^2}{2h} \|x^1 - x^0\|^2.
\end{aligned}$$

Proof. To abbreviate the notation we introduce the sequence $(d_k)_{k\geq 0}$, with

$$d_k := \frac{1}{2h} \|x^k - x^{k-1}\|^2,$$

for $k \geq 0$. Inequality (5.16) gives for all $k \geq 1$

$$\varphi_{z,k+1} - \varphi_{z,k} \leq \left(\frac{k-1}{k+\alpha-1} \right)^2 d_k - d_{k+1}.$$

For $\beta = \frac{(\alpha-2)^2}{\alpha-3} + \frac{1}{2}$ with $\alpha \geq 3$, we have

$$\frac{k-1}{k+\alpha-1} \leq \frac{k+\beta-2}{k+\beta},$$

and hence

$$\varphi_{z,k+1} - \varphi_{z,k} \leq \left(\frac{k+\beta-2}{k+\beta}\right)^2 d_k - d_{k+1}.$$

We rewrite this into

$$(k+\beta)^2 d_{k+1} - (k+\beta-2)^2 d_k \leq (k+\beta)^2(\varphi_{z,k} - \varphi_{z,k+1})$$

We use $(k+\beta)^2 = (k+\beta-1)^2 + 2\left(k+\beta-\frac{1}{2}\right)$ and hence

$$\begin{aligned}
&(k+\beta-1)^2 d_{k+1} - (k+\beta-2)^2 d_k + 2\left(k+\beta-\frac{1}{2}\right) d_{k+1} \leq \\
&(k+\beta-1)^2 \varphi_{z,k} - (k+\beta)^2 \varphi_{z,k+1} + 2\left(k+\beta-\frac{1}{2}\right) \varphi_{z,k}.
\end{aligned}$$

We use $\beta - \frac{1}{2} = \frac{(\alpha-2)^2}{\alpha-3}$ and $k+\beta-\frac{1}{2} \geq 1$ with $d_{k+1} \geq 0$ to conclude

$$\begin{aligned}
&(k+\beta-1)^2 d_{k+1} - (k+\beta-2)^2 d_k + 2\,(k+1)\, d_{k+1} \leq \\
&(k+\beta-1)^2 \varphi_{z,k} - (k+\beta)^2 \varphi_{z,k+1} + 2\left(k + \frac{(\alpha-2)^2}{\alpha-3}\right) \varphi_{z,k}.
\end{aligned}$$

We sum this inequality from $k = 1, \ldots, K$ to follow

$$\begin{aligned}
&(K+\beta-1)^2 d_{K+1} - (\beta-1)^2 d_1 + 2\sum_{k=1}^{K} (k+1)\, d_{k+1} \leq \\
&\beta^2 \varphi_{z,1} - (K+\beta)^2 \varphi_{z,K+1} + 2\sum_{k=1}^{K} \left(k + \frac{(\alpha-2)^2}{\alpha-3}\right) \varphi_{z,k}.
\end{aligned}$$

By Lemma 5.2.11 we follow

$$\begin{aligned}
&(K+\beta-1)^2 d_{K+1} - (\beta-1)^2 d_1 + 2\sum_{k=1}^{K} (k+1)\, d_{k+1} \leq \\
&\left(\beta^2 + 2(\alpha-1)^2\right) \varphi_{z,1} - \left((K+\beta)^2 + 2(K+\alpha-1)^2\right) \varphi_{z,K+1} + \frac{(\alpha-1)^2}{h}\|x^1 - z\|^2.
\end{aligned}$$

We rewrite this into

$$\begin{aligned}
&(K+\beta-1)^2 d_{K+1} + 2\sum_{k=1}^{K} (k+1)\, d_{k+1} \leq \\
&\left(\beta^2 + 2(\alpha-1)^2\right) \varphi_{z,1} - \left((K+\beta)^2 + 2(K+\alpha-1)^2\right) \varphi_{z,K+1} \\
&+ \frac{(\alpha-1)^2}{h}\|x^1 - z\|^2 + (\beta-1)^2 d_1,
\end{aligned}$$

which completes the proof. □

Building on the previous Lemma, in the following theorem we state asymptotic bounds on $\|x^k - x^{k-1}\|$ as $k \to +\infty$.

Theorem 5.2.13. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 3$ and $0 < h \leq \frac{1}{L}$. Let the sequence $(x^k)_{k\geq 0}$ be given by* (MNAG) *and assume it is bounded. Then the following statements hold.*

i) $\|x^{k+1} - x^k\| = \mathcal{O}\left(k^{-1}\right)$ *as* $k \to +\infty$*;*

ii) $\displaystyle\sum_{k=1}^{+\infty} k\|x^{k+1} - x^k\|^2 < +\infty$.

Proof. We use the notation $d_k := \frac{1}{2h}\|x^k - x^{k-1}\|^2$ for $k \geq 0$. Lemma 5.2.12 with $z = x^{K+1}$, gives for all $K \geq 1$

$$\begin{aligned}
&(K+\beta-1)^2 d_{K+1} + 2\sum_{k=1}^{K}(k+1)\,d_{k+1} \leq \\
&\left(\beta^2 + 2(\alpha-1)^2\right)\min_{i=1,\dots,m} f_i(x^1) - f_i(x^{K+1}) + \frac{(\alpha-1)^2}{h}\left\|x^1 - x^{K+1}\right\|^2 + (\beta-1)^2 d_1.
\end{aligned} \tag{5.32}$$

Since $(x^k)_{k\geq 0}$ is bounded by assumption, the right hand side of (5.32) is bounded by a constant $C > 0$ and we follow for all $k, K \geq 0$

$$\|x^{k+1} - x^k\|^2 \leq \frac{C}{k^2} \quad \text{and} \quad \sum_{k=1}^{K} k\|x^{k+1} - x^k\|^2 \leq C,$$

which completes the proof. □

Lemma 5.2.14. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 3$ and $0 < h \leq \frac{1}{L}$. Let $(x^k)_{k\geq 0}$ and $(y^k)_{k\geq 0}$ be the sequences generated by* (MNAG). *Define the set*

$$S := \left\{ z \in \mathcal{H} \,:\, f_i(z) \leq \liminf_{k\to+\infty} f_i(x^k) \quad \textit{for all} \quad i = 1, \dots, m \right\}.$$

If $(x^k)_{k\geq 0}$ is bounded then $S \neq \emptyset$. Further, for all $z \in S$ and all $k \geq 1$

$$\left\langle x^{k+1} - y^k, x^{k+1} - z \right\rangle \leq \frac{1}{2}\|x^{k+1} - x^k\|^2 + \frac{1}{2}\|x^{k+1} - y^k\|^2. \tag{5.33}$$

Proof. By convexity and weak lower semicontinuity, we know that every weak limit point of x^k belongs to S and by boundedness of x^k there exists at least one cluster point and hence S is nonempty. Let $z \in S$. We start from (5.11) and write

$$f_i(x^{k+1}) - f_i(z) \leq \langle \nabla f_i(y^k), x^{k+1} - z\rangle + \frac{L}{2}\|x^{k+1} - y^k\|^2. \tag{5.34}$$

Since $z \in S$ we conclude by Proposition 5.2.1 for all $i = 1, \dots, m$ and all $k \geq 0$

$$f_i(x^{k+1}) - f_i(z) \geq -\frac{1}{2h}\|x^{k+1} - x^k\|^2. \tag{5.35}$$

Combining (5.34) and (5.35) gives

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

and in total we observe

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

□

Theorem 5.2.15. *Assume the objective functions f_i are convex and continuously differentiable with L-Lipschitz continuous gradients ∇f_i for all $i = 1, \dots, m$. Let $\alpha > 3$ and $0 < h \leq \frac{1}{L}$. Assume $(x^k)_{k \geq 0}$ is bounded and define*

$$S := \left\{ z \in \mathcal{H} \, : \, f_i(z) \leq \liminf_{k \to +\infty} f_i(x^k) \quad \text{for all} \quad k \geq 0 \right\}.$$

Then $x^k \rightharpoonup x^\infty \in S$ converges weakly in $\mathcal{H}$ as $k \to +\infty$ and x^∞ is weakly Pareto optimal.

Proof. By Lemma 5.2.14 it holds that $S \neq \emptyset$. Let $z \in S$ and define

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

Simple manipulations give

$$\begin{aligned} h_{z,k+1} - h_{z,k} = & \frac{1}{2}\|x^{k+1} - z\|^2 - \frac{1}{2}\|x^k - z\|^2 \\ = & \langle x^{k+1} - x^k, x^{k+1} - z \rangle - \frac{1}{2}\|x^{k+1} - x^k\|^2 \\ = & \langle x^{k+1} - y^k, x^{k+1} - z \rangle - \frac{1}{2}\|x^{k+1} - x^k\|^2 + \langle y^k - x^k, x^{k+1} - z \rangle. \end{aligned} \tag{5.36}$$

We apply (5.33) and bound this by

$$\begin{aligned} \leq & \frac{1}{2}\|x^{k+1} - y^k\|^2 - \|x^{k+1} - x^k\|^2 + \langle y^k - x^k, x^{k+1} - z \rangle \\ = & \frac{1}{2}\|x^{k+1} - y^k\|^2 - \|x^{k+1} - x^k\|^2 + \frac{k-1}{k+\alpha-1}\langle x^k - x^{k-1}, x^{k+1} - z \rangle. \end{aligned} \tag{5.37}$$

Analogously, we have

$$h_{z,k} - h_{z,k-1} = \langle x^k - x^{k-1}, x^k - z \rangle - \frac{1}{2}\|x^k - x^{k-1}\|^2. \tag{5.38}$$

Combining (5.36), (5.37) and (5.38), we get

$$\begin{aligned} h_{z,k+1} - h_{z,k} - \frac{k-1}{k+\alpha-1}(h_{z,k} - h_{z,k-1}) &\le \frac{1}{2}\|x^{k+1} - y^k\|^2 - \|x^{k+1} - x^k\|^2 \\ + \frac{k-1}{k+\alpha-1}\left[\langle x^k - x^{k-1}, x^{k+1} - x^k\rangle + \frac{1}{2}\|x^k - x^{k-1}\|^2\right]. \end{aligned} \tag{5.39}$$

We use the following identity in (5.39)

$$\begin{aligned} \|x^{k+1} - y^k\|^2 =& \|x^{k+1} - x^k\|^2 - 2\frac{k-1}{k+\alpha-1}\langle x^{k+1} - x^k, x^k - x^{k-1}\rangle \\ &+ \left(\frac{k-1}{k+\alpha-1}\right)^2 \|x^k - x^{k-1}\|^2, \end{aligned}$$

to obtain

$$\begin{aligned} & h_{z,k+1} - h_{z,k} - \frac{k-1}{k+\alpha-1}(h_{z,k} - h_{z,k-1}) \\ \le & -\frac{1}{2}\|x^{k+1} - x^k\|^2 + \frac{1}{2}\left(\frac{k-1}{k+\alpha-1}\right)^2 \|x^k - x^{k-1}\|^2 + \frac{1}{2}\frac{k-1}{k+\alpha-1}\|x^k - x^{k-1}\|^2 \\ \le & \|x^k - x^{k-1}\|^2. \end{aligned} \tag{5.40}$$

For a real number $a \in \mathbb{R}$ we define the positive part of a as $[a]_+ := \max(a, 0)$. Using this notion we derive from (5.40)

$$[h_{z,k+1} - h_{z,k}]_+ - \frac{k-1}{k+\alpha-1}[h_{z,k} - h_{z,k-1}]_+ \le \|x^k - x^{k-1}\|^2.$$

The sequence $[h_{z,k+1} - h_{z,k}]_+$ is nonnegative and part *ii)* of Theorem 5.2.13 states $\sum_{k=1}^{+\infty} k\|x^k - x^{k-1}\|^2 < +\infty$. Then, by Lemma 2.2.13

$$\sum_{k=1}^{+\infty} [h_{z,k+1} - h_{z,k}]_+ < +\infty.$$

As $h_{z,k} \ge 0$ for all $k \ge 1$, we follow $\lim_{k\to+\infty} h_{z,k}$ exists. In the beginning of the proof $z \in S$ was chosen arbitrarily and hence $\lim_{k\to+\infty}\|x^k - z\|$ exists for all $z \in S$. By the weak lower semicontinuity of the objective functions each weak sequential cluster point of $(x^k)_{k\ge0}$ belongs to S. Then by Opial's Lemma x^k converges weakly to an element in $x^\infty \in S$. By Corollary 5.2.10 the element x^∞ is weakly Pareto optimal. □

5.3 Numerical experiments

In this section we conduct numerical experiments on the multiobjective Nesterov accelerated gradient method (MNAG). The goal of this experiments is to verify the asymptotic results of the previous subsections. Foremost, we want to verify the bound $\varphi(x^k) \le \frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ given by

Theorem 5.2.9 and compare the behavior of (MNAG) for different values of α. We start with a comparison of the method (MNAG) with its continuous counter part (MAVD) by reconsidering the numerical experiments from Subsection 4.5.5. Additionally, we include comparisons to the multiobjective steepest descent method (MGD), presented in Subsection 2.3.4, to show that the accelerated method (MNAG) in fact converges faster. All numerical experiments in this section where implemented using Matlab R2021b.

5.3.1 Comparison with the continuous system (MAVD)

In the first two numerical experiments, we compare the convergence of the merit function values of the sequence $(x^k)_{k\geq 0}$ given by (MNAG) with the trajectory $x(\cdot)$ given by (MAVD). To this end, we revisit the numerical experiments from Subsection 4.5.5 on (MAVD), where a quadratic multiobjective optimization problem and a nonquadratic convex multiobjective optimization problem get examined.

The two experiments share the following joint parameters. Both problems have two convex continuously differentiable objective functions $f_i : \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto f_i(x)$ for $i = 1, 2$. For both problems we perform $k_{\max} = 1\,000$ iterations of (MNAG) with initial iterate $x^0 = x^{-1} \in \mathbb{R}^2$. The function definitions and the initial iterates get specified in the respective parts. We use step size $h = 1\text{e}{-3}$ and four different acceleration parameters $\alpha \in \{3, 10, 50, 100\}$. We compute the constant $R > 0$ from Assumption $(\mathcal{A}_2)$ using the fact that the weak Pareto sets $\mathcal{P}_w$ can be computed explicitly for the problems (see Subsection 4.5.5). Additionally, we solve $\varphi(x^k) = \sup_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(x^k) - f_i(z)$ for all iterates $k = 1, \dots, k_{\max} = 1\,000$ using Lemma 2.3.15 and the known Pareto set.

A quadratic multiobjective optimization problem

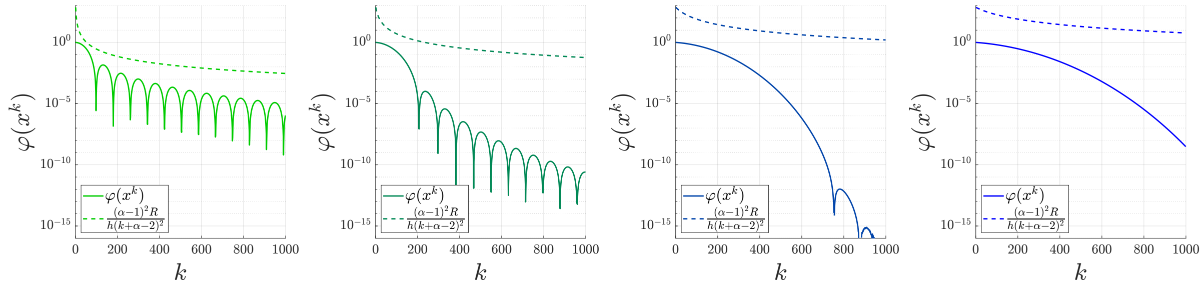
In the first example, we reconsider the following multiobjective optimization problem with two quadratic objective functions

$$f_i : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \frac{1}{2}(x - c^i)^\top Q_i (x - c^i), \tag{5.41}$$

for $i = 1, 2$, given symmetric and positive definite matrices and vectors

$$Q_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad c^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad c^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Using $x^0 = x^{-1} = (-.2, -.1)^\top$ as an initial iterate and the remaining parameters as specified in the beginning of this subsection, we compute $1\,000$ iterations of (MNAG) and the corresponding merit function values $\varphi(x^k)$.



(a) $\alpha = 3$ (b) $\alpha = 10$ (c) $\alpha = 50$ (d) $\alpha = 100$

Figure 5.1: Values of merit function $\varphi(x^k)$ and bound $\frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ for sequence $(x^k)_{k\geq 0}$ given by (MNAG) generated for different values of $\alpha \in \{3, 10, 50, 100\}$.

In Figure 5.1, the evolution of the merit function values $\varphi(x^k)$ and the theoretical bounds $\frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ are shown for $k = 0, \dots, 1\ 000$. The figure includes four subfigures, one for each value of $\alpha \in \{3, 10, 50, 100\}$. For all values of $\alpha \in \{3, 10, 50, 100\}$ the bound $\varphi(x^k) \leq \frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ as stated in Theorem 5.2.9 holds. The inertial features in (MNAG) introduced by the acceleration step lead to oscillations in $\varphi(x^k)$. This is typical for accelerated first-order methods. For small values of α the merit function values $\varphi(x^k)$ experience faster decay in the beginning, but have more oscillations and therefore an over all slower convergence. For bigger values of α the merit function values converge slower in the beginning but have less oscillations in the considered regime and decay faster later on. Depending on the desired tolerance for $\varphi(x^k)$, different values of $\alpha \geq 3$ are preferable. Since the objective functions f_i defined in (5.41) are strongly convex for $i = 1, 2$, the merit function values $\varphi(x^k)$ decay faster than the theoretical bound. The plots are similar to the ones from Subsection 4.5.5 where an experiment on the same objective functions and with the same initial iterate was conducted on trajectory solutions $x(\cdot)$ of the multiobjective gradient system with asymptotic vanishing damping (MAVD). As described in Subsection 5.1.2, the scheme (MNAG) can be interpreted as a discretization of the system (MAVD). Therefore, the similarity in the results is expected.

A nonquadratic multiobjective optimization problem

In the following, we reconsider the second example from Subsection 4.5.5. Define

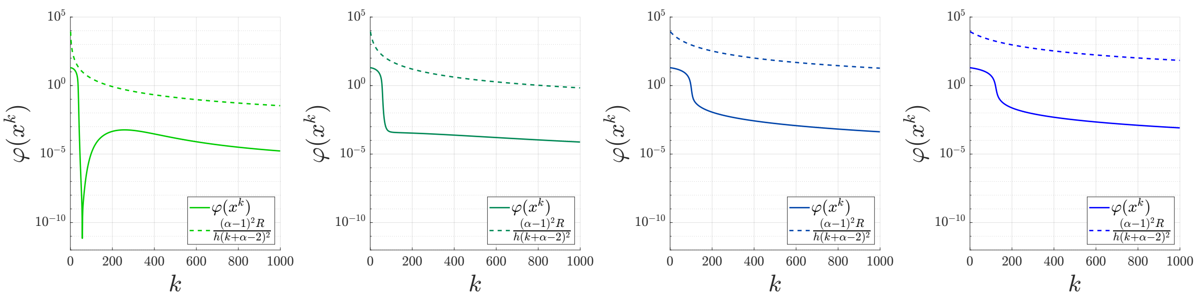
$$f_i : \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \ln\left(\sum_{j=1}^{p} \exp\left(\left(a_j^{(i)}\right)^\top x - b_j^{(i)}\right)\right), \tag{5.42}$$

for $i = 1, 2$ and $p = 4$ with given matrices and vectors

$$A^{(1)} = \begin{pmatrix} \left(a_1^{(1)}\right)^\top \\ \vdots \\ \left(a_4^{(1)}\right)^\top \end{pmatrix} = \begin{pmatrix} 10 & 10 \\ 10 & -10 \\ -10 & -10 \\ -10 & 10 \end{pmatrix}, \quad b^{(1)} = \begin{pmatrix} 0 \\ -20 \\ 0 \\ 20 \end{pmatrix},$$

$$A^{(2)} = \begin{pmatrix} \left(a_1^{(2)}\right)^\top \\ \vdots \\ \left(a_4^{(2)}\right)^\top \end{pmatrix} = \begin{pmatrix} 10 & 10 \\ 10 & -10 \\ -10 & -10 \\ -10 & 10 \end{pmatrix}, \quad b^{(2)} = \begin{pmatrix} 0 \\ 20 \\ 0 \\ -20 \end{pmatrix}.$$

The objective functions defined in (5.42) are convex but not strongly convex. The initial iterate is set to $x^0 = x^{-1} = (0, 3)^\top$ and we choose the remaining parameters as specified in the beginning of this subsection.



(a) $\alpha = 3$ (b) $\alpha = 10$ (c) $\alpha = 50$ (d) $\alpha = 100$

Figure 5.2: Values of merit function $\varphi(x^k)$ and bound $\frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ for sequence $(x^k)_{k \geq 0}$ given by (MNAG) generated for different values of $\alpha \in \{3, 10, 50, 100\}$.

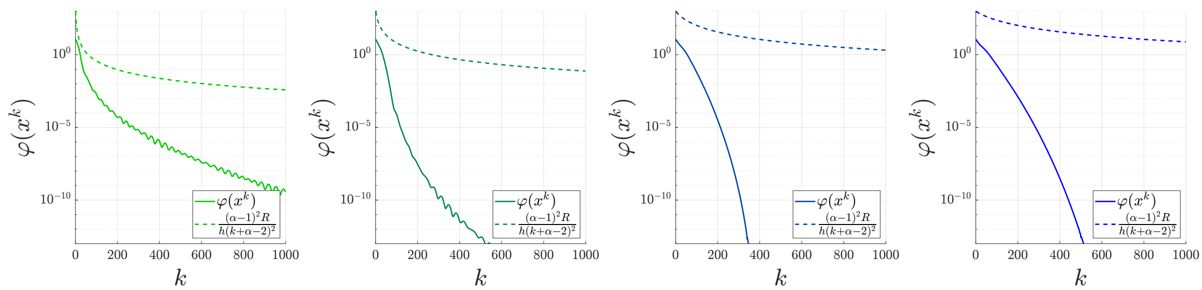
Similar to the last experiment, we compute the sequence $(x^k)_{k \geq 0}$ defined by (MNAG) and plot $\varphi(x^k)$ and the bound $\frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ from Theorem 5.2.9 in Figure 5.2. For all values of $\alpha \in \{3, 10, 50, 100\}$ the inequality $\varphi(x^k) \leq \frac{(\alpha-1)^2 R}{h(k+\alpha-2)^2}$ holds for all $k = 0, \dots, 1\,000$. Because the objective functions in this example are not strongly convex, we see slower decay of the merit function values $\varphi(x^k)$ compared to the last experiment. For the smallest values $\alpha = 3$ the merit function values $\varphi(x^k)$ are not monotone but decay fast in the beginning. For higher values of α we have slower decay and do not observe oscillations. The plots are consistent with the observations made in Subsection 4.5.5, where a similar experiment is conducted for trajectory solutions $x(\cdot)$ of the multiobjective gradient system with asymptotic vanishing damping (MAVD).

5.3.2 Finite dimensional convex multiobjective optimization

Compared to the first two experiments, in this subsection we consider a higher-dimensional example with three objective functions ($m = 3$) defined on $\mathbb{R}^n$ with $n = 100$. We define the objective functions using the following parameters. For $p = 200$, let $A^{(i)} = \left(a_1^{(i)}, \dots, a_p^{(i)}\right)^\top \in \mathbb{R}^{p\times n}$ with $a_j^{(i)} \in \mathbb{R}^n$ for $j = 1, \dots, p$ and $b^{(i)} \in \mathbb{R}^p$ for $i = 1, 2, 3$. Then, for $i = 1, 2, 3$, we define the objective functions

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

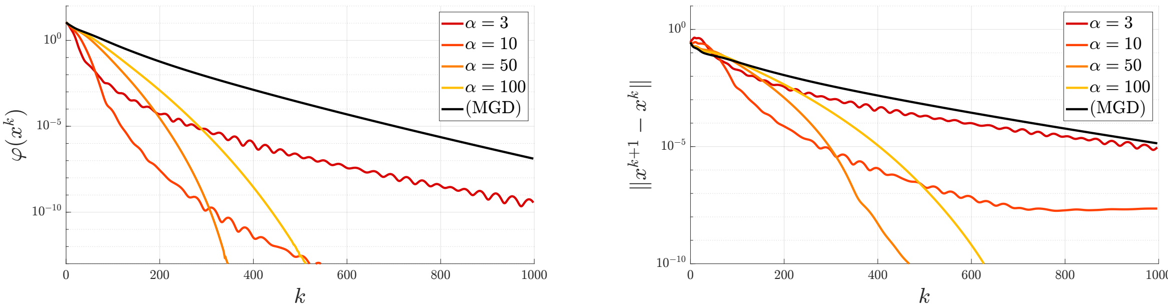
We randomly generate matrices $A^{(i)} \in \mathbb{R}^{p\times n}$ and vectors $b^{(i)} \in \mathbb{R}^p$ with entries uniformly sampled in $[-1, 1]$ for $i = 1, 2, 3$. The initial iterate is set to $x^0 = x^{-1} = [1, \dots, 1]^\top \in \mathbb{R}^n$ and the step size is chosen as $h = 1\mathrm{e}{-1}$. For this parameter settings we execute $k_{\max} = 1\,000$ iterations of (MNAG) for $\alpha \in \{3, 10, 50, 100\}$. For comparison we execute $1\,000$ iterations of the multiobjective steepest descent method (MGD) with the same initial iterate x^0 and with step size $h = 1\mathrm{e}{-1}$. For each iterate we compute $\varphi(x^k) = \sup_{z\in\mathcal{H}} \max_{i=1,\dots,m} f_i(x^k) - f_i(z)$ with a SQP method with sufficient accuracy using the function `fmincon` implemented in Matlab. Additionally, we approximate $R > 0$ defined in Assumption $(\mathcal{A}_2)$ by approximating the Pareto set using a weighted sum approach (see [91]), i.e., for 1275 equidistant weights $\theta^l \in \Delta^m$, we solve the problem $\min_{x\in\mathbb{R}^n} \theta_1^l f_1(x) + \theta_2^l f_2(x) + \theta_3^l f_3(x)$ where each weighted sum problem is solved with a Quasi-Newton method using the function `fminunc` implemented in Matlab.



(a) $\alpha = 3$ (b) $\alpha = 10$ (c) $\alpha = 50$ (d) $\alpha = 100$

Figure 5.3: Values of merit function $\varphi(x^k)$ and bound $\frac{(\alpha-1)^2R}{h(k+\alpha-2)^2}$ for sequence $(x^k)_{k\geq 0}$ given by (MNAG) generated for different values of $\alpha \in \{3, 10, 50, 100\}$.

In Figure 5.3 the merit function values $\varphi(x^k)$ get visualized with one subfigure for each value of $\alpha \in \{3, 10, 50100\}$, respectively. The bound $\varphi(x^k) \leq \frac{(\alpha-1)^2R}{h(k+\alpha-2)^2}$ from Theorem 5.2.9 is satisfied for all $\alpha \in \{3, 10, 50, 100\}$. Depending on the desired tolerance for $\varphi(x^k)$, different values of α are advantageous. For smaller values of α we see faster decay of the merit function values $\varphi(x^k)$ in the beginning but slower decay later on and overall more oscillations. For $\alpha = 50$ the merit function values decay the fastest for the given example. Additionally, Figure 5.4 includes a comparison of (MNAG) and (MGD). Subfigure 5.4a shows the merit function values $\varphi(x^k)$ generated by (MNAG) for $\alpha \in \{3, 10, 50, 100\}$ and (MGD). For all choices of α the merit function values for (MNAG) decay faster than the ones obtained for (MGD). Subfigure 5.4b contains a comparison on the discrete velocities $\|x^{k+1} - x^k\|$. For the accelerated method (MNAG) the velocity increases in the beginning for all values of $\alpha \in \{3, 10, 50, 100\}$, while for (MGD) the velocity is monotonically decreasing. In both subfigures oscillations can be observed in the plots for $\alpha = 3$ and $\alpha = 10$ as a consequence of the inertial features in (MNAG) introduced by the acceleration step. For $\alpha = 50$ and $\alpha = 100$ the merit function values and velocities decay rapidly in comparison to (MGD).



(a) Merit function values $\varphi(x^k)$ (b) Discrete velocities $\|x^{k+1} - x^k\|$

Figure 5.4: Comparison of merit function values $\varphi(x^k)$ and discrete velocities $\|x^{k+1} - x^k\|$ for sequences $(x^k)_{k\geq 0}$ given by (MNAG) for different values of $\alpha \in \{3, 10, 50, 100\}$ and for a sequence given by (MGD).

5.3.3 Infinite dimensional convex multiobjective optimization in $H_0^1(\Omega)$

In the last numerical experiment, we consider a multiobjective optimization problem defined on an infinite dimensional vector space. Given the domain $\Omega = (0,1) \subset \mathbb{R}$ we consider the Sobolev space $H_0^1(\Omega)$ with topological dual $H^{-1}(\Omega)$ (see e.g. [132]). On this space, we define the two objective functions

$$\begin{aligned} f_1 &: H_0^1(\Omega) \to \mathbb{R}, \quad u \mapsto f_1(u) \coloneqq \frac{c_1}{2}\|u - u^{\text{ref}}\|_{L^2(\Omega)}^2, \\ f_2 &: H_0^1(\Omega) \to \mathbb{R}, \quad u \mapsto f_2(u) \coloneqq \frac{c_2}{2}\|\nabla u\|_{L^2(\Omega)}^2, \end{aligned} \tag{5.43}$$

with $c_1, c_2 > 0, u^{\text{ref}} \in L^2(\Omega)$ and where $\nabla : H_0^1(\Omega) \to L^2(\Omega)$ is the bounded linear operator which assigns to a vector $u \in H_0^1(\Omega)$ its weak derivative $\nabla u \in L^2(\Omega)$. Both objective functions are convex and continuously Fréchet differentiable with Lipschitz continuous derivatives

$$\begin{aligned} \mathrm{D}f_1(u) &= c_1\langle u - u^{\text{ref}}, \cdot \rangle_{L^2(\Omega)} \in H^{-1}(\Omega), \\ \mathrm{D}f_2(u) &= c_2\langle \nabla u, \nabla \cdot \rangle_{L^2(\Omega)} \in H^{-1}(\Omega), \end{aligned}$$

with Lipschitz constants c_1 and c_2, respectively. Therefore, we can use (MNAG) to solve the multiobjective optimization problem

$$(\text{MOP-}H_0^1) \qquad \min_{u \in H_0^1(\Omega)} \begin{bmatrix} \frac{c_1}{2}\|u - u^{\text{ref}}\|_{L^2(\Omega)}^2 \\ \frac{c_2}{2}\|\nabla u\|_{L^2(\Omega)}^2 \end{bmatrix} =: \begin{bmatrix} f_1(u) \\ f_2(u) \end{bmatrix}.$$

Problem (MOP-H_0^1) can be interpreted as follows. Given a reference function $u^{\text{ref}} \in L^2(\Omega)$, we want to find a function $u \in H_0^1(\Omega)$ which is as close to u^{ref} as possible and has low H^1-seminorm $\|\nabla u\|_{L^2(\Omega)}$. These objectives are in general conflicting, especially if $u^{\text{ref}} \notin H_0^1(\Omega)$ or if $u^{\text{ref}} \in H_0^1(\Omega)$ but the norm $\|\nabla u^{\text{ref}}\|_{L^2(\Omega)}$ is very large.

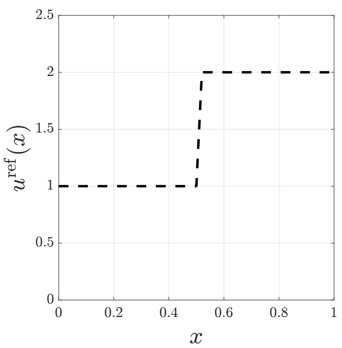
To numerically treat this problem, we use a FEM discretization with Lagrangian $P1$ finite elements on a uniform discretization of $\Omega = (0,1)$ with 49 elements. We use the following reference function in the definition of (5.43). Define

$$u^{\text{ref}} : \Omega \to \mathbb{R}, \quad x \mapsto u^{\text{ref}}(x) := \begin{cases} 1, & \text{if} \quad x \le \frac{1}{2}, \\ 2, & \text{else.} \end{cases} \tag{5.44}$$

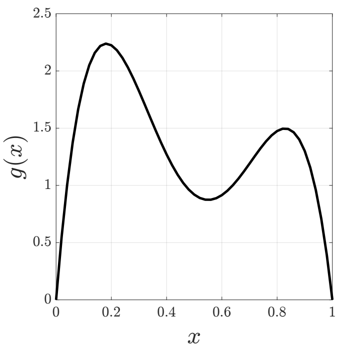
The function $u^{\text{ref}} \in L^2(\Omega)$ is shown in Subfigure 5.5a. Further, we choose constants $c_1 = c_2 = 1\text{e}{-1}$ and step size $h = 1\text{e}{-1}$ and $\alpha \in \{3, 10, 50, 100\}$. As the initial iterate $u^0 = u^{-1} \in H_0^1(\Omega)$ we use a piecewise linear function which nodally interpolates the function

$$g : \Omega \to \mathbb{R}, \quad x \mapsto 80(1-x)x\left(x - \frac{1}{2}\right)^2 + 8(1-x)x\exp(-2x)\,. \tag{5.45}$$

The function g is shown in Subfigure 5.5b. For all values of $\alpha \in \{3, 10, 50, 100\}$ we perform $k_{\max} = 4\,000$ iterations of (MNAG). For comparison, we compute $4\,000$ iterations of (MGD) with step size $h = 1\text{e}{-1}$ and the same initial iterate u^0.

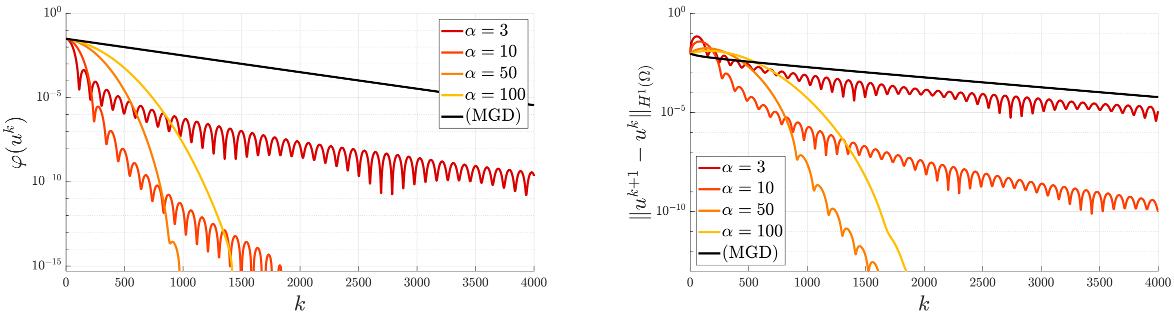


(a) Reference function $u^{\text{ref}} \in L^2(\Omega)$ defined in (5.44)



(b) Function g defined in (5.45), which is used to define u^0 by piecewise linear interpolation

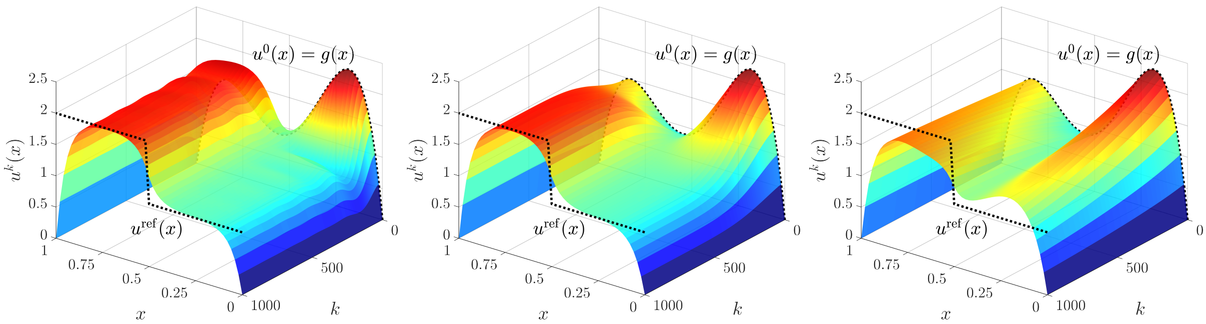
For all iterates, we compute $\varphi(u^k)$ using an interior-point method via the function `fmincon` that is implemented in MATLAB. Additionally, we evaluate the discrete velocity $\|u^{k+1} - u^k\|_{H^1(\Omega)}$. Subfigure 5.6a shows $\varphi(u^k)$ for the different values of $\alpha \in \{3, 10, 50, 100\}$ and (MGD). For small values of α, the merit function values $\varphi(u^k)$ decay faster in the beginning but exhibit more oscillations and converge more slowly later. For bigger values of α the merit function values $\varphi(u^k)$ converge slower in the beginning but overall faster in the considered range. For all values of $\alpha \in \{3, 10, 50, 100\}$, the accelerated method is faster than (MGD). For the discrete velocities $\|u^{k+1} - u^k\|_{H^1(\Omega)}$ plotted in Subfigure 5.6b we can make similar observations. For (MGD) the velocities are monotonically decreasing. For the accelerated methods for all values of $\alpha \in \{3, 10, 50, 100\}$, the velocities initially grow but decay more rapidly later. For $\alpha = 3$ and $\alpha = 5$, we observe significant oscillations and slower decay. Within the considered ranges, the merit function values and velocities decay the fastest for $\alpha = 50$.



(a) Merit function values $\varphi(u^k)$ (b) Discrete velocities $\|u^{k+1} - u^k\|_{H^1(\Omega)}$

Figure 5.6: Comparison of merit function values $\varphi(u^k)$ and discrete velocities $\|u^{k+1} - u^k\|_{H^1(\Omega)}$ for sequences $(u^k)_{k\geq 0}$ given by (MNAG) for different values of $\alpha \in \{3, 10, 50, 100\}$ and for a sequence given by (MGD).

Figure 5.7 shows 1 000 iterates $(u^k)_{k\geq 0}$ of (MNAG) for $\alpha \in \{3, 50\}$ and (MGD), separately plotted in three subfigures. In Subfigures 5.7a and 5.7b the final iterates of (MNAG) appear to be smooth approximations of u^{ref}, as expected from the interpretation of (MOP-H_0^1). In comparison, the final iterate of (MGD), shown in Subfigure 5.7c, does not appear to have converged to a solution. These observations are consistent with those made for the merit function values $\varphi(u^k)$.



(a) $\alpha = 3$ (b) $\alpha = 50$ (c) (MGD)

Figure 5.7: Iterates $(u^k)_{k\geq 0}$ of (MNAG) for different values of $\alpha \in \{3, 50\}$ and (MGD) for $k = 0, \ldots, 1\,000$.

Chapter 6

Conclusion

A descent method for nonconvex locally Lipschitz continuous multiobjective optimization

In Chapter 3, we investigate multiobjective optimization problems with locally Lipschitz continuous objective functions defined on a general Hilbert space. In this context, we present a common descent method that neither requires an a priori discretization of the infinite-dimensional decision space nor relies on scalarization techniques. Avoiding early discretization is crucial, as naive approaches can lead to inconsistency with the topology of the underlying infinite-dimensional problem, resulting in mesh-dependent behavior, such as non-uniform convergence or inconsistent stopping criteria across different discretizations. Similarly, straightforward scalarization strategies can be inadequate in the presence of nonconvex objectives or when the number of objective functions exceeds two.

The proposed method overcomes these difficulties and, to the best of our knowledge, represents the first approach of its kind applicable to such a general setting. It extends a common descent method originally designed for finite-dimensional problems based on subgradient sampling. To facilitate this extension to Hilbert spaces, we introduce a generalized multiobjective ε-subdifferential, which serves as a foundational tool for our algorithmic design. We rigorously analyze its key analytical properties, including a generalized closedness result in the strong $\times$ weak* topology, which is an essential ingredient for establishing convergence toward Pareto critical points in a subsequential sense.

Before formally introducing the common descent algorithm, we show that approximate descent directions satisfying a sufficient decrease condition can be constructed from numerical approximations of the ε-subdifferential. Moreover, we prove that such directions can be obtained in finitely many steps via a subgradient sampling procedure. The descent method is then defined using these directions in combination with an Armijo-backtracking-type step size strategy. The main theoretical results are contained in two theorems: The first establishes that any cluster point of the generated sequence is Pareto critical. The second demonstrates that, under alternative parameter choices, the algorithm attains an approximate criticality condition within finitely many iterations.

Several promising research directions emerge from our current framework. Given the abstract formulation of both the theoretical results and the algorithm, a natural extension would be their application in general Banach spaces. It is particularly important to investigate which Banach spaces allow the multiobjective ε-subdifferential to retain its generalized closedness properties. Such a generalization would broaden the applicability of the method, particularly in more complex settings in optimal control and PDE-constrained optimization. Additionally, there is potential to enhance convergence properties by incorporating higher-order gradient sampling techniques, such as those explored in [111].

Gradient dynamical systems for convex multiobjective optimization

Chapter 4 is dedicated to gradient dynamical systems associated with multiobjective optimization problems with convex and continuously differentiable objective functions with Lipschitz continuous gradients. As a starting point, we revisit the classical steepest descent dynamical system for scalar optimization, followed by a survey of existing gradient dynamical systems in the multiobjective setting.

In Section 4.4, we introduce the inertial multiobjective gradient system (IMOG') and establish the existence of solutions in finite-dimensional Hilbert spaces. We show that solutions to (IMOG') converge weakly to weakly Pareto optimal points.

We improve the system (IMOG') in Section 4.5 and derive the multiobjective gradient system with asymptotically vanishing damping (MAVD). We prove existence of solutions in finite-dimensional spaces and carry out a detailed asymptotic analysis. In particular, we establish improved convergence rates of the merit function values and confirm weak convergence to weak Pareto optima. Theoretical results are supported by a series of numerical experiments.

In Section 4.6, we propose the multiobjective Tikhonov regularized inertial gradient system (MTRIGS). Prior to presenting the system, we discuss the role of Tikhonov regularization in multiobjective optimization and show that an associated generalized regularization path converges strongly to weak Pareto optimal solutions. We prove existence of solutions to (MTRIGS) in the finite-dimensional setting and analyze their asymptotic behavior under varying parameter choices. For a broad class of parameter regimes, we demonstrate strong convergence of trajectories to weak Pareto optima, alongside rapid decay of the merit function values. The section concludes with extensive numerical experiments that validate the theoretical findings and examine the convergence behavior of trajectories in detail.

The results presented in this chapter provide a solid foundation for investigating more advanced dynamical systems in the context of multiobjective optimization. Through the different systems proposed, we demonstrate that the underlying framework is flexible and robust enough to support the development of more advanced gradient dynamical systems. Several research directions naturally arise from this work. One key direction is the extension to constrained multiobjective optimization problems. In particular, integrating primal-dual dynamics, as explored in [48, 124, 125, 241]. Another promising direction is the generalization of these systems to multiobjective min-max optimization problems [46, 64, 121], which are relevant in robust optimization, game theory, and machine learning. Additionally, incorporating time-rescaling techniques [20, 47] and Hessian-driven damping [6, 25] could further improve convergence behavior and dynamic stability. Finally, an exciting direction lies in the study of high-resolution ODE models [212], which have recently been applied in the single-objective setting and may carry over valuable insights into the multiobjective case.

An accelerated gradient method for convex multiobjective optimization

In Chapter 5, we propose an accelerated gradient method for multiobjective optimization problems with convex and continuously differentiable objective functions with Lipschitz continuous gradients. After revisiting Nesterov's accelerated gradient method for scalar optimization, we derive the accelerated multiobjective gradient method (MNAG) via a discretization of the multiobjective gradient system with asymptotic vanishing damping (MAVD). Following the method's definition, we compare (MNAG) to existing first-order methods in multiobjective optimization.

The main results of this chapter are contained in the section on the asymptotic analysis of the method (MNAG). We show that the function values of the iterates converge at a fast rate to an optimal function value, as measured by a suitable merit function for multiobjective optimization. Moreover, we prove weak convergence of the sequence of iterates to weakly Pareto optimal points.

The chapter concludes with numerical experiments demonstrating the practical efficiency of the accelerated multiobjective gradient method (MNAG). First, we test the method on two finite-dimensional problems, confirming that the observed convergence behavior aligns with the theoretical results. Then, we extend the experiments to an infinite-dimensional Hilbert space setting, illustrating the method's applicability beyond finite dimensions. For the infinite-dimensional problem (MNAG) exhibits substantially faster convergence compared to the classical multiobjective steepest descent method.

Several promising directions for future research emerge naturally from this work, reflecting those outlined in Chapter 4. The derivation of the method (MNAG) from the system (MAVD) shows that it is fruitful to further investigate these research directions. The next sensible step is the derivation of fast gradient and proximal point methods with Tikhonov regularization for multiobjective optimization from the system (MTRIGS), similar to [134, 144, 145, 146, 147]. In addition, it would be valuable to explore fast primal-dual algorithms for constrained multiobjective optimization [48, 124, 125, 241], as well as Newton-type methods inspired by fast gradient systems with Hessian-driven damping [6, 25]. Other interesting topics include time-rescaling techniques [20, 47], min-max optimization methods [46, 64, 121], and the investigation of accelerated gradient systems with restart mechanisms, building on preliminary studies such as [29, 156, 185].

References

[1] Absil, P.-A., Mahony, R., and Andrews, B. *Convergence of the iterates of descent methods for analytic cost functions.* In: *SIAM Journal on Optimization* 16 (2) (2005), pp. 531–547. doi: 10.1137/040605266.

[2] Adly, S., Haddad, T., and Le, B. K. *State-dependent implicit sweeping process in the framework of quasistatic evolution quasi-variational inequalities.* In: *Journal of Optimization Theory and Applications* 182 (2019), pp. 473–493. doi: 10.1007/s10957-018-1427-x.

[3] Aliprantis, C. D. and Border, K. C. *Infinite dimensional analysis: A hitchhiker's guide.* Berlin, Heidelberg: Springer, 2006. doi: 10.1007/3-540-29587-9.

[4] Alvarez, F. *On the Minimizing Property of a Second Order Dissipative System in Hilbert Spaces.* In: *SIAM Journal on Control and Optimization* 38 (4) (2000), pp. 1102–1119. doi: 10.1137/S0363012998335802.

[5] Alvarez, F. and Attouch, H. *An Inertial Proximal Method for Maximal Monotone Operators via Discretization of a Nonlinear Oscillator with Damping.* In: *Set-Valued Analysis* 9 (2001), pp. 3–11. doi: 10.1023/A:1011253113155.

[6] Alvarez, F., Attouch, H., Bolte, J., and Redont, P. *A second-order gradient-like dissipative dynamical system with Hessian-driven damping: Application to optimization and mechanics.* In: *Journal de Mathématiques Pures et Appliquées* 81 (8) (2002), pp. 747–779. doi: 10.1016/S0021-7824(01)01253-3.

[7] Antipin, A. S. *Minimization of convex functions on convex sets by means of differential equations.* In: *Differential equations* 30 (9) (1994), pp. 1365–1375.

[8] Armijo, L. *Minimization of functions having Lipschitz continuous first partial derivatives.* In: *Pacific Journal of mathematics* 16 (1) (1966), pp. 1–3. doi: 10.2140/pjm.1966.16.1.

[9] Attouch, H. and Alvarez, F. *The heavy ball with friction dynamical system for convex constrained minimization problems.* In: *Optimization: Proceedings of the 9th Belgian-French-German Conference on Optimization Namur, September 7–11, 1998.* Ed. by Nguyen, V. H., Strodiot, J.-J., and Tossings, P. Springer Berlin Heidelberg, 2000, pp. 25–35. doi: 10.1007/978-3-642-57014-8_2.

[10] Attouch, H., Balhag, A., Chbani, Z., and Riahi, H. *Accelerated Gradient Methods Combining Tikhonov Regularization with Geometric Damping Driven by the Hessian.* In: *Applied Mathematics & Optimization* 88 (29) (2 2023). doi: 10.1007/s00245-023-09997-x.

[11] Attouch, H., Balhag, A., Chbani, Z., and Riahi, H. *Damped inertial dynamics with vanishing Tikhonov regularization: Strong asymptotic convergence towards the minimum norm solution.* In: *Journal of Differential Equations* 311 (2022), pp. 29–58. doi: 10.1016/j.jde.2021.12.005.

[12] Attouch, H., Balhag, A., Chbani, Z., and Riahi, H. *Damped inertial dynamics with vanishing Tikhonov regularization: Strong asymptotic convergence towards the minimum norm solution.* In: *Journal of Differential Equations* 311 (2022), pp. 29–58. doi: 10.1016/j.jde.2021.12.005.

[13] Attouch, H., Chbani, Z., Peypouquet, J., and Redont, P. *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity.* In: *Mathematical Programming* 168 (1) (2018), pp. 123–175. doi: 10.1007/s10107-016-0992-8.

[14] Attouch, H., Chbani, Z., and Riahi, H. *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$.* In: *ESAIM: Control, Optimisation and Calculus of Variations* 25 (2019), p. 2. doi: 10.1051/cocv/2017083.

[15] Attouch, H. and Fadili, J. *From the ravine method to the Nesterov method and vice versa: A dynamical system perspective.* In: *SIAM Journal on Optimization* 32 (3) (2022), pp. 2074–2101. doi: 10.1137/22M1474357.

[16] Attouch, H. and Garrigos, G. *Multiobjective optimization: An inertial dynamical approach to Pareto optima.* In: *arXiv preprint arXiv:1506.02823* (2015). doi: 10.48550/arXiv.1506.02823.

[17] Attouch, H., Garrigos, G., and Goudou, X. *A dynamic gradient approach to Pareto optimization with nonsmooth convex objective functions.* In: *Journal of Mathematical Analysis and Applications* 422 (1) (2015), pp. 741–771. doi: 10.1016/j.jmaa.2014.09.001.

[18] Attouch, H. and Goudou, X. *A Continuous Gradient-like Dynamical Approach to Pareto-Optimization in Hilbert Spaces.* In: *Set-Valued and Variational Analysis* 22 (1) (2014), pp. 189–219. doi: 10.1007/s11228-013-0245-4.

[19] Attouch, H., Goudou, X., and 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.* In: *Communications in Contemporary Mathematics* 2 (01) (2000), pp. 1–34. doi: 10.1142/S0219199700000025.

[20] Attouch, H., Ioan Boţ, R., and Nguyen, D.-K. *Fast convex optimization via time scale and averaging of the steepest descent.* In: *Mathematics of Operations Research* (2024). doi: 10.1287/moor.2023.0186.

[21] Attouch, H. and László, S. *Convex optimization via inertial algorithms with vanishing Tikhonov regularization: Fast convergence to the minimum norm solution.* In: *Mathematical Methods of Operations Research* 99 (2024), pp. 307–347. doi: 10.1007/s00186-024-00867-y.

[22] Attouch, H. and László, S. C. *Continuous Newton-like inertial dynamics for monotone inclusions.* In: *Set-valued and variational analysis* 29 (3) (2021), pp. 555–581. doi: 10.1007/s11228-020-00564-y.

[23] Attouch, H. and Peypouquet, J. *Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators.* In: *Mathematical Programming* 174 (2019), pp. 391–432. doi: 10.1007/s10107-018-1252-x.

[24] Attouch, H. and Peypouquet, J. *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$.* In: *SIAM Journal on Optimization* 26 (2015). doi: 10.1137/15M1046095.

[25] Attouch, H., Peypouquet, J., and Redont, P. *Fast convex optimization via inertial dynamics with Hessian driven damping.* In: *Journal of Differential Equations* 261 (10) (2016), pp. 5734–5783. doi: 10.1016/j.jde.2016.08.020.

[26] Aubin, J.-P. and Cellina, A. *Differential Inclusions: Set-Valued Maps and Viability Theory.* Vol. 264. Berlin: Springer, 2012. doi: 10.1007/978-3-642-69512-4.

[27] Aubin, J.-P. and Frankowska, H. *Set-Valued Analysis.* Boston: Springer, 2009. doi: 10.1007/978-0-8176-4848-0.

[28] Bagirov, A., Karmitsa, N., and Mäkelä, M. *Introduction to Nonsmooth Optimization.* Cham, Heidelberg: Springer, 2014. doi: 10.1007/978-3-319-08114-4.

[29] Bao, C., Chen, L., Li, J., and Shen, Z. *Accelerated gradient methods with gradient restart: Global linear convergence.* In: *arXiv preprint arXiv:2401.07672* (2024). doi: 10.48550/arXiv.2401.07672.

[30] Barzilai, J. and Borwein, J. M. *Two-point step size gradient methods.* In: *IMA journal of numerical analysis* 8 (1) (1988), pp. 141–148. doi: 10.1093/imanum/8.1.141.

[31] Bauschke, H. and Combettes, P. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* 1st ed. Springer, 2011. doi: 10.1007/978-1-4419-9467-7.

[32] Bauschke, H. and Combettes, P. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* 2nd ed. Cham: Springer, 2017. doi: 10.1007/978-3-319-48311-5.

[33] Beck, A. and Teboulle, M. *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems.* In: *SIAM Journal on Imaging Sciences* 2 (1) (2009), pp. 183–202. doi: 10.1137/080716542.

[34] Beer, G. *On Mosco convergence of convex sets.* In: *Bulletin of the Australian Mathematical Society* 38 (2) (1988), pp. 239–253. doi: 10.1017/S0004972700027519.

[35] Beer, G., Rockafellar, R., and Wets, R. J.-B. *A characterization of epi-convergence in terms of convergence of level sets.* In: *Proceedings of the American Mathematical Society* 116 (3) (1992), pp. 753–761. doi: 10.2307/2159443.

[36] Bello-Cruz, J. Y. *A subgradient method for vector optimization problems.* In: *SIAM Journal on Optimization* 23 (4) (2013), pp. 2169–2182. doi: 10.1137/120866415.

[37] Bernreuther, M., Müller, G., and Volkwein, S. *Efficient scalarization in multiobjective optimal control of a nonsmooth PDE.* In: *Computational Optimization and Applications* 83 (2) (2022), pp. 435–464. doi: 10.1007/s10589-022-00390-y.

[38] Bertsekas, D. P. *Nonlinear Programming.* Belmont, MA: Athena Scientific, 1999.

[39] Bieker, K. *Optimization Techniques for Data-Based Control and Machine Learning.* PhD thesis. Paderborn, Germany: Paderborn University, 2023. doi: 10.17619/UNIPB/1-1762.

[40] Bonnans, J. F. and Shapiro, A. *Optimization problems with perturbations: A guided tour.* In: *SIAM review* 40 (2) (1998), pp. 228–264. doi: 10.1137/S0036144596302644.

[41] Bonnel, H., Iusem, A., and Svaiter, B. *Proximal methods in vector optimization.* In: *SIAM Journal on Optimization* 15 (4) (2005), pp. 953–970. doi: 10.1137/s1052623403429093.

[42] Borwein, J. M. and Fitzpatrick, S. *Mosco convergence and the Kadec property.* In: *Proceedings of the American Mathematical Society* 106 (3) (1989), pp. 843–851. doi: 10.1090/S0002-9939-1989-0969313-4.

[43] Boţ, R.-I. and Grad, S.-M. *Inertial forward–backward methods for solving vector optimization problems.* In: *Optimization* 67 (7) (2018), pp. 959–974. doi: 10.1080/02331934.2018.1440553.

[44] Boţ, R. I., Csetnek, E. R., and Nguyen, D.-K. *Fast Augmented Lagrangian Method in the convex regime with convergence guarantees for the iterates.* In: *Mathematical Programming* 200 (2022), pp. 147–197. doi: 10.1007/s10107-022-01879-4.

[45] Boţ, R. I., Csetnek, E. R., and Nguyen, D.-K. *Fast optimistic gradient descent ascent (OGDA) method in continuous and discrete time.* In: *Foundations of Computational Mathematics* (2023), pp. 1–60. doi: 10.1007/s10208-023-09636-5.

[46] Boţ, R. I., Csetnek, E. R., and Sedlmayer, M. *An accelerated minimax algorithm for convex-concave saddle point problems with nonsmooth coupling function.* In: *Computational Optimization and Applications* 86 (3) (2023), pp. 925–966. doi: 10.1007/s10589-022-00378-8.

[47] Boţ, R. I. and Karapetyants, M. A. *A fast continuous time approach with time scaling for nonsmooth convex optimization.* In: *Advances in Continuous and Discrete Models* 2022 (1) (2022). doi: 10.1186/s13662-022-03744-2.

[48] Boţ, R. I. and Nguyen, D.-K. *Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping.* In: *Journal of Differential Equations* 303 (2021), pp. 369–406. doi: 10.1016/j.jde.2021.09.021.

[49] Boţ, R. I. and Sonntag, K. *Inertial dynamics with vanishing Tikhonov regularization for multiobjective optimization.* In: *Journal of Mathematical Analysis and Applications* 554 (2) (2025). doi: 10.1016/j.jmaa.2025.129940.

[50] Brezis, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* New York, NY: Springer, 2011. doi: 10.1007/978-0-387-70914-7.

[51] Brezis, H. *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert.* Amsterdam: North Holland, 1973. doi: 10.1016/S0304-0208(08)72386-7.

[52] Brézis, H. *Asymptotic behavior of some evolution systems.* In: *Nonlinear evolution equations.* Elsevier, 1978, pp. 141–154.

[53] Brézis, H. *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations.* In: *Contributions to nonlinear functional analysis.* Elsevier, 1971, pp. 101–156.

[54] Brown, A. A. and Bartholomew-Biggs, M. C. *Some effective methods for unconstrained optimization based on the solution of systems of ordinary differential equations.* In: *Journal of Optimization Theory and Applications* 62 (2) (1989), pp. 211–224. doi: 10.1007/BF00941054.

[55] Broyden, C. G. *The convergence of a class of double-rank minimization algorithms 1. general considerations.* In: *IMA Journal of Applied Mathematics* 6 (1) (1970), pp. 76–90. doi: 10.1093/imamat/6.1.76.

[56] Bruck Jr, R. E. *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space.* In: *Journal of Functional Analysis* 18 (1) (1975), pp. 15–26. doi: 10.1016/0022-1236(75)90027-0.

[57] Burke, J. V., Lewis, A. S., and Overton, M. L. *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization.* In: *SIAM Journal on Optimization* 15 (2005), pp. 751–779. doi: 10.1137/030601296.

[58] Cabot, A., Engler, H., and Gadat, S. *On the long time behavior of second order differential equations with asymptotically small dissipation.* In: *Transactions of the American Mathematical Society* 361 (11) (2009), pp. 5983–6017. doi: 10.1090/S0002-9947-09-04785-0.

[59] Cabot, A., Engler, H., and Gadat, S. *Second-order differential equations with asymptotically small dissipation and piecewise flat potentials.* In: *Electronic Journal of Differential Equations* 2009 (2009), pp. 33–38.

[60] Cai, Y., Oikonomou, A., and Zheng, W. *Accelerated Algorithms for Constrained Nonconvex-Nonconcave Min-Max Optimization and Comonotone Inclusion.* In: *Proceedings of the 41st International Conference on Machine Learning (ICML).* Vol. 235. Proceedings of Machine Learning Research, 2024, pp. 5312–5347. doi: 10.48550/arXiv.2206.05248.

[61] Campbell, P. J. *The origin of "Zorn's Lemma".* In: *Historia Mathematica* 5 (1) (1978), pp. 77–89. doi: 10.1016/0315-0860(78)90136-2.

[62] Cauchy, A. et al. *Méthode générale pour la résolution des systemes d'équations simultanées.* In: *Comp. Rend. Sci. Paris* 25 (1847) (1847), pp. 536–538.

[63] Chambolle, A. and Dossal, C. H. *On the convergence of the iterates of "FISTA".* In: *Journal of Optimization Theory and Applications* 166 (2015), pp. 968–982. doi: 10.1007/s10957-015-0746-4.

[64] Chambolle, A. and Pock, T. *A first-order primal-dual algorithm for convex problems with applications to imaging.* In: *Journal of mathematical imaging and vision* 40 (2011), pp. 120–145. doi: 10.1007/s10851-010-0251-1.

[65] Chankong, V. and Haimes, Y. Y. *Multiobjective Decision Making: Theory and Methodology.* New York: North-Holland, 2008.

[66] Chen, G., Goh, C., and Yang, X. *The gap function of a convex multicriteria optimization problem.* In: *European Journal of Operational Research* 111 (1) (1998), pp. 142–151. doi: 10.1016/S0377-2217(97)00300-7.

[67] Chen, G.-Y., Goh, C.-J., and Yang, X. Q. *On Gap Functions for Vector Variational Inequalities.* In: *Vector Variational Inequalities and Vector Equilibria.* Ed. by Giannessi, F. Springer, 2000, pp. 55–72. doi: 10.1007/978-1-4613-0299-5_4.

[68] Chen, Z. *Generalized viscosity approximation methods in multiobjective optimization problems.* In: *Computational Optimization and Applications* 49 (1) (2011), pp. 179–192. doi: 10.1007/s10589-009-9282-1.

[69] Chen, Z., Xiang, C., Zhao, K., and Liu, X. *Convergence analysis of Tikhonov-type regularization algorithms for multiobjective optimization problems.* In: *Applied Mathematics and Computation* 211 (1) (2009), pp. 167–172. doi: 10.1016/j.amc.2009.01.037.

[70] Cheney, W. and Goldstein, A. *Proximity maps for convex sets.* In: *Proceedings of the American Mathematical Society* 10 (3) (1959), pp. 448–450. doi: 10.2307/2032864.

[71] Christof, C. and Müller, G. *Multiobjective optimal control of a non-smooth semilinear elliptic partial differential equation.* In: *European Series in Applied and Industrial Mathematics (ESAIM) : Control, Optimisation and Calculus of Variations* (2020). doi: 10.1051/cocv/2020060.

[72] Christof, C. and Wachsmuth, G. *On second-order optimality conditions for optimal control problems governed by the obstacle problem.* In: *Optimization* 70 (10) (2020), pp. 2247–2287. doi: 10.1080/02331934.2020.1778686.

[73] Chuong, T. D. *Tikhonov-type regularization method for efficient solutions in vector optimization.* In: *Journal of Computational and Applied Mathematics* 234 (3) (2010), pp. 761–766. doi: 10.1016/j.cam.2010.01.040.

[74] Chuong, T. D. and Yao, J.-C. *Viscosity-type approximation method for efficient solutions in vector optimization.* In: *Taiwanese Journal of Mathematics* 14 (6) (2010), pp. 2329–2342. doi: 10.11650/twjm/1500406078.

[75] Clarke, F. H. *Optimization and Nonsmooth Analysis.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 1990. doi: 10.1137/1.9781611971309.

[76] Clarkson, J. A. *Uniformly Convex Spaces.* In: *Transactions of the American Mathematical Society* 40 (1936), pp. 396–414. doi: 10.2307/1989630.

[77] Conn, A. R., Gould, N. I., and Toint, P. L. *Trust region methods.* SIAM, 2000.

[78] Cornet, B. *Contributions to the mathematical theory of dynamical mechanisms of resource allocation.* In: *These de doctorat d'état, Université Paris IX Dauphine* (1981).

[79] Cornet, B. *Existence of slow solutions for a class of differential inclusions.* In: *Journal of Mathematical Analysis and Applications* 96 (1) (1983), pp. 130–147. doi: 10.1016/0022-247X(83)90032-X.

[80] Curry, H. B. *The method of steepest descent for non-linear minimization problems.* In: *Quarterly of Applied Mathematics* 2 (3) (1944), pp. 258–261. doi: 10.1090/qam/10667.

[81] Curtis, F. E. and Que, X. *An adaptive gradient sampling algorithm for non-smooth optimization.* In: *Optimization Methods and Software* 28 (6) (2013), pp. 1302–1324. doi: 10.1080/10556788.2012.714781.

[82] Da Cruz Neto, J., Da Silva, G., Ferreira, O. P., and Lopes, J. O. *A subgradient method for multiobjective optimization.* In: *Computational Optimization and Applications* 54 (2013), pp. 461–472. doi: 10.1007/s10589-012-9494-7.

[83] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. *A fast and elitist multiobjective genetic algorithm: NSGA-II.* In: *IEEE Transactions on Evolutionary Computation* 6 (2) (2002), pp. 182–197. doi: 10.1109/4235.996017.

[84] Deb, K. *Multi-Objective Optimization using Evolutionary Algorithms.* Chichester: John Wiley & Sons, 2001.

[85] Deimling, K. *Multivalued Differential Equations.* Vol. 1. Berlin and New York, NY: Walter de Gruyter, 1992. doi: 10.1515/9783110874228.

[86] Dellnitz, M., Schütze, O., and Hestermeyer, T. *Covering Pareto Sets by Multilevel Subdivision Techniques.* In: *Journal of Optimization Theory and Applications* 124 (1) (2005), pp. 113–136. doi: 10.1007/s10957-004-6468-7.

[87] Diestel, J. and Uhl, J. J. *Vector Measures.* Mathematical surveys. Providence, RI: American Mathematical Society, 1977. doi: 10.1090/surv/015.

[88] Dinh, T. L. *Theory of Vector Optimization.* Springer Berlin, Heidelberg, 1989. doi: 10.1007/978-3-642-50280-4.

[89] Dutta, J., Kesarwani, P., and Gupta, S. *Gap functions and error bounds for non-smooth convex vector optimization problem.* In: *Optimization* 66 (11) (2017), pp. 1807–1836.

[90] Ehrgott, M. *Multicriteria Optimization.* Berlin, Heidelberg: Springer, 2005. doi: 10.1007/3-540-27659-9.

[91] Eichfelder, G. *Adaptive Scalarization Methods in Multiobjective Optimization (Vector Optimization).* Berlin: Springer, 2008. doi: 10.1007/978-3-540-79159-1.

[92] Eichfelder, G. *Twenty years of continuous multiobjective optimization in the twenty-first century.* In: *EURO Journal on Computational Optimization* 9 (2021), p. 100014. doi: 10.1016/j.ejco.2021.100014.

[93] Ekeland, I. and Témam, R. *Convex Analysis and Variational Problems.* Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1999. doi: 10.1137/1.9781611971088.

[94] El Moudden, M. and El Mouatasim, A. *Accelerated Diagonal Steepest Descent Method for Unconstrained Multiobjective Optimization.* In: *Journal of Optimization Theory and Applications* 188 (1) (2021), pp. 220–242. doi: 10.1007/s10957-020-01785-9.

[95] Ern, A. and Guermond, J.-L. *Theory and Practice of Finite Elements.* Berlin: Springer, 2004. doi: 10.1007/978-1-4757-4355-5.

[96] Fabian, M., Habala, P., Hájek, P., Montesinos, V., and Zizler, V. *Banach Space Theory: The Basis for Linear and Nonlinear Analysis.* New York, NY: Springer, 2011. doi: 10.1007/978-1-4419-7515-7.

[97] Fiacco, A. V. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* New York: Academic Press, 1983.

[98] Fiacco, A. V. and Ishizuka, Y. *Sensitivity and stability analysis for nonlinear programming.* In: *Annals of Operations Research* 27 (1) (1990), pp. 215–235. doi: 10.1007/BF02055196.

[99] Fletcher, R. and Reeves, C. M. *Function minimization by conjugate gradients.* In: *The Computer Journal* 7 (2) (1964), pp. 149–154. doi: 10.1093/comjnl/7.2.149.

[100] Fletcher, R. *A new approach to variable metric algorithms.* In: *The Computer Journal* 13 (3) (1970), pp. 317–322. doi: 10.1093/comjnl/13.3.317.

[101] Fletcher, R. *Practical methods of optimization.* Chichester: John Wiley & Sons, 2000.

[102] Fliege, J., Drummond, L. M. G., and Svaiter, B. F. *Newton's method for multiobjective optimization.* In: *SIAM Journal on Optimization* 20 (2) (2009), pp. 602–626. doi: 10.1137/08071692X.

[103] Fliege, J. and Svaiter, B. *Steepest descent methods for multicriteria optimization.* In: *Mathematical Methods of Operations Research* 51 (2000), pp. 479–494. doi: 10.1007/s001860000043.

[104] Fliege, J. and Svaiter, B. F. *Steepest descent methods for multicriteria optimization.* In: *Mathematical Methods of Operations Research* 51 (3) (2000), pp. 479–494. doi: 10.1007/s001860000043.

[105] Garrigos, G. *Descent dynamical systems and algorithms for tame optimization, and multi-objective problems.* PhD thesis. Valparaiso, Chili: Université Montpellier; Universidad Técnica Federíco Santa María, 2015.

[106] Gebken, B. *A note on the convergence of deterministic gradient sampling in nonsmooth optimization.* In: *Computational Optimization and Applications* (2024). doi: 10.1007/s10589-024-00552-0.

[107] Gebken, B. and Peitz, S. *An efficient descent method for locally Lipschitz multiobjective optimization problems.* In: *Journal of Optimization Theory and Applications* 188 (2021), pp. 696–723. doi: 10.1007/s10957-020-01803-w.

[108] Gebken, B., Peitz, S., and Dellnitz, M. *A Descent Method for Equality and Inequality Constrained Multiobjective Optimization Problems.* In: *Numerical and Evolutionary Optimization – NEO 2017.* Ed. by Trujillo, L., Schütze, O., Maldonado, Y., and Valle, P. Springer, 2019, pp. 29–61. doi: 10.1007/978-3-319-96104-0_2.

[109] Gebken, B., Peitz, S., and Dellnitz, M. *On the hierarchical structure of Pareto critical sets.* In: *Journal of Global Optimization* 73 (4) (2019), pp. 891–913. doi: 10.1007/s10898-019-00737-6.

[110] Gebken, B. *Computation and analysis of Pareto critical sets in smooth and nonsmooth multiobjective optimization.* PhD thesis. Paderborn, Germany: Paderborn University, 2022. doi: 10.17619/UNIPB/1-1327.

[111] Gebken, B. *Using second-order information in gradient sampling methods for nonsmooth optimization.* In: *arXiv preprint arXiv:2210.04579* (2022). doi: 10.48550/arXiv.2210.04579.

[112] Goldfarb, D. *A family of variable-metric methods derived by variational means.* In: *Mathematics of computation* 24 (109) (1970), pp. 23–26. doi: 10.1090/S0025-5718-1970-0258249-6.

[113] Goldstein, A. *Optimization of Lipschitz continuous functions.* In: *Mathematical Programming* 13 (1977), pp. 14–22. doi: 10.1007/BF01584320.

[114] Goudou, X. and Munier, J. *The gradient and heavy ball with friction dynamical systems: The quasiconvex case.* In: *Mathematical Programming* 116 (2009), pp. 173–191. doi: 10.1007/s10107-007-0109-5.

[115] Grad, S.-M. *A survey on proximal point type algorithms for solving vector optimization problems.* In: *Splitting Algorithms, Modern Operator Theory, and Applications.* Ed. by Bauschke, H. H., Burachik, R. S., and Luke, D. R. Cham: Springer, 2019, pp. 269–308. doi: 10.1007/978-3-030-25939-6_11.

[116] Gräser, C. and Kornhuber, R. *Multigrid methods for obstacle problems.* In: *Journal of Computational Mathematics* 27 (1) (2009), pp. 1–44. doi: 10.4208/jcm.2009-JCM-8559.

[117] Gräser, C., Sack, U., and Sander, O. *Truncated nonsmooth Newton multigrid methods for convex minimization problems.* In: *Lecture Notes in Computational Science and Engineering.* Berlin, Heidelberg: Springer, 2009, pp. 129–136. doi: 10.1007/978-3-642-02677-5_12.

[118] Güler, O. *New proximal point algorithms for convex minimization.* In: *SIAM Journal on Optimization* 2 (4) (1992), pp. 649–664. doi: 10.1137/S1052623491193500.

[119] Güler, O. *On the convergence of the proximal point algorithm for convex minimization.* In: *SIAM journal on control and optimization* 29 (2) (1991), pp. 403–419. doi: 10.1137/0329022.

[120] Halmos, P. R. *Naive Set Theory.* Undergraduate Texts in Mathematics. New York, NY: Springer, 1974. doi: 10.1007/978-1-4757-1645-0.

[121] Hamedani, E. Y. and Aybat, N. S. *A primal-dual algorithm with line search for general convex-concave saddle point problems.* In: *SIAM Journal on Optimization* 31 (2) (2021), pp. 1299–1329. doi: 10.1137/18M1213488.

[122] Haraux, A. *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities.* In: *Journal of the Mathematical Society of Japan* 29 (4) (1977), pp. 615–631. doi: 10.2969/jmsj/02940615.

[123] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning.* New York: Springer, 2009. doi: 10.1007/978-0-387-84858-7.

[124] He, X., Hu, R., and Fang, Y.-P. *Inertial primal-dual dynamics with damping and scaling for linearly constrained convex optimization problems.* In: *Applicable Analysis* 102 (15) (2023), pp. 4114–4139. doi: 10.1080/00036811.2022.2104260.

[125] He, X., Hu, R., and Fang, Y. P. *Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems.* In: *SIAM Journal on Control and Optimization* 59 (5) (2021), pp. 3278–3301. doi: 10.1137/20M1355379.

[126] Helou, E. S., Santos, S., and Simões, L. *On the differentiability check in gradient sampling methods.* In: *Optimization Methods and Software* 31 (5) (2016), pp. 983–1007. doi: 10.1080/10556788.2016.1178262.

[127] Henry, C. *An existence theorem for a class of differential equations with multivalued right-hand side.* In: *Journal of Mathematical Analysis and Applications* 41 (1) (1973), pp. 179–186. doi: 10.1016/0022-247X(73)90192-3.

[128] Hertlein, L. *Inexact bundle methods in Hilbert space with applications to optimal control problems governed by variational inequalities.* PhD thesis. Munich, Germany: Technische Universität München, 2022.

[129] Hertlein, L. and Ulbrich, M. *An inexact bundle algorithm for nonconvex nonsmooth minimization in Hilbert space.* In: *SIAM Journal on Control and Optimization* 57 (5) (2019), pp. 3137–3165. doi: 10.1137/18M1221849.

[130] Hestenes, M. R. and Stiefel, E. *Methods of conjugate gradients for solving linear systems.* In: *Journal of research of the National Bureau of Standards* 49 (6) (1952), pp. 409–436. doi: 10.6028/jres.049.044.

[131] Hillermeier, C. *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach.* Vol. 135. Basel: Birkhäuser, 2012. doi: 10.1007/978-3-0348-8280-4.

[132] Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S. *Optimization with PDE Constraints.* Mathematical Modelling: Theory and Applications. Dordrecht: Springer, 2009. doi: 10.1007/978-1-4020-8839-1.

[133] Jahn, J. *Vector Optimization.* Springer Berlin, Heidelberg, 2011. doi: 10.1007/978-3-642-17005-8.

[134] Karapetyants, M. and László, S. C. *A Nesterov type algorithm with double Tikhonov regularization: fast convergence of the function values and strong convergence to the minimal norm solution.* In: *Applied Mathematics & Optimization* 90 (1) (2024), p. 17. doi: 10.1007/s00245-024-10163-0.

[135] Karush, W. *Minima of functions of several variables with inequalities as side conditions.* PhD thesis. Chicago, IL: University of Chicago, 1939.

[136] Kinderlehrer, D. and Stampacchia, G. *An Introduction to Variational Inequalities and Their Applications.* New York: Academic Press, 1980. doi: 10.1137/1.9780898719451.

[137] Kiwiel, K. *A descent method for non-smooth convex multi-objective minimization.* In: *Large Scale Systems* 8 (2) (1985), pp. 119–129.

[138] Kiwiel, K. *A nonderivative version of the gradient sampling algorithm for nonsmooth nonconvex optimization.* In: *SIAM Journal on Optimization* 20 (2010), pp. 1983–1994. doi: 10.1137/090748408.

[139] Kiwiel, K. *Methods of Descent for Nondifferentiable Optimization.* Berlin Heidelberg: Springer, 1985. doi: 10.1007/BFb0074500.

[140] Kiwiel, K. *Proximity control in bundle methods for convex nondifferentiable minimization.* In: *Mathematical Programming* 46 (1990), pp. 105–122. doi: 10.1007/BF01585731.

[141] Kuhn, H. W. and Tucker, A. W. *Nonlinear Programming.* In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, Calif.: University of California Press, 1951, pp. 481–492.

[142] Kuhn, H. W. and Tucker, A. W. *Nonlinear programming.* In: *Traces and emergence of nonlinear programming.* Ed. by Giorgi, G. and Kjeldsen, T. H. Springer, 2013, pp. 247–258. doi: 10.1007/978-3-0348-0439-4.

[143] Ladas, G. E. and Lakshmikantham, V. *Differential Equations in Abstract Spaces*. Vol. 85. Mathematics in Science and Engineering. New York, London: Academic Press, 1972.

[144] László, S. C. *A proximal-gradient inertial algorithm with Tikhonov regularization: strong convergence to the minimal norm solution*. In: *arXiv preprint arXiv:2407.10350* (2024). doi: 10.48550/arXiv.2407.10350.

[145] László, S. C. *On the convergence of an inertial proximal algorithm with a Tikhonov regularization term*. In: *Communications in Nonlinear Science and Numerical Simulation* (2025), p. 108924. doi: 10.48550/arXiv.2302.02115.

[146] László, S. C. *On the strong convergence of the trajectories of a Tikhonov regularized second order dynamical system with asymptotically vanishing damping*. In: *Journal of Differential Equations* 362 (2023), pp. 355–381. doi: 10.1016/j.jde.2023.03.014.

[147] László, S. C. *Solving convex optimization problems via a second order dynamical system with implicit Hessian damping and Tikhonov regularization*. In: *Computational Optimization and Applications* 90 (1) (2025), pp. 113–149. doi: 10.1007/s10589-024-00620-5.

[148] Lemaréchal, C. *Chapter VII. Nondifferentiable optimization*. In: *Handbooks in Operations Research and Management Science*. Elsevier, 1989, pp. 529–572. doi: 10.1016/s0927-0507(89)01008-x.

[149] Lemaréchal, C. *Cauchy and the gradient method*. In: *Documenta Mathematica, Extra Volume: Optimization Stories* (2012), pp. 251–254. doi: 10.4171/DMS/6/27.

[150] Lin, Z., Li, H., and Fang, C. *Accelerated Optimization for Machine Learning: First-Order Algorithms*. Singapore: Springer, 2020. doi: 10.1007/978-981-15-2910-8.

[151] Lions, J. and Stampacchia, G. *Variational inequalities*. In: *Communications on Pure and Applied Mathematics* 20 (3) (1967), pp. 493–519. doi: 10.1002/cpa.3160200302.

[152] Lions, P.-L. and Mercier, B. *Splitting algorithms for the sum of two nonlinear operators*. In: *SIAM Journal on Numerical Analysis* 16 (6) (1979), pp. 964–979. doi: 10.1137/0716071.

[153] Liu, C. G., Ng, K. F., and Yang, W. H. *Merit functions in vector optimization*. In: *Mathematical Programming* 119 (2) (2009), pp. 215–237. doi: 10.1007/s10107-008-0208-y.

[154] Lucambio Pérez, L. R. and Prudente, L. F. *Nonlinear conjugate gradient methods for vector optimization*. In: *SIAM Journal on Optimization* 28 (3) (2018), pp. 2690–2720. doi: 10.1137/17M1126588.

[155] Luenberger, D. G. *Optimization by Vector Space Methods*. New York: John Wiley & Sons, 1997.

[156] Luo, H., Tang, L., and Yang, X. *An accelerated gradient method with adaptive restart for convex multiobjective optimization problems*. In: *arXiv preprint arXiv:2501.07863* (2025). doi: 10.48550/arXiv.2501.07863.

[157] Mahdavi-Amiri, N. and Yousefpour, R. *An effective nonsmooth optimization algorithm for locally Lipschitz functions*. In: *Journal of Optimization Theory and Applications* 155 (2012), pp. 180–195. doi: 10.1007/s10957-012-0024-7.

[158] Mäkelä, M. *Multiobjective proximal bundle method for nonconvex nonsmooth optimization: Fortran subroutine MPBNGC 2.0*. In: *Reports of the Department of Mathematical Information Technology, Series B. Scientific Computing, B* 13 (2003), p. 2003.

[159] Mäkelä, M., Eronen, V.-P., and Karmitsa, N. *On nonsmooth multiobjective optimality conditions with generalized convexities*. In: *Optimization in Science and Engineering: In Honor of the 60th Birthday of P.M. Pardalos*. Ed. by Rassias, T., Floudas, C., and Butenko, S. New York, NY: Springer, 2014, pp. 333–357. doi: 10.1007/978-1-4939-0808-0_17.

[160] Mäkelä, M., Karmitsa, N., and Wilppu, O. *Multiobjective proximal bundle method for nonsmooth optimization*. In: *TUCS Technical Report No 1120, Turku Centre for Computer Science, Turku* (2014).

[161] Mäkelä, M. and Neittaanmäki, P. *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control*. World Scientific, 1992. doi: 10.1142/1493.

[162] Mäkelä, M. M., Karmitsa, N., and Wilppu, O. *Proximal bundle method for nonsmooth and nonconvex multiobjective optimization*. In: *Mathematical Modeling and Optimization of Complex Structures*. Ed. by Neittaanmäki, P., Repin, S., and Tuovinen, T. Cham: Springer International Publishing, 2016, pp. 191–204. doi: 10.1007/978-3-319-23564-6_12.

[163] Mankau, J. and Schuricht, F. *A nonsmooth nonconvex descent algorithm*. In: *arXiv preprint arXiv:1910.11199* (2019). doi: 10.48550/arXiv.1910.11199.

[164] Mankau, J. and Schuricht, F. *Gradients on sets*. In: *Journal of Convex Analysis* 26 (2019), pp. 1059–1070. doi: 10.4171/JCA/26057.

[165] Martinet, B. *Brève communication. Régularisation d'inéquations variationnelles par approximations successives*. In: *Revue française d'informatique et de recherche opérationnelle. Série rouge* 4 (R3) (1970), pp. 154–158. doi: 10.1051/m2an/197004R301541.

[166] May, R. *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*. In: *Turkish Journal of Mathematics* 41 (2017), pp. 681–685. doi: 10.3906/mat-1512-28.

[167] Meyer, C., Rademacher, A., and Wollner, W. *Adaptive optimal control of the obstacle problem*. In: *SIAM Journal on Scientific Computing* 37 (2) (2015), A918–A945. doi: 10.1137/140975863.

[168] Miettinen, K. and Mäkelä, M. *An interactive method for nonsmooth multiobjective optimization with an application to optimal control*. In: *Optimization Methods and Software* 2 (1) (Jan. 1993), pp. 31–44. doi: 10.1080/10556789308805533.

[169] Miettinen, K. *Nonlinear Multiobjective Optimization*. New York: Springer, 1998. doi: 10.1007/978-1-4615-5563-6.

[170] Miglierina, E. *Slow Solutions of a Differential Inclusion and Vector Optimization*. In: *Set-Valued Analysis* 12 (3) (2004), pp. 345–356. doi: 10.1023/B:SVAN.0000031332.10564.f0.

[171] Mignot, F. *Contrôle dans les inéquations variationelles elliptiques*. In: *Journal of Functional Analysis* 22 (2) (1976), pp. 130–185. doi: 10.1016/0022-1236(76)90017-3.

[172] Montonen, O., Karmitsa, N., and Mäkelä, M. *Multiple subgradient descent bundle method for convex nonsmooth multiobjective optimization.* In: *Optimization* 67 (1) (2018), pp. 139–158. doi: 10.1080/02331934.2017.1387259.

[173] Mordukhovich, B. *Multiobjective optimization problems with equilibrium constraints.* In: *Optimization Methods and Software* 117 (2008), pp. 331–354. doi: 10.1007/s10107-007-0172-y.

[174] Mordukhovich, B. S. *Variational Analysis and Generalized Differentiation I: Basic Theory.* Berlin, Heidelberg: Springer, 2006. doi: 10.1007/3-540-31247-1.

[175] Moré, J. J. and Sorensen, D. C. *Computing a trust region step.* In: *SIAM Journal on scientific and statistical computing* 4 (3) (1983), pp. 553–572. doi: 10.1137/0904038.

[176] Mosco, U. *Convergence of convex sets and of solutions of variational inequalities.* In: *Advances in Mathematics* 3 (4) (1969), pp. 510–585. doi: 10.1016/0001-8708(69)90009-7.

[177] Mukai, H. *Algorithms for Multicriterion Optimization.* In: *IEEE Transactions on Automatic Control* 25 (1980), pp. 177–186. doi: 10.1109/TAC.1980.1102298.

[178] Nemirovskij, A. S. and Yudin, D. B. *Problem complexity and method efficiency in optimization.* Chichester: John Wiley & Sons, 1983.

[179] Nenov, R., Nguyen, D.-K., Balazs, P., and Boţ, R. I. *Accelerated Griffin-Lim algorithm: A fast and provably converging numerical method for phase retrieval.* In: *IEEE Transactions on Signal Processing* 72 (2023), pp. 190–202. doi: 10.1109/TSP.2023.3234567.

[180] Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course.* New York: Springer, 2004. doi: 10.1007/978-1-4419-8853-9.

[181] Nesterov, Y. *Lectures On Convex Optimization.* Cham: Springer, 2018. doi: 10.1007/978-3-319-91578-4.

[182] Nesterov, Y. E. *A Method for Solving the Convex Programming Problem with Convergence Rate $O(1/k^2)$.* In: *Dokl. Akad. Nauk SSSR* 269 (3) (1983), pp. 543–547.

[183] Nocedal, J. *Updating quasi-Newton matrices with limited storage.* In: *Mathematics of computation* 35 (151) (1980), pp. 773–782. doi: 10.1090/S0025-5718-1980-0572855-7.

[184] Nocedal, J. and Wright, S. J. *Numerical optimization.* New York, NY: Springer, 1999. doi: 10.1007/978-0-387-40065-5.

[185] O'donoghue, B. and Candes, E. *Adaptive restart for accelerated gradient schemes.* In: *Foundations of Computational Mathematics* 15 (3) (2015), pp. 715–732. url: https://doi.org/10.1007/s10208-013-9150-3.

[186] Opial, Z. *Weak convergence of the sequence of successive approximations for nonexpansive mappings.* In: *Bulletin of the American Mathematical Society* 73 (4) (1967), pp. 591–597. doi: 10.1090/S0002-9904-1967-11761-0.

[187] Pachpatte, B. *Inequalities for Differential and Integral Equations.* In: ed. by Pachpatte, B. Vol. 197. Mathematics in Science and Engineering. Elsevier, 1998. doi: 10.1016/S0076-5392(98)80002-7. url: https://www.sciencedirect.com/science/article/pii/S0076539298800027.

[188] Palis, J. J. and De Melo, W. *Geometric Theory of Dynamical Systems: An Introduction.* New York, NY: Springer, 2012. doi: 10.1007/978-1-4612-5703-5.

[189] Pareto, V. *Manuale Di Economia Politica.* Piccola biblioteca scientifica. Societa Editrice, 1906.

[190] Peano, G. *Démonstration de l'intégrabilité des équations différentielles ordinaires.* In: *Math. Ann.* 37 (2) (1890), pp. 471–484.

[191] Peano, G. *Sull'integrabilità delle equazioni differenziali di primo ordine.* In: *Atti. Accad. Sci. Torino* 21 (1886), pp. 677–685.

[192] Peitz, S. *Exploiting structure in multiobjective optimization and optimal control.* PhD thesis. Paderborn, Germany: Paderborn University, 2017. doi: 10.17619/UNIPB/1-176.

[193] Penot, J.-P. *Calculus Without Derivatives.* Vol. 266. New York, NY: Springer, 2013. doi: 10.1007/978-1-4614-4538-8.

[194] Perraudin, N., Balazs, P., and Søndergaard, P. L. *A fast Griffin-Lim algorithm.* In: *2013 IEEE workshop on applications of signal processing to audio and acoustics.* IEEE. 2013, pp. 1–4. doi: 10.1109/WASPAA.2013.6701851.

[195] Polak, E. and Ribiere, G. *Note sur la convergence de méthodes de directions conjuguées.* In: *Revue française d'informatique et de recherche opérationnelle. Série rouge* 3 (16) (1969), pp. 35–43.

[196] Polyak, B. T. *Some methods of speeding up the convergence of iteration methods.* In: *USSR Computational Mathematics and Mathematical Physics* 4 (5) (1964), pp. 1–17. doi: 10.1016/0041-5553(64)90137-5.

[197] Polyak, B. T. *Introduction to Optimization.* New York: Optimization Software, 1987.

[198] Polyak, B. T. *Subgradient Methods: A Survey of Soviet Research.* In: *Nonsmooth Optimization.* Ed. by Lemaréchal, C. and Mifflin, R. Vol. 3. Oxford: Pergamon Press, 1978, pp. 5–29. doi: 10.1007/978-3-030-86664-8_4.

[199] Polyak, B. T. *Gradient methods for minimizing functionals.* In: *USSR Computational Mathematics and Mathematical Physics* 3 (4) (1963), pp. 643–653. doi: 10.1016/0041-5553(63)90382-3.

[200] Rauls, A.-T. and Ulbrich, S. *Computation of a Bouligand generalized derivative for the solution operator of the obstacle problem.* In: *SIAM Journal on Control and Optimization* 57 (5) (2019), pp. 3223–3248. doi: 10.1137/18M1187283.

[201] Rauls, A.-T. and Wachsmuth, G. *Generalized derivatives for the solution operator of the obstacle problem.* In: *Set-Valued and Variational Analysis* 28 (2) (2019), pp. 259–285. doi: 10.1007/s11228-019-0506-y.

[202] Reed, M. and Simon, B. *Methods of Modern Mathematical Physics. Volume I: Functional Analysis.* New York: Academic Press, 1980. doi: 10.1016/B978-0-12-585001-8.50001-5.

[203] Rockafellar, R. T. *Monotone operators and the proximal point algorithm.* In: *SIAM journal on control and optimization* 14 (5) (1976), pp. 877–898. doi: 10.1137/0314056.

[204] Rockafellar, R. T. and Wets, R. J.-B. *Variational analysis.* Berlin, Heidelberg: Springer, 2009. doi: 10.1007/978-3-642-02431-3.

[205] ROCKAFELLAR, R. T. *Convex Analysis*. Princeton, NJ: Princeton University Press, 1997. DOI: 10.1515/9781400873173.

[206] RUDIN, W. *Functional Analysis*. McGraw Hill, 1973.

[207] SCHÄFFLER, S., SCHULTZ, R., and WEINZIERL, K. *Stochastic method for the solution of unconstrained vector optimization problems*. In: *Journal of Optimization Theory and Applications* 114 (1) (2002), pp. 209–222. DOI: 10.1023/A:1015472306888.

[208] SCHROPP, J. *Using dynamical systems methods to solve minimization problems*. In: *Applied Numerical Mathematics* 18 (1-3) (1995), pp. 321–335. DOI: 10.1016/0168-9274(95)00065-3.

[209] SEDLMAYER, M., NGUYEN, D.-K., and BOŢ, R. I. *A fast optimistic method for monotone variational inequalities*. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 30406–30438.

[210] SENER, O. and KOLTUN, V. *Multi-Task Learning as Multi-Objective Optimization*. In: *Advances in Neural Information Processing Systems*. Ed. by BENGIO, S., WALLACH, H., LAROCHELLE, H., GRAUMAN, K., CESA-BIANCHI, N., and GARNETT, R. Vol. 31. 2018. URL: https://papers.nips.cc/paper_files/paper/2018/file/432aca3a1e345e339f35a30c8f65edce-Paper.pdf.

[211] SHANNO, D. F. *Conditioning of quasi-Newton methods for function minimization*. In: *Mathematics of computation* 24 (111) (1970), pp. 647–656. DOI: 10.1090/S0025-5718-1970-0274029-X.

[212] SHI, B., DU, S. S., JORDAN, M. I., and SU, W. J. *Understanding the acceleration phenomenon via high-resolution differential equations*. In: *Mathematical Programming* 195 (1) (2022), pp. 79–148. DOI: 10.1007/s10107-021-01681-8.

[213] SHOR, N. Z. *Minimization Methods for Non-Differentiable Function*. Berlin, Heidelberg: Springer, 1985. DOI: 10.1007/978-3-642-82118-9.

[214] SMALE, S. *Global Analysis and Economics I: Pareto Optimum and a Generalization of Morse Theory*. In: *Dynamical Systems*. Ed. by PEIXOTO, M. Elsevier, 1973, pp. 531–544. DOI: 10.1016/B978-0-12-550350-1.50044-8.

[215] SONNTAG, K., GEBKEN, B., MÜLLER, G., PEITZ, S., and VOLKWEIN, S. *A descent method for nonsmooth multiobjective optimization in Hilbert spaces*. In: *Journal of Optimization Theory and Applications* 203 (1) (2024), pp. 455–487. DOI: 10.1007/s10957-024-02520-4.

[216] SONNTAG, K. and PEITZ, S. *Fast convergence of inertial multiobjective gradient-like systems with asymptotic vanishing damping*. In: *SIAM Journal on Optimization* 34 (3) (2024), pp. 2259–2286. DOI: 10.1137/23M1588512.

[217] SONNTAG, K. and PEITZ, S. *Fast Multiobjective Gradient Methods with Nesterov Acceleration via Inertial Gradient-Like Systems*. In: *Journal of Optimization Theory and Applications* 201 (2024), pp. 539–582. DOI: 10.1007/s10957-024-02389-3.

[218] SU, W., BOYD, S., and CANDÈS, E. J. *A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights*. In: *Journal of Machine Learning Research* 17 (153) (2016), pp. 1–43. URL: http://jmlr.org/papers/v17/15-084.html.

[219] Svaiter, B. F. *The multiobjective steepest descent direction is not Lipschitz continuous, but is Hölder continuous.* In: *Operations Research Letters* 46 (4) (2018), pp. 430–433. doi: 10.1016/j.orl.2018.05.008.

[220] Tanabe, H. *Composite Multi-Objective Optimization: Theory and Algorithms.* PhD thesis. Kyoto, Japan: Kyoto University, 2022.

[221] Tanabe, H., Fukuda, E. H., and Yamashita, N. *A globally convergent fast iterative shrinkage-thresholding algorithm with a new momentum factor for single and multi-objective convex optimization.* In: *arXiv preprint arXiv:2205.05262* (2022). doi: 10.48550/arXiv.2205.05262.

[222] Tanabe, H., Fukuda, E. H., and Yamashita, N. *An accelerated proximal gradient method for multiobjective optimization.* In: *Computational Optimization and Applications* (2023), pp. 1–35. doi: 10.1007/s10589-023-00497-w.

[223] Tanabe, H., Fukuda, E. H., and Yamashita, N. *Convergence rates analysis of a multiobjective proximal gradient method.* In: *Optimization Letters* (2022), pp. 1–18. doi: 10.1007/s11590-022-01877-7.

[224] Tanabe, H., Fukuda, E. H., and Yamashita, N. *New merit functions for multiobjective optimization and their properties.* In: *Optimization* (2023), pp. 1–38. doi: 10.1080/02331934.2023.2232794.

[225] Tanabe, H., Fukuda, E. H., and Yamashita, N. *Proximal gradient methods for multiobjective optimization and their applications.* In: *Computational Optimization and Applications* 72 (2019), pp. 339–361. doi: 10.1007/s10589-018-0043-x.

[226] Terazono, Y. and Matani, A. *Continuity of optimal solution functions and their conditions on objective functions.* In: *SIAM Journal on Optimization* 25 (4) (2015), pp. 2050–2060. doi: 10.1137/110850189.

[227] Tibshirani, R. *Regression shrinkage and selection via the lasso.* In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1) (1996), pp. 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x.

[228] Tikhonov, A. N. *Solution of incorrectly formulated problems and the regularization method.* In: *Soviet Mathematics Doklady* 4 (1963), pp. 1035–1038.

[229] Tikhonov, A. N. *Regularization of incorrectly posed problems.* In: *Soviet Mathematics Doklady* 4 (1963), pp. 1624–1627.

[230] Toledano, J. A. *Measures of Noncompactness in Metric Fixed Point Theory.* Birkhauser, Basel, 1997. doi: 10.1007/978-3-0348-8920-9.

[231] Wachsmuth, G. *Strong stationarity for optimal control of the obstacle problem with control constraints.* In: *SIAM Journal on Optimization* 24 (4) (2014), pp. 1914–1932. doi: 10.1137/130925827.

[232] Weiss, A. and Wohlmuth, B. *A posteriori error estimator for obstacle problems.* In: *SIAM Journal on Scientific Computing* 32 (5) (2010), pp. 2627–2658. doi: 10.1137/090773921.

[233] Werner, D. *Funktionalanalysis.* Berlin, Heidelberg: Springer-Verlag, 2011. doi: 10.1007/978-3-642-21017-4.

[234] Whitley, R. *An elementary proof of the Eberlein-Šmulian theorem.* In: *Mathematische Annalen* 172 (1967), pp. 116–118. doi: 10.1007/BF01350091.

[235] Witting, K. *Numerical Algorithms for the Treatment of Parametric Multiobjective Optimization Problems and Applications.* PhD thesis. Paderborn, Germany: Paderborn University, 2012.

[236] Wolfe, P. *Convergence conditions for ascent methods.* In: *SIAM review* 11 (2) (1969), pp. 226–235. doi: 10.1137/1011036.

[237] Yang, X. Q. and Yao, J. C. *Gap Functions and Existence of Solutions to Set-Valued Vector Variational Inequalities.* In: *Journal of Optimization Theory and Applications* 115 (2) (2002), pp. 407–417. doi: 10.1023/A:1020844423345.

[238] Yoon, T. and Ryu, E. K. *Accelerated algorithms for smooth convex-concave minimax problems with O $(1/k^2)$ rate on squared gradient norm.* In: *International Conference on Machine Learning.* PMLR. 2021, pp. 12098–12109.

[239] Young, N. *An introduction to Hilbert space.* Cambridge: Cambridge University Press, 1988.

[240] Zaidman, S. *Functional analysis and differential equations in abstract spaces.* Boca Raton, FL: Chapman Hall/CRC, 1999.

[241] Zeng, X., Lei, J., and Chen, J. *Dynamical primal-dual Nesterov accelerated method and its application to network optimization.* In: *IEEE Transactions on Automatic Control* 68 (3) (2023), pp. 1760–1767. doi: 10.1109/TAC.2022.3152720.