Proximal Splitting Methods in Nonsmooth Convex Optimization

An der Fakultät für Mathematik der Technischen Universität Chemnitz eingereichte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

Vorgelegt von

Dipl.-Math. oec. Christopher Hendrich

Fachbereich für Angewandte Mathematik (Approximationstheorie)

Gefördert durch ein Landesstipendium des Freistaates Sachsen

Gutachter: Prof. Dr. Radu Ioan Boț (Betreuer)
Prof. Dr. Heinz H. Bauschke
Prof. Dr. Gabriele Steidl

Tag der Einreichung: 28. April 2014
Tag der Verteidigung: 17. Juli 2014
Bibliographical description

Christopher Hendrich

Proximal Splitting Methods in Nonsmooth Convex Optimization

Dissertation, 142 pages, Technische Universität Chemnitz, Department of Mathematics, 2014

Report

This thesis is concerned with the development of novel numerical methods for solving nondifferentiable convex optimization problems in real Hilbert spaces and with the investigation of their asymptotic behavior. To this end, we are also making use of monotone operator theory as some of the provided algorithms are originally designed to solve monotone inclusion problems.

After introducing basic notations and preliminary results in convex analysis, we derive two numerical methods based on different smoothing strategies for solving nondifferentiable convex optimization problems. The first approach, known as the double smoothing technique, solves the optimization problem with some given a priori accuracy by applying two regularizations to its conjugate dual problem. A special fast gradient method then solves the regularized dual problem such that an approximate primal solution can be reconstructed from it. The second approach affects the primal optimization problem directly by applying a single regularization to it and is capable of using variable smoothing parameters which lead to a more accurate approximation of the original problem as the iteration counter increases.

We then derive and investigate different primal-dual methods in real Hilbert spaces. In general, one considerable advantage of primal-dual algorithms is that they are providing a complete splitting philosophy in that the resolvents, which arise in the iterative process, are only taken separately from each maximally monotone operator occurring in the problem description. We firstly analyze the forward-backward-forward algorithm of Combettes and Pesquet in terms of its convergence rate for the objective of a nondifferentiable convex optimization problem. Additionally, we propose accelerations of this method under the additional assumption that certain monotone operators occurring in the problem formulation are strongly monotone. Subsequently, we derive two Douglas–Rachford type primal-dual methods for solving monotone inclusion problems involving finite sums of linearly composed parallel sum type monotone operators. To prove their asymptotic convergence, we use a common product Hilbert space strategy by reformulating the corresponding inclusion problem reasonably such that the Douglas–Rachford algorithm can be applied to it. Finally, we propose two primal-dual algorithms relying on forward-backward and forward-backward-forward approaches for solving monotone inclusion problems involving parallel sums of linearly composed monotone operators.

The last part of this thesis deals with different numerical experiments where we intend to compare our methods against algorithms from the literature. The problems which arise in this part are manifold and they reflect the importance of this field of research as convex optimization problems appear in lots of applications of interest.
Keywords

primal-dual algorithm, smoothing technique, monotone inclusions, conjugate duality, convex optimization, nonsmooth optimization, proximal point mapping, resolvent, projection, imaging, location problems, machine learning, portfolio optimization, clustering
Acknowledgment

It is a pleasure for me to thank the people who supported me during my research.

This thesis would not have been possible without the assistance and advice of Prof. Dr. Radu Ioan Boţ. I want to thank him for proposing this topic to me, for his commitment, and for being available for countless discussions. I also want to express my gratitude to Prof. Dr. Gert Wanka for the continuous support during my doctoral study. Many thanks go to our research group, our guests, and my office colleagues for the warmhearted and friendly atmosphere at the department.

I would like to thank Prof. Dr. Heinz H. Bauschke and Prof. Dr. Gabriele Steidl for their contribution in reviewing this thesis and for their valuable remarks and comments which improved the quality of this thesis.

I am grateful to the Free State of Saxony for the financial support of my research by granting me with a graduate fellowship.

I am also grateful to the Department of Mathematics, Technische Universität Chemnitz, for providing me a good research environment.

Finally, I would like to thank my family for love, patience and understanding.
# Contents

1 Introduction
   1.1 A description of the content ................................................. 3
   1.2 Notation and preliminaries ................................................... 5

2 Smoothing techniques in convex optimization  ................................ 13
   2.1 Double Smoothing Technique ................................................. 13
      2.1.1 Problem description ...................................................... 14
      2.1.2 First and second smoothing ............................................. 16
      2.1.3 An appropriate fast gradient method ................................ 21
      2.1.4 Convergence of $\theta(p_k)$ to $\theta(p^*)$ .......................... 22
      2.1.5 Convergence of $\|\nabla_{\rho,\mu}(p_k)\|$ to 0 ....................... 24
      2.1.6 How to construct an approximately primal optimal solution .... 26
      2.1.7 Existence of an optimal solution ..................................... 27
      2.1.8 Improving the convergence rates ..................................... 27
   2.2 Variable Smoothing .............................................................. 32
      2.2.1 Problem description ...................................................... 33
      2.2.2 The smoothing of the problem $(P)$ .................................. 34
      2.2.3 The variable and the constant smoothing algorithm ............... 35
      2.2.4 The case when $f$ is differentiable with Lipschitz continuous gradient .................................................. 41
      2.2.5 The optimization problem with the sum of more than two functions in the objective .................................................. 44

3 Primal-dual algorithms for inclusion problems ................................ 47
   3.1 Convergence analysis of a forward-backward-forward method ........ 47
      3.1.1 Problem description ...................................................... 48
      3.1.2 Convergence estimates for convex minimization problems ....... 50
      3.1.3 Zeros of sums of monotone operators ................................ 56
   3.2 A Douglas–Rachford type primal-dual method ............................ 64
      3.2.1 Problem description ...................................................... 65
      3.2.2 A first primal-dual algorithm ............................................. 66
      3.2.3 A second primal-dual algorithm ....................................... 73
      3.2.4 Application to convex minimization problems ....................... 78
   3.3 Solving inclusions with parallel sums of linearly composed monotone operators .................................................. 82
      3.3.1 Problem description ...................................................... 83
      3.3.2 An algorithm of forward-backward type ............................... 85
      3.3.3 An algorithm of forward-backward-forward type .................... 91
      3.3.4 Application to convex minimization .................................. 94
### 4 Numerical experiments

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Image processing</td>
<td>101</td>
</tr>
<tr>
<td>4.1.1 TV-based image denoising</td>
<td>101</td>
</tr>
<tr>
<td>4.1.2 TV-based image denoising involving higher-order derivatives</td>
<td>104</td>
</tr>
<tr>
<td>4.1.3 TV-based image deblurring</td>
<td>108</td>
</tr>
<tr>
<td>4.1.4 TV-based image inpainting</td>
<td>110</td>
</tr>
<tr>
<td>4.2 Kernel based machine learning</td>
<td>110</td>
</tr>
<tr>
<td>4.3 The generalized Heron problem</td>
<td>113</td>
</tr>
<tr>
<td>4.4 Portfolio optimization under different risk measures</td>
<td>115</td>
</tr>
<tr>
<td>4.5 Clustering</td>
<td>121</td>
</tr>
</tbody>
</table>

### Theses

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theses</td>
<td>124</td>
</tr>
</tbody>
</table>

### Symbols and notation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbols and notation</td>
<td>128</td>
</tr>
</tbody>
</table>

### Bibliography

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliography</td>
<td>131</td>
</tr>
</tbody>
</table>

### Index

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>138</td>
</tr>
</tbody>
</table>

### Curriculum Vitae

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum Vitae</td>
<td>141</td>
</tr>
</tbody>
</table>
# List of Figures

2.1 Illustration of the sum problem appearing in Example 2.11 ........... 41

4.1 An image denoising problem on the lichtenstein test image ........... 103
4.2 An image denoising problem with first- and second-order total variation
functionals on the lichtenstein test image ........................................ 106
4.3 An ISNR value comparison for the problem with first- and second-order
total variation functionals .......................................................... 107
4.4 An image deblurring problem on the cameraman test image ............ 108
4.5 Performance evaluation of an image deblurring problem ............... 109
4.6 An image inpainting problem on the fruits test image .................... 110
4.7 Sample data for the kernel-based classification problem ............... 111
4.8 The generalized Heron problem in two dimensions with numerical results 114
4.9 The generalized Heron problem in three dimensions with numerical results .......................................................... 115
4.10 The clustering problem of two interlocking half moons showing the correct affiliations ....................................................... 121
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Performance evaluation for the image denoising problem</td>
<td>104</td>
</tr>
<tr>
<td>4.2</td>
<td>Misclassification rate in percentage for different model parameters in the classification problem</td>
<td>112</td>
</tr>
<tr>
<td>4.3</td>
<td>Performance evaluation for the classification problem</td>
<td>113</td>
</tr>
<tr>
<td>4.4</td>
<td>Performance evaluation for the portfolio optimization problem</td>
<td>120</td>
</tr>
<tr>
<td>4.5</td>
<td>Performance evaluation for the clustering problem</td>
<td>122</td>
</tr>
</tbody>
</table>
Introduction

In the last couple of years, a special attention within applied mathematics was given on the development of numerical methods for solving structured convex optimization problems having nonsmooth terms in their objectives. This effort is motivated by numerous applications, for example in fields like signal and image processing, portfolio optimization, cluster analysis, and location theory, this enumeration being by far not complete. When characterizing optimality, the convexity allows to make use of powerful results in convex analysis, separation theorems and the (Fenchel) conjugate theory here included. Literature on this topic can be found in [11,22,30,67,82,119,127] and in the seminal work by Rockafellar from 1970 (cf. [106]).

By considering the active and competitive branch of research for solving convex optimization problems numerically, this thesis is concerned with the development of efficient splitting algorithms for approximating optimal solutions of nondifferentiable convex optimization problems, and, more generally, with the solving of monotone inclusions in real Hilbert spaces. To this end, we propose and investigate a number of numerical methods relying on first-order information which either belong to the class of smoothing or primal-dual algorithms.

The principal character of smoothing algorithms is to apply appropriate smoothing techniques as the ones discussed in [99–101] to the problem under investigation, in order to approximate nondifferentiable convex functions by continuously differentiable ones. The resulting problem is then solved via some of the accelerated gradient methods by Nesterov (see [97,98,100]). In fact, we propose and discuss two different smoothing algorithms, one of them being able to reduce the smoothing impact as the iteration counter increases, which, in contrast to the approach involving constant smoothing parameters, leads to a more accurate approximation of the original objective function. Then, instead of determining gradients, the calculation of proximal point mappings of the functions in the objective arises in the numerical scheme. However, it is worth to mention that the proximal point mappings are applied to the functions in the objective separately and that bounded linear operators occuring in the problem formulation (respectively their adjoints) are evaluated explicitly via forward steps. This uncovers an important distinction when compared with the majority of numerical methods relying on augmented Lagrangian (cf. [1,42,68,74,77,81,124–126]) and iterative tresholding (cf. [15,16,19,56,62]) approaches, themselves only featuring
an unsatisfactory splitting behavior.

Primal-dual algorithms, on the other hand, build the second important class of numerical methods that we address in this thesis. These methods are predominantly designed to solve primal-dual systems of monotone inclusion problems where the dual inclusion is formulated in the sense of Attouch–Théra (cf. [7]). They can, however, give rise to the solving of convex optimization problems by taking into account appropriate qualification conditions. In general, primal-dual algorithms are more attractive from a conceptual point of view than those relying on smoothing strategies. This is justified by the property of primal-dual methods to solve systems of first-order optimality conditions, hence they solve the original optimization problem rather than a perturbed version of it. In addition, primal-dual algorithms provide convergence statements for the sequences of iterates whereas smoothing algorithms only guarantee convergence with respect to the function values.

To the oldest and most popular methods for solving monotone inclusion problems belong the proximal point algorithm by Martinet in [92], which was further generalized by Rockafellar in [109], and the Douglas–Rachford splitting algorithm (cf. [64]). The latter was, together with the related Peaceman–Rachford algorithm (cf. [104]), shown to be applicable to evolution equations in [88] and also recovers, in convex optimization, the popular alternating direction method of multipliers (ADMM, cf. [68, 75]) when applied to the conjugate dual problem as is discussed in [76]. In general, the Douglas–Rachford algorithm is designed to find a point in the set of zeros with respect to the sum of two maximally monotone operators as is detailed in Problem 1.1 below.

**Problem 1.1** Let $\mathcal{H}$ be a real Hilbert space and let $A$, $B$ be maximally monotone operators mapping from $\mathcal{H}$ to $2^{\mathcal{H}}$. The problem is to

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx. \quad (1.1)$$

This method, whose roots go back to the year 1956, processes the resolvents of the maximally monotone operators occurring in Problem 1.1 separately in each iteration. The resolvent of a maximally monotone operator corresponds to a proximal point mapping when applied to the subdifferential of a proper, convex, and lower semicontinuous function (itself being a maximally monotone operator, cf. Rockafellar [107]). Demonstrating the importance of operator splitting, we can consider the situation when the sum $A + B$ in Problem 1.1 is maximally monotone. Then, the proximal point algorithm can also be applied to it, but it requires the determination of the resolvent of $A + B$ which may be very hard in general, even in the case when the resolvents of $A$ and $B$ have a simple representation taken separately.

The monotone inclusion in Problem 1.1 corresponds, under certain qualification conditions, to the solving of a convex optimization problem involving the sum of two proper, convex, and lower semicontinuous functions. However, many optimization problems appearing in real-world applications take more sophisticated representations. As this gives rise via convex duality and optimality statements to the solving of monotone inclusions involving mixtures of linear composite, single-valued Lipschitzian, and/or cocoercive and parallel sum type operators, the question is to find easily implementable schemes for these more intricate formulations. Taking this into account, one considerable aim is to avoid asking for the resolvents of sums, parallel
sums, and compositions with bounded linear operators, for which in general no exact formulae exist. On the other hand, one aims to include the single-valued operators via forward evaluations into the iterative process of the algorithm as these steps are in general easier to identify than backward evaluations which require the calculation of resolvents.

As the three fundamental splitting approaches in the shape of Tseng’s forward-backward-forward algorithm (cf. [121]), the forward-backward algorithm (cf. [88, 103]), and the Douglas–Rachford algorithm (cf. [64]) showed to have substantial limitations in this context, first fruitful ideas were developed by Chambolle and Pock in [48] for solving convex optimization problems and by Combettes and Briceño-Arias in [43] for solving monotone inclusions. There, in addition to Problem 1.1, the authors also consider some linear compositions within the problem formulation and derive numerical schemes in the sense of an appropriate splitting. The ideas in [43] were further extended by Combettes and Pesquet in [58] to a monotone inclusion problem involving a single-valued monotone Lipschitzian operator and arbitrary finite sums of linearly composed parallel sum type monotone operators. By reformulating this problem in an appropriate Hilbert product space, the authors reduce the monotone inclusion to the one of finding a zero in the sum of a maximally monotone with a monotone Lipschitzian operator. The latter is then solved via some variant of Tseng’s forward-backward-forward-forward algorithm featuring a tolerance towards errors in the shape of summable sequences, which may arise within the implementation. This method also allows to access the single-valued Lipschitzian operators explicitly via forward steps and it is capable of processing maximally monotone and bounded linear operators separately whenever they arise in the shape of precompositions within the problem description.

This concept was further employed by Võ in [122] in order to solve systems of monotone inclusions with comparable structural complexity. There, instead of assuming monotone Lipschitzian operators, attention is given to monotone cocoercive operators. As a consequence, instead of making use of Tseng’s splitting algorithm, an error sensitive forward-backward method is used, which, however, relies on the use of considerably different technical arguments. If we return to the popular primal-dual method due to Chambolle and Pock described and analyzed in [48, Algorithm 1] and its extension proposed by Condat in [60], it can be shown that these methods are particular instances of Võ’s algorithm.

In addition to the important paper by Combettes in [53], which discusses the solving of monotone inclusions in a general framework via nonexpansive averaged operators, other recently introduced splitting algorithms for solving monotone inclusions can be found in [12, 17, 23, 25, 54, 66, 105, 128] and for the inertial case in [2, 3, 24, 27], while coupled monotone inclusions are analyzed in [5, 6, 29, 44, 55, 123] and convergence rate estimates are provided in [26, 86]. By taking into account constrained convex optimization problems, first fruitful ideas involving easily implementable epigraphical projection approaches are provided in [51, 80].

1.1 A description of the content

In the remainder of this chapter we introduce our notation and preliminary results in convex analysis.
In Chapter 2, we propose two different smoothing algorithms for solving nondifferentiable convex optimization problems. These two approaches rely, on the one hand, on our articles [34, 35] for the so-called double smoothing algorithm including its acceleration strategies, and, on the other hand, on the article [40] for the so-called variable smoothing algorithm. Both algorithms share similarities as they make use of the concept of the Moreau envelope in order to approximate nondifferentiable convex objectives by continuously differentiable ones. Therefore, instead of involving gradient or subdifferential evaluations, these methods rely on the use of the proximity operator. However, the two approaches are completely different to each other in view of the technical assumptions, the smoothing strategies, the fast gradient methods involved, the rates of convergence for the primal objective function, and the reconstruction of approximate primal solutions.

In Chapter 3, we investigate a number of primal-dual algorithms for solving structured monotone inclusion problems in real Hilbert spaces. This technique allows it to process the functions within the objectives separately via their resolvents (resp. proximal point mappings) without being obliged to apply smoothing strategies as in Chapter 2 which may distort the underlying optimization problem by approximating it via some continuously differentiable one. Additionally, primal-dual algorithms are well-suited to solve highly structured monotone inclusion problems as they are able to perform a complete splitting, explicit evaluations of bounded linear operators or their adjoints present in the problem formulation here included. We firstly analyze the forward-backward-forward algorithm of Combettes and Pesquet (cf. [58]), as we did in our article [39], in terms of its convergence rate for the objective of a nondifferentiable convex optimization problem. Subsequently, we propose accelerations of this method under the additional assumption that certain monotone operators occurring in the problem formulation are strongly monotone. The second section in this chapter is concerned with the solving of structured monotone inclusions via some Douglas–Rachford type primal-dual method, a novel approach which was subject to our article in [38]. Finally, the last section in Chapter 3 is devoted to our article in [37] where we derive two different primal-dual algorithms for solving monotone inclusion problems involving parallel sums of linearly composed monotone operators.

In Chapter 4, we analyze the numerical performance of the methods established in Chapter 2 and Chapter 3 by solving a range of nondifferentiable convex optimization problems coming from different areas of optimization. The first numerical experiments are applied to image processing problems, themselves representing a broad class of problems inclosing those in the fields of image denoising, deblurring and inpainting which are the ones discussed in this thesis. Thereafter, by making use of the theory of support vector machines, we solve an image classification problem where handwritten digits are classified by their corresponding number. Additionally, we solve the generalized Heron problem which can be seen as a generalized location problem where one minimizes the sum of distances to closed convex sets subject to a hard constraint on the optimal solution. Based on our article in [36], we also demonstrate the applicability of primal-dual methods to portfolio optimization problems where one minimizes a convex risk functional (being associated with the so-called Optimized Certainty Equivalent) subject to a constraint on the expected return of the portfolio. Finally, our last numerical example is concerned with the solving of a problem which arises in clustering.
1.2 Notation and preliminaries

This section collects the most important notations, definitions and preliminary results which are used throughout this thesis. Appropriate literature on this topic can be found in [11,22,30,82,106,127] and the references therein.

In the following, we are considering real Hilbert spaces \( \mathcal{H} \) which are endowed with the inner product \( \langle \cdot, \cdot \rangle \) and associated norm \( \| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle} \). When for \( i = 1, \ldots, m \), the real Hilbert spaces \( \mathcal{H}_i \) are endowed with inner product \( \langle \cdot, \cdot \rangle_{\mathcal{H}_i} \) and norm \( \| \cdot \|_{\mathcal{H}_i} = \sqrt{\langle \cdot, \cdot \rangle_{\mathcal{H}_i}} \), we denote by

\[
\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m
\]

their Hilbert direct sum. For \( \mathbf{v} = (v_1, \ldots, v_m), \mathbf{q} = (q_1, \ldots, q_m) \in \mathcal{H} \), this real Hilbert space is endowed with inner product and associated norm, respectively defined via

\[
\langle \mathbf{v}, \mathbf{q} \rangle_{\mathcal{H}} = \sum_{i=1}^{m} \langle v_i, q_i \rangle_{\mathcal{H}_i} \quad \text{and} \quad \| \mathbf{v} \|_{\mathcal{H}} = \left( \sum_{i=1}^{m} \| v_i \|_{\mathcal{H}_i}^2 \right)^{1/2}.
\]  \hspace{1cm} (1.2)

Note that \( \mathcal{H} \) can be equivalently written as \( \mathcal{H} = \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \) with scalar product and norm defined via (1.2). By \( \rightharpoonup \) and \( \rightarrow \) we denote weak and strong convergence, respectively. The Cauchy–Schwarz inequality, which follows next, plays a central role in all of mathematics.

**Fact 1.2** (Cauchy–Schwarz inequality) Let \( x \) and \( y \) be in \( \mathcal{H} \). Then it holds

\[
|\langle x, y \rangle| \leq \|x\|\|y\|.
\]  \hspace{1cm} (1.3)

The above inequality is fulfilled as equality if and only if there exists \( \alpha \in \mathbb{R} \) such that \( x = \alpha y \) or \( y = \alpha x \).

By \( \mathbb{N} \), we denote the set of natural numbers \( \{1, 2, \ldots\} \), by \( \mathbb{R} \) the set of real numbers, and by \( \mathbb{R}_+ \) the set of positive real numbers. We shall also use the set of strictly positive real numbers denoted by \( \mathbb{R}_+ = \mathbb{R}_+ \setminus \{0\} \) and the extended real line \( \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm \infty\} \). By \( \mathbb{R}^n \), \( n \in \mathbb{N} \), we denote the \( n \)-dimensional Euclidean space and we let \( \mathbb{1}^n \) be the vector in \( \mathbb{R}^n \) with all entries equal to 1. By \( B_\mathcal{H} \subseteq \mathcal{H} \) and \( B(x, r) \subseteq \mathcal{H} \), we denote the closed unit ball in \( \mathcal{H} \) and the closed ball centered at \( x \in \mathcal{H} \) with radius \( r \in \mathbb{R}_+ \), respectively.

In the sequel we write \( \min \) and \( \max \) instead of \( \inf \) and \( \sup \) when the infimum or supremum is attained, respectively. By \( \arg \min \), we denote the set of minimizers (or optimal solutions) to a scalar optimization problem. For a primal optimization problem \( (P) \), we denote by \( v(P) \) its optimal objective value. For the associated dual optimization problem \( (D) \), the notation \( v(D) \) has a similar meaning. Having a primal-dual pair of optimization problems, then weak duality is fulfilled, if \( v(P) \geq v(D) \) holds. On the other hand, we say that strong duality holds, when \( v(P) = v(D) \) and \( (D) \) has an optimal solution.

Having two nonempty sets \( C \) and \( D \) in \( \mathcal{H} \), their Minkowski sum is defined as \( C + D = \{ c + d : c \in C, d \in D \} \), while for \( \lambda \in \mathbb{R} \), we introduce the scaled
set $\lambda C = \{\lambda c : c \in C\}$, and let $C^\perp = \{x \in \mathcal{H} : \langle x, y \rangle = 0 \ \forall y \in C\}$ be the orthogonal complement of $C$. A nonempty set $K \subseteq \mathcal{H}$ is said to be a cone, if $\lambda K \subseteq K$ for all $\lambda \in \mathbb{R}_+$ while the normal cone of a set $C \subseteq \mathcal{H}$ is defined as $N_C(x) = \{p \in \mathcal{H} : \langle p, y - x \rangle \leq 0 \ \forall y \in C\}$ for arbitrary $x \in C$, and being the empty set for $x \not\in C$. The affine hull of the set $C \subseteq \mathcal{H}$ is nothing else than

$$\text{aff } C = \left\{ \sum_{i=1}^n \lambda_i x_i : n \in \mathbb{N}, x_i \in C, \lambda_i \in \mathbb{R}, i = 1, \ldots, n, \sum_{i=1}^n \lambda_i = 1 \right\}.$$  

We say that $C \subseteq \mathcal{H}$ is convex, if

$$\lambda C + (1 - \lambda)C \subseteq C \ \text{for all } \lambda \in [0, 1].$$

Let $C \subseteq \mathcal{H}$ be convex. By $\text{cl } C$, $\text{int } C$, $\text{ri } C$, and $\text{sqri } C$, we denote the closure, the interior, the relative interior, and the strong quasi-relative interior of the convex set $C$, where the latter is defined as

$$\text{sqri } C = \left\{ x \in C : \text{cone}(C - x) \text{ is a closed linear subspace} \right\}.$$  

In finite dimensional spaces, one has $\text{sqri } C = \text{ri } C$, which is the interior of $C$ relative to its affine hull.

For a linear operator $L : \mathcal{H} \to \mathcal{G}$, the operator $L^* : \mathcal{G} \to \mathcal{H}$ denotes the adjoint operator of $L$ and fulfills the relation $\langle L^* y, x \rangle = \langle y, Lx \rangle$ for all $x \in \mathcal{H}$ and all $y \in \mathcal{G}$. By $\ker L = \{x \in \mathcal{H} : Lx = 0\}$ and $\|L\|$, we denote the kernel of $L$ and its operator norm, respectively. We call the linear operator $L$ bounded (or continuous), if $\|L\| < +\infty$. For a given matrix $A \in \mathbb{R}^{m \times n}$, $m, n \in \mathbb{N}$, the matrix $A^T$ is the transpose of $A$.

**Convex functions**

For $D \subseteq \mathbb{R}$, we say that the function $f : D \to \mathbb{R}$ is increasing on $D$, if for every $a, b \in D$ such that $a > b$, we have $f(a) \geq f(b)$. On the other hand, we call $f$ strictly increasing on $D$, if this inequality is fulfilled as strict inequality, i.e., $f(a) > f(b)$. Consider the function $f : \mathcal{H} \to \mathbb{R}$. By $\text{dom } f := \{x \in \mathcal{H} : f(x) < +\infty\}$, we denote its effective domain and call $f$ proper, if $\text{dom } f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in \mathcal{H}$. The set $\text{epi } f = \{(x, r) \in \mathcal{H} \times \mathbb{R} : f(x) \leq r\}$ is called the epigraph of $f$. In the following, we introduce convex functions.

**Definition 1.3** (Convex function) Let $f : X \to \mathbb{R}$ be a given function and $X$ a nonempty convex set in $\mathcal{H}$. The function $f$ is said to be convex on $X$, if

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

for all $x_1, x_2 \in X$ and for all $\lambda \in (0, 1)$. The function $f$ is called strictly convex on $X$, if the above inequality is fulfilled as a strict inequality for each distinct $x_1$ and $x_2$ in $X$ and for each $\lambda \in (0, 1)$. The function $f : X \to \mathbb{R}$ is called concave (strictly concave) on $X$, if $-f$ is convex (strictly convex) on $X$. An even stronger property is the existence of some $\gamma \in \mathbb{R}_{++}$ such that

$$f(\lambda x_1 + (1 - \lambda)x_2) + \gamma \lambda (1 - \lambda) \|x_1 - x_2\|^2 \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

for each $x_1, x_2 \in X$ and for each $\lambda \in (0, 1)$. In this case, the function $f$ is said to be $\gamma$-strongly convex on $X$. 


By taking into account the conventions \((+\infty) + (-\infty) = +\infty,\ 0 \cdot (+\infty) = +\infty\)
and \(0 \cdot (-\infty) = 0\) (cf. [127, p. 39]), the definition of convexity can be extended to functions \(\hat{f} : \mathcal{H} \to \mathbb{R}\) mapping to the extended real line. We say that \(\hat{f}\) is convex, if for all \(x_1, x_2 \in \mathcal{H}\) and for all \(\lambda \in [0, 1]\)
\[
\hat{f}(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda\hat{f}(x_1) + (1 - \lambda)\hat{f}(x_2).
\]
Therefore, the function \(f : \mathcal{H} \to \mathbb{R}\) is convex on the nonempty, convex set \(X \subseteq \mathcal{H}\), if and only if the function \(\tilde{f} : \mathcal{H} \to \mathbb{R}, \tilde{f}(x) = f(x)\) for \(x \in X\), and \(\tilde{f}(x) = +\infty\), otherwise, is convex. Moreover, an important connection between functions and epigraphs is that the convexity of \(\tilde{f}\) implies the convexity of \(\text{epi} \tilde{f}\) and vice versa, as can be found in [22, 30, 106, 127].

We call \(f\) lower semicontinuous at \(\tau \in \mathcal{H}\), if
\[
\liminf_{y \to \tau} f(y) \geq f(\tau).
\]
Moreover, the function \(f\) is called lower semicontinuous, if this holds for all \(x \in \mathcal{H}\), or, equivalently, if \(\text{epi} f\) is closed. An important class of functions is considered by the set \(\Gamma(\mathcal{H})\), which is defined as
\[
\Gamma(\mathcal{H}) := \{f : \mathcal{H} \to \mathbb{R} : f\text{ is proper, convex, and lower semicontinuous}\}. \tag{1.4}
\]
Properness, convexity, and lower semicontinuity are our standard assumptions for the functions occurring in the optimization problems within this thesis.

The following definition of conjugate functions, which goes back to Fenchel (cf. [69, 70]), plays a central role in convex analysis.

**Definition 1.4** (Conjugate function) Let \(f : \mathcal{H} \to \mathbb{R}\) be a given function. Then the (Fenchel) conjugate function of \(f\) is \(f^* : \mathcal{H} \to \mathbb{R}\),
\[
f^*(p) = \sup_{x \in \mathcal{H}} \{\langle p, x \rangle - f(x)\} \forall p \in \mathcal{H}. \tag{1.5}
\]

If conjugation is applied to \(f^*\), we obtain \(f^{**} := (f^*)^*\) and call it the biconjugate of \(f\). The conjugate function \(f^*\) of some arbitrary (not necessarily convex or lower semicontinuous) function \(f : \mathcal{H} \to \mathbb{R}\) is always convex and lower semicontinuous and it holds the so-called Young–Fenchel inequality, i.e.,
\[
f(x) + f^*(p) \geq \langle p, x \rangle \forall x, p \in \mathcal{H}.
\]

One of the most important theorems in conjugate duality is the Fenchel–Moreau Theorem.

**Theorem 1.5** (Fenchel–Moreau) Let \(f \in \Gamma(\mathcal{H})\). Then the conjugate function \(f^*\) is proper, convex, and lower semicontinuous and \(f^{**} = f\).

Another fundamental instrument in convex analysis is the (convex) subdifferential.

**Definition 1.6** (Convex subdifferential) Let \(f : \mathcal{H} \to \mathbb{R}\) and \(\tau \in \mathcal{H}\) be such that \(f(\tau) \in \mathbb{R}\). An element \(p \in \mathcal{H}\) is called a subgradient of the function \(f\) at \(\tau\), if
\[
f(x) - f(\tau) \geq \langle p, x - \tau \rangle \forall x \in \mathcal{H}. \tag{1.6}
\]
The set of all subgradients of the function \( f \) at \( \bar{x} \) is denoted by \( \partial f(\bar{x}) \) and is called the (convex) subdifferential of \( f \) at \( \bar{x} \). The subdifferential of \( f \) at \( \bar{x} \) is considered to be empty, if \( f(\bar{x}) \in \{\pm \infty\} \). We say that \( f \) is subdifferentiable at \( \bar{x} \), if \( \partial f(\bar{x}) \neq \emptyset \).

By taking into account convex functions, the subdifferential replaces the gradient in a more general sense by gathering similar properties. Therefore, the importance of the subdifferential is pointed out by the fact that functions in this work are generally considered to be convex and nondifferentiable. One also has the equivalent characterization

\[
p \in \partial f(\bar{x}) \iff f(\bar{x}) + f^*(p) = \langle p, \bar{x} \rangle,
\]

i.e., for \( p, \bar{x} \in H \) fulfilling \( p \in \partial f(\bar{x}) \), the Young–Fenchel inequality is fulfilled as equality and vice versa.

A fundamental result in view of convex optimization is the following theorem which is known as Fermat’s rule.

**Theorem 1.7** (Fermat’s rule) Let \( f : H \to \bar{R} \) be proper. Then

\[
\text{arg min } f = \{x \in H : 0 \in \partial f(x)\}.
\]

The next definition concerns the notation of infimal convolutions.

**Definition 1.8** (Infimal convolution) Having two proper functions \( f, g : H \to \mathbb{R} \), their infimal convolution is defined by

\[
(f \boxplus g)(x) = \inf_{y \in H} \{f(y) + g(x - y)\} \quad \forall x \in H. \tag{1.7}
\]

The infimal convolution in Definition 1.8 is a convex function when \( f \) and \( g \) are convex, while, for \( f, g \in \Gamma(H) \), the function \( f \boxplus g \) is not necessarily in \( \Gamma(H) \). However, whenever the condition (cf. \([11, 22, 45, 85]\))

\[
0 \in \text{sqrt}(\text{dom } f^* - \text{dom } g^*) \tag{1.8}
\]

is fulfilled, one has \( f \boxplus g \in \Gamma(H) \) and the infimal convolution is exact, i.e., the infimum in (1.7) is attained for all \( x \in H \). For other conditions guaranteeing (1.8), we refer the reader to \([11, \text{Proposition 15.7}]\).

Next, we introduce the notation of the Moreau envelope, which plays a central role in the development of our smoothing algorithms in Chapter 2.

**Definition 1.9** (Moreau envelope) The Moreau envelope of parameter \( \gamma \in \mathbb{R}_{++} \) of \( f \in \Gamma(H) \) is the function \( \gamma f : H \to \mathbb{R} \) defined as

\[
\gamma f(x) := f \left( \frac{1}{2\gamma} \| \cdot \|^2 \right)(x) = \inf_{y \in H} \left\{ f(y) + \frac{1}{2\gamma} \| x - y \|^2 \right\} \quad \forall x \in H. \tag{1.9}
\]

On the other hand, the proximal point mapping, which is closely connected with the Moreau envelope, plays a key role in smoothing and primal-dual algorithms.
**Definition 1.10** (Proximal point) Let $f \in \Gamma(H)$, $\gamma \in \mathbb{R}_{++}$, and $x \in H$. We denote by $\text{Prox}_{\gamma f}(x)$ the proximal point (also called proximal point mapping) of $\gamma f$ at $x$, representing the unique optimal solution of the minimization problem in (1.9), i.e.,

$$\text{Prox}_{\gamma f}(x) = \arg \min_{y \in H} \left\{ \gamma f(y) + \frac{1}{2} \| y - x \|^2 \right\}.$$  \hspace{1cm} (1.10)

Notice that $\text{Prox}_{\gamma f} : H \rightarrow H$ is single-valued and firmly nonexpansive (cf. [11, Proposition 12.27]), i.e., it holds

$$\| \text{Prox}_{\gamma f}(x) - \text{Prox}_{\gamma f}(y) \|^2 + \| (x - \text{Prox}_{\gamma f}(x)) - (y - \text{Prox}_{\gamma f}(y)) \|^2 \leq \| x - y \|^2$$  \hspace{1cm} (1.11)

for all $x, y \in H$. Therefore, the proximal point mapping is also Lipschitz continuous with Lipschitz constant equal to 1. For a large class of functions arising in different fields of applications, the proximal point mappings are given by exact formulae (cf. [57, 59]). We additionally have (cf. [11, Theorem 14.3 (i)])

$$\gamma f(x) + \frac{1}{\gamma} f^*(\frac{x}{\gamma}) = \frac{\| x \|^2}{2\gamma} \forall x \in H,$$  \hspace{1cm} (1.12)

and the extended Moreau’s decomposition formula (cf. [11, Theorem 14.3 (ii)])

$$\text{Prox}_{\gamma f}(x) + \gamma \text{Prox}_{\frac{1}{\gamma} f^*}(\frac{x}{\gamma}) = x \forall x \in H.$$  \hspace{1cm} (1.13)

The function $\gamma f$ is (Fréchet) differentiable on $H$ and its gradient $\nabla(\gamma f) : H \rightarrow H$ fulfills (cf. [11, Proposition 12.29])

$$\nabla(\gamma f)(x) = \frac{1}{\gamma} (x - \text{Prox}_{\gamma f}(x)) \forall x \in H,$$  \hspace{1cm} (1.14)

being in the light of (1.11) $\gamma^{-1}$-Lipschitz continuous.

The indicator function of the set $C \subseteq H$ is the function

$$\delta_C : H \rightarrow \mathbb{R}, \quad \delta_C(x) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{otherwise} \end{cases}.$$  \hspace{1cm} (1.15)

For a nonempty, convex, and closed set $C \subseteq H$ and $\gamma \in \mathbb{R}_{++}$, we have $\delta_C \in \Gamma(H)$, and, furthermore, $\text{Prox}_{\gamma \delta_C} = \mathcal{P}_C$, where

$$\mathcal{P}_C : H \rightarrow C, \quad \mathcal{P}_C(x) = \arg \min_{z \in C} \| x - z \|^2$$

denotes the projection operator onto $C$.

When $f : H \rightarrow \mathbb{R}$ is convex and differentiable with $L_{\gamma f}$-Lipschitz continuous gradient, then for all $x, y \in H$, it holds (see, for instance, [11, 97, 98])

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L_{\gamma f}}{2} \| y - x \|^2.$$  \hspace{1cm} (1.16)

**Single- and set-valued operators**

Let $T : H \rightarrow H$ be a single-valued operator. By $\text{Fix} T = \{ x \in H : x = Tx \}$, we denote the set of fixed points of $T$. We call $T$ a $\beta$-Lipschitz continuous operator with $\beta \in \mathbb{R}_{++}$, if $\| Tx - Ty \| \leq \beta \| x - y \|$ for all $x, y \in H$. 

**Definition 1.11** (Nonexpansiveness) Let \( D \subseteq \mathcal{H} \) be a nonempty subset of \( \mathcal{H} \) and \( T : D \to \mathcal{H} \). We call \( T \)

(i) *nonexpansive*, if it is 1-Lipschitz continuous, i.e.,
\[
\|Tx - Ty\| \leq \|x - y\| \quad \forall x, y \in D,
\]

(ii) *firmly nonexpansive*, if
\[
\|Tx - Ty\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2 \quad \forall x, y \in D.
\]

The following definition of cocoercive operators plays a central role in the analysis of primal-dual algorithms having forward-backward characteristics.

**Definition 1.12** (cocoercive operator) Let \( D \subseteq \mathcal{H} \) be a nonempty subset of \( \mathcal{H} \), let \( T : D \to \mathcal{H} \), and let \( \beta \in \mathbb{R}_{++} \). Then \( T \) is called \( \beta \)-cocoercive (or \( \beta \)-inverse strongly monotone), if \( \beta T \) is firmly nonexpansive, that is
\[
\langle x - y, Tx - Ty \rangle \geq \beta \|Tx - Ty\|^2 \quad \text{for all } x, y \in D.
\]

Now, let \( M : \mathcal{H} \to 2^\mathcal{H} \) be a set-valued operator. We denote by \( \text{dom} M = \{x \in \mathcal{H} : Mx \neq \emptyset\} \) its domain, by \( \text{zer} M = \{x \in \mathcal{H} : 0 \in Mx\} \) its set of zeros, by \( \text{ran} M = \{u \in \mathcal{H} : \exists x \in \mathcal{H}, u \in Mx\} \) its range, by \( \text{gra} M = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Mx\} \) its graph, and by \( M^{-1} : \mathcal{H} \to 2^\mathcal{H} \), \( u \mapsto \{x \in \mathcal{H} : u \in Mx\} \) its inverse.

**Definition 1.13** (Monotone operator) Let \( M : \mathcal{H} \to 2^\mathcal{H} \) be set-valued. We call \( M \)

(i) *monotone*, if
\[
\langle x - y, u - v \rangle \geq 0 \quad \text{for all } (x, u), (y, v) \in \text{gra} M,
\]

(ii) *maximally monotone*, if there exists no monotone operator \( M' : \mathcal{H} \to 2^\mathcal{H} \) such that \( \text{gra} M' \) properly contains \( \text{gra} M \),

(iii) *uniformly monotone* with modulus \( \phi_M : \mathbb{R}_+ \to [0, +\infty] \), if \( \phi_M \) is increasing, vanishes only at 0, and
\[
\langle x - y, u - v \rangle \geq \phi_M (\|x - y\|) \quad \text{for all } (x, u), (y, v) \in \text{gra} M,
\]

(iv) \( \beta \)-*strongly monotone* with \( \beta \in \mathbb{R}_{++} \), if it is uniformly monotone with modulus \( \phi_M : \mathbb{R}_+ \to [0, +\infty] \), \( \phi_M(t) = \beta t^2 \), i.e.,
\[
\langle x - y, u - v \rangle \geq \beta \|x - y\|^2 \quad \text{for all } (x, u), (y, v) \in \text{gra} M.
\]

The *resolvent* and the *reflected resolvent* of an operator \( M : \mathcal{H} \to 2^\mathcal{H} \) are, respectively,
\[
J_M = (\text{Id} + M)^{-1} \quad \text{and} \quad R_M = 2J_M - \text{Id},
\]
the operator \( \text{Id} : \mathcal{H} \to \mathcal{H} \) denoting the *identity* on the underlying Hilbert space. When \( M \) is maximally monotone, its resolvent (and, consequently, its reflected resolvent) is a single-valued operator being defined everywhere on \( \mathcal{H} \), which is a classical result due to Minty (cf. [93]). Furthermore, in this configuration, the resolvent is firmly nonexpansive, and, by [11, Proposition 23.18], we have for \( \gamma \in \mathbb{R}_{++} \)
\[
\text{Id} = J_{\gamma M} + \gamma J_{\gamma^{-1} M^{-1}} \circ \gamma^{-1} \text{Id}.
\]
(1.17)
The following result (cf. [11, Proposition. 23.16]) plays a key role in Chapter 3.
Proposition 1.14 For each $i = 1, \ldots, m$, let $\mathcal{H}_i$ be a real Hilbert space, let $A_i : \mathcal{H}_i \to 2^{\mathcal{H}_i}$ be a maximally monotone operator, and let $\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m$. We set $A : \mathcal{H} \to 2^\mathcal{H}, A = \bigtimes_{i=1}^m A_i$. Then $A$ is maximally monotone and
\[
J_A = J_{A_1} \times J_{A_2} \times \ldots \times J_{A_m}.
\] (1.18)

Moreover, for $f \in \Gamma(\mathcal{H})$ and $\gamma \in \mathbb{R}_{++}$, the subdifferential $\partial(\gamma f)$ is a maximally monotone set-valued operator (cf. [107]). By taking this into account, the important connection between resolvent and proximal point mapping reads
\[
J_{\gamma \partial f} = (\text{Id} + \gamma \partial f)^{-1} = \text{Prox}_{\gamma f}.
\] (1.19)

The sum and the parallel sum of two set-valued operators $M_1, M_2 : \mathcal{H} \to 2^\mathcal{H}$ are defined as $M_1 + M_2 : \mathcal{H} \to 2^\mathcal{H}, (M_1 + M_2)(x) = M_1(x) + M_2(x)$, for all $x \in \mathcal{H}$, and
\[
M_1 \circ M_2 : \mathcal{H} \to 2^\mathcal{H}, M_1 \circ M_2 = \left(M_1^{-1} + M_2^{-1}\right)^{-1},
\]
respectively. If $M_1$ and $M_2$ are monotone, then $M_1 + M_2$ and $M_1 \circ M_2$ are monotone as well. However, if $M_1$ and $M_2$ are maximally monotone, this property is in general neither for $M_1 + M_2$ nor for $M_1 \circ M_2$ true (see [22, 108]). The maximality can, however, be guaranteed, if appropriate qualification conditions are fulfilled, which can be found in the above cited literature.

Probability spaces and risk measures

Now (cf. [36, Sec. 1.2]) we let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where the elements $\omega$ of $\Omega$ represent future states, or individual scenarios (and are allowed to be only finitely many), $\mathcal{F}$ is a $\sigma$-algebra on measurable subsets of $\Omega$, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. For a measurable random variable $X : \Omega \to \mathbb{R} \cup \{+\infty\}$, the expectation value with respect to $\mathbb{P}$ is defined by $E[X] := \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega)$. Whenever $X$ takes the value $+\infty$ on a subset of positive measure, we have $E[X] = +\infty$. Equalities between random variables are to be interpreted in an almost surely (a.s.) way. Random variables $X : \Omega \to \mathbb{R} \cup \{+\infty\}$, which take a constant value $\lambda \in \mathbb{R}$, i.e., $X = \lambda$ a.s., will be identified with the real number $\lambda$. Similarly, inequalities of the form $X \geq \lambda$, $X \leq \lambda$, $X \leq Y$, etc., are to be viewed in the sense of holding almost surely. By $F_X$, we denote the distribution function of $X$, i.e., $F_X(\lambda) = \mathbb{P}(X \leq \lambda)$. By taking this into account, essential supremum and essential infimum of a random variable $X$ are, respectively,
\[
\text{esssup}(X) = \inf \{a \in \mathbb{R} : \mathbb{P}(X > a) = 0\} = \inf \{a \in \mathbb{R} : X \leq a\},
\]
\[
\text{essinf}(X) = -\text{esssup}(-X) = \sup \{a \in \mathbb{R} : X \geq a\}.
\]
Each random variable $X$ can be represented as $X = X_+ - X_-$, where $X_+, X_-$ are random variables defined via $X_+(\omega) = \max\{X(\omega), 0\}$ and $X_-(\omega) = \max\{-X(\omega), 0\}$ for all $\omega \in \Omega$.

Consider further the real Hilbert space $L^2 := L^2(\Omega, \mathcal{F}, \mathbb{P})$, that is the space
\[
L^2 = \left\{X : \Omega \to \mathbb{R} \cup \{+\infty\} : X \text{ is measurable, } \int_{\Omega} |X(\omega)|^2 \, d\mathbb{P}(\omega) < +\infty\right\},
\]
which is endowed with *inner product* and *norm* defined for arbitrary \(X, Y \in L^2\) via
\[
\langle X, Y \rangle = \int_\Omega X(\omega)Y(\omega) \, dP(\omega) \quad \text{and} \quad \|X\| = \left(\int_\Omega (X(\omega))^2 \, dP(\omega)\right)^{\frac{1}{2}},
\]
respectively.

**Definition 1.15** (Risk functions) A proper function \(\rho : L^2 \to \mathbb{R}\) is called *risk function.* The risk function \(\rho\) is said to be

(i) *convex,* if \(\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda)\rho(Y)\) for all \(\lambda \in (0, 1), X, Y \in L^2\);

(ii) *positively homogeneous,* if \(\rho(0) = 0\) and \(\rho(\lambda X) = \lambda \rho(X)\) for all \(\lambda \in \mathbb{R}^+, X \in L^2\);

(iii) *monotone,* if \(X \geq Y\) implies \(\rho(X) \leq \rho(Y)\) for all \(X, Y \in L^2\);

(iv) *cash-invariant,* if \(\rho(X + c) = \rho(X) - c\) for all \(c \in \mathbb{R}, X \in L^2\);

(v) a *convex risk measure,* if \(\rho\) is convex, monotone and cash-invariant,

(vi) a *coherent risk measure,* if \(\rho\) is a positively homogeneous convex risk measure.

Axioms for coherent risk measures were first given in the literature by Artzner, Delbaen, Eber, and Heath in [44], while later, Föllmer and Schied considered in [73] convex risk measures by replacing the sublinearity with the weaker assumption of convexity. Since the value \(\rho(X)\) can be understood as a capital requirement for the future net worth \(X\), a convex risk measure guarantees that the capital requirement of the convex combination of two positions does not exceed the convex combination of the capital requirements of the positions taken separately. For properties and examples of coherent and convex risk measures, we refer to [4, 14, 31, 71, 72, 89, 110–113].

In Section 4.4, a central role will be played by a generalized convex risk measure associated to the so-called Optimized Certainty Equivalent, which was introduced for concave utility functions in [13] and adapted to convex utility functions in [31]. For the utility functions considered here, we make the following assumption.

**Assumption 1.16** (Convex utility function) Let \(u : \mathbb{R} \to \mathbb{R}\) be a proper, convex, lower semicontinuous, and nonincreasing function such that \(u(0) = 0\) and \(-1 \in \partial u(0)\).

In the literature, the two conditions imposed on \(u\) are known as the *normalization conditions* and are equivalent to \(u(0) = 0\) and \(u(t) \geq -t\) for all \(t \in \mathbb{R}\). The generalized convex risk measure we use in order to quantify the risk was given under the name Optimized Certainty Equivalent (OCE) in [14] and is defined as (see, also, [31])
\[
\rho_u : L^2 \to \mathbb{R} \cup \{+\infty\}, \quad \rho_u(X) = \inf_{\lambda \in \mathbb{R}} \{\lambda + \mathbb{E}[u(X + \lambda)]\}. \tag{1.20}
\]

By Assumption 1.16, it follows that \(\rho_u(X) \geq -\mathbb{E}[X]\) for every \(X \in L^2\) and that \(\rho_u\) fulfills the requirements of being a convex risk measure.
Smooth techniques in convex optimization

In this chapter we investigate two different approaches for solving nondifferentiable convex optimization problems by making use of appropriate smoothing techniques. Here we approximate nonsmooth functions occurring in the objectives by their Moreau envelopes which are known to be (Fréchet) differentiable, and thereafter apply accelerated gradient methods to solve the resulting smooth problem. A possible drawback of this approach when using constant smoothing parameters as in the double smoothing approach below is that one solves an approximated optimization problem but not the original one and therefore has to choose the smoothing parameters appropriately small from the beginning. However, within the second section of this chapter we are going to propose a smoothing algorithm where the smoothing parameters are variable and tend to zero as the iteration counter increases, which lets us solve the original optimization problem.

2.1 Double Smoothing Technique

In this section, whose content relies on the two articles [34, 35], we are interested in solving a specific class of unconstrained convex optimization problems in finite dimensional spaces. In convex optimization, separation theorems and the (Fenchel) conjugate theory (see [11, 106, 127]) are the ingredients for assigning a dual optimization problem via the perturbation approach to a primal one. Under the premise that strong duality holds, solving the dual optimization problem instead is a natural approach to obtain an optimal solution to the primal optimization problem, too. Be aware that weak duality is always fulfilled for these problems. However, in order to guarantee strong duality, so-called regularity conditions are needed (see, for example, [22, 30, 127]).

If one intends to solve an unconstrained, convex, and differentiable minimization problem, then there are already plenty of promising methods available (such as the steepest descent method, fast gradient methods, or, in an appropriate setting, Newton’s method, cf. [98]) which can be applied. However, the situation changes
dramatically when the considered objective function proves to be nondifferentiable as is the case in lots of real-world applications. Then, methods which involve gradient evaluations are not anymore well-defined and one has to find appropriate solution strategies. In view of this conflict, the convex subdifferential is used instead, not exclusively as a tool for theoretically characterizing optimality, but also as the counterpart of the gradient in different numerical methods. However, the convergence properties of classical subgradient methods which solve unconstrained, convex, and nondifferentiable minimization problems are known to be slow (cf. [98]).

Our aim in this section is to develop an efficient algorithm for solving an unconstrained convex optimization problem having as objective the sum of two proper, convex, and lower semicontinuous functions, one of them being composed with a linear operator. To this end we are not relying on subgradient type methods, since their complexity can not be better than $O\left(\frac{1}{\varepsilon^2}\right)$ iterations, where $\varepsilon > 0$ is the desired accuracy for the primal objective value (see [98]). Instead of this, we solve the associated Fenchel dual problem efficiently and show that it is possible to reconstruct from it an approximately optimal solution to the primal one. To this end, we make use of a double smoothing technique, in fact a generalization of the double smoothing approach employed by Devolder, Glineur and Nesterov in [63] for a special class of convex constrained optimization problems. By employing the structure of the dual problem, this technique assumes the regularization of its objective function into a differentiable strongly convex one with Lipschitz continuous gradient. The doubly regularized dual problem is then solved via some special fast gradient method from [98], its numerical scheme giving rise to a sequence of primal variables for solving the primal optimization problem with $\varepsilon$-accuracy in $O\left(\frac{1}{\varepsilon} \ln \left(\frac{1}{\varepsilon}\right)\right)$ iterations.

The double smoothing algorithm provided in this section performs two matrix-vector multiplications in each iteration and requires the determination of the proximal point mappings of the functions occurring in the primal objective. Furthermore we manage to avoid expensive linear operator inversions. This aspect represents an important distinction when compared to the majority of the splitting algorithms, exceptions in this sense being modern primal-dual algorithms (see, for instance, [38, 43, 48, 58, 122]), where some of them are discussed in Chapter 3. In general the determination of the proximal point mappings can be seen as restrictive. Though, for a large class of functions arising in different applications, exact formulae for these are available (cf. [57, 59]).

2.1.1 Problem description

In this section we are dealing with optimization problems of the type

\[
(P) \quad \inf_{x \in \mathcal{H}} \{ f(x) + g(Kx) \},
\]

where $\mathcal{H}$ is a real Hilbert space, $f \in \Gamma(\mathcal{H})$ and $g \in \Gamma(\mathbb{R}^m)$, and $K : \mathcal{H} \rightarrow \mathbb{R}^m$ is a linear operator fulfilling $K(\text{dom } f) \cap \text{dom } g \neq \emptyset$. Furthermore, we assume that $\text{dom } f$ and $\text{dom } g$ are bounded.

Remark 2.1 The assumption that $\text{dom } f$ and $\text{dom } g$ are bounded can be weakened by only considering the boundedness of $\text{dom } f$. In this situation, in the formulation of $(P)$, the function $g$ can be replaced by $g + \delta_{\text{cl}(\text{dom } f)}$, which is a proper, convex,
and lower semicontinuous function with bounded effective domain. The drawback of this approach is that the proximal point mapping of $g + \delta_{\text{cl}(\text{dom } f)}$ may be hard to implement. For another iterative method designed for optimization problems which assumes the minimization of a convex function over a convex and bounded feasible set, and relying on the use of proximal point mappings, we refer the reader to [96].

Taking into account that our method is a generalization of the double smoothing approach in [63], one should also notice that the counterparts of the assumptions considered there, in our setting would ask for closedness regarding the effective domains of the functions $f$ and $g$, too. However, we will be able to employ the double smoothing technique for $(P)$ without being obliged to impose this assumption.

According to [22, 30], the Fenchel dual problem to $(P)$ is nothing else than

\[
(D) \quad \sup_{p \in \mathbb{R}^m} \{-f^*(K^*p) - g^*(-p)\},
\]

where $f^*: \mathcal{H} \to \mathbb{R}$ and $g^*: \mathbb{R}^m \to \mathbb{R}$ denote the conjugate functions of $f$ and $g$, respectively. The conjugate functions of $f$ and $g$ can then be written as

\[
f^*(q) = \sup_{x \in \text{dom } f} \{\langle q, x \rangle - f(x)\} = -\inf_{x \in \text{dom } f} \{-\langle q, x \rangle + f(x)\} \quad \forall q \in \mathcal{H},
\]

and

\[
g^*(p) = \sup_{x \in \text{dom } g} \{\langle p, x \rangle - g(x)\} = -\inf_{x \in \text{dom } g} \{-\langle p, x \rangle + g(x)\} \quad \forall p \in \mathbb{R}^m,
\]

respectively. According to [11, Theorem 11.9], the optimization problems arising in the formulation of both $f^*(q)$ for all $q \in \mathcal{H}$ and $g^*(p)$ for all $p \in \mathbb{R}^m$ are solvable, some fact which implies that $\text{dom } f^* = \mathcal{H}$ and $\text{dom } g^* = \mathbb{R}^m$.

By writing the dual problem $(D)$ equivalently as the infimum optimization problem

\[
\inf_{p \in \mathbb{R}^m} \{f^*(K^*p) + g^*(-p)\},
\]

one can easily see that the Fenchel dual problem of the latter is

\[
\sup_{x \in \mathcal{H}} \{-f^{**}(x) - g^{**}(Kx)\},
\]

which, by the Fenchel–Moreau Theorem, is nothing else than

\[
\sup_{x \in \mathbb{R}^n} \{-f(x) - g(Kx)\}.
\]

In order to guarantee strong duality for this primal-dual pair, it is sufficient to ensure that (see, for instance, [22]) $0 \in \text{ri}(K^*(\text{dom } g^*) + \text{dom } f^*)$. As $f^*$ has full domain, this regularity condition is automatically fulfilled, which means that $v(D) = v(P)$, and the primal optimization problem $(P)$ has an optimal solution. Due to the fact that $f$ and $g$ are proper and $A(\text{dom } f) \cap \text{dom } g \neq \emptyset$, this further implies $v(D) = v(P) \in \mathbb{R}$.

Later we will assume that the dual problem $(D)$ has an optimal solution, too, and that an upper bound of its norm is known.

In the following we denote by $\theta: \mathbb{R}^m \to \mathbb{R}$, $\theta(p) = f^*(K^*p) + g^*(-p)$, the objective function of the dual problem $(D)$. Hence, the latter can be equivalently written as

\[
(D) \quad -\inf_{p \in \mathbb{R}^m} \theta(p).
\]
Since in general we can neither guarantee the smoothness of \( p \mapsto f^*(K*p) \) nor of \( p \mapsto g^*(-p) \), the dual problem \((D)\) is a nondifferentiable convex optimization problem. Our goal is to solve this problem efficiently and to obtain from here an optimal solution to \((P)\). To this end, we are not relying on subgradient type schemes, due to their nonsatisfactory complexity being \( O \left( \frac{1}{\varepsilon} \right) \), but, instead, we are applying some smoothing techniques introduced in [99–101]. More precisely, we firstly regularize the functions \( p \mapsto f^*(K*p) \) and \( p \mapsto g^*(-p) \), by taking into account the definitions of the two conjugates, in order to obtain a smooth approximation of the objective of \((2.3)\) having a Lipschitz continuous gradient. Then we solve the regularized dual problem by making use of a fast gradient method for smooth and strongly convex functions given in [98] for solving the regularized dual, which implicitly will solve both the dual problem \((D)\) and the primal problem \((P)\) approximately in \( O \left( \frac{1}{\varepsilon} \ln \left( \frac{1}{\varepsilon} \right) \right) \) iterations. More than that, we will show that this rate of convergence can be improved when strengthening the assumptions imposed on \( f \) and \( g \).

### 2.1.2 First and second smoothing

**First smoothing**

For a positive real number \( \rho \in \mathbb{R}_{++} \), the function \( p \mapsto f^*(K*p) = \sup_{x \in \mathcal{H}} \{ \langle K*p, x \rangle - f(x) \} \) can be approximated by

\[
f^*_\rho(K*p) = \sup_{x \in \mathcal{H}} \left\{ \langle K*p, x \rangle - f(x) - \frac{\rho}{2} \|x\|^2 \right\},
\]

while, given \( \mu \in \mathbb{R}_{++} \), the function \( p \mapsto g^*(-p) = \sup_{x \in \mathbb{R}^n} \{ \langle -p, x \rangle - g(x) \} \) can be approximated by

\[
g^*_\mu(-p) = \sup_{x \in \mathbb{R}^n} \left\{ \langle -p, x \rangle - g(x) - \frac{\mu}{2} \|x\|^2 \right\}.
\]

For each \( p \in \mathbb{R}^m \), the maximization problems which occur in the formulations of \( f^*_\rho(K*p) \) and \( g^*_\mu(-p) \) have unique solutions (see, for instance, [20, Lemma 2.33]), since their objectives are proper, strongly concave and upper semicontinuous functions.

In order to determine the gradient of the functions \( p \mapsto f^*_\rho(K*p) \) and \( p \mapsto g^*_\mu(-p) \), we are going to make use of the Moreau envelope of the functions \( f \) and \( g \), respectively. Indeed, for all \( p \in \mathbb{R}^m \), we have

\[
-f^*_\rho(K*p) = -\sup_{x \in \mathcal{H}} \left\{ \langle K*p, x \rangle - f(x) - \frac{\rho}{2} \|x\|^2 \right\}
= \inf_{x \in \mathcal{H}} \left\{ -\langle K*p, x \rangle + f(x) + \frac{\rho}{2} \|x\|^2 \right\}
= \inf_{x \in \mathcal{H}} \left\{ f(x) + \frac{\rho}{2} \left\| \frac{K*p}{\rho} - x \right\|^2 \right\} - \frac{\|K*p\|^2}{2\rho} = \frac{1}{\rho} f\left( \frac{K*p}{\rho} \right) - \frac{\|K*p\|^2}{2\rho}.
\]
As the Moreau envelope is continuously differentiable (see Section 1.2), \( p \mapsto -f^*_\rho(K^*p) \) is continuously differentiable, as well, and it holds for all \( p \in \mathbb{R}^m \)

\[
-\nabla (f^*_\rho \circ K^*)(p) = \frac{K}{\rho} \nabla \frac{1}{2} f \left( \frac{K^*p}{\rho} \right) - \frac{KK^*p}{\rho} = \frac{K}{\rho} \left( \frac{K^*p}{\rho} - x_{\rho,p} \right) - \frac{KK^*p}{\rho} = -Kx_{\rho,p},
\]

which means that

\[\nabla (f^*_\rho \circ K^*)(p) = Kx_{\rho,p},\]

where

\[x_{\rho,p} = \text{Prox} \frac{1}{\rho} f \left( \frac{K^*p}{\rho} \right) .\]

By taking into account the nonexpansiveness of the proximal point mapping as discussed in (1.11), for \( p, q \in \mathbb{R}^m \), it holds

\[
\left\| \nabla (f^*_\rho \circ K^*)(p) - \nabla (f^*_\rho \circ K^*)(q) \right\| = \|Kx_{\rho,p} - Kx_{\rho,q}\| \leq \|K\| \|x_{\rho,p} - x_{\rho,q}\|
\]

\[
\leq \|K\| \left\| \frac{K^*p}{\rho} - \frac{K^*q}{\rho} \right\| \leq \|K\|^2 \rho \|p - q\|,
\]

thus \( p \mapsto \nabla (f^*_\rho \circ K^*)(p) \) is \( \frac{\|K\|^2}{\rho} \)-Lipschitz continuous.

For the function \( p \mapsto g^*(\rho, -p) \), one can proceed analogously. For all \( p \in \mathbb{R}^m \), one has

\[
-g^*_\rho (-p) = \inf_{x \in \mathbb{R}^m} \left\{ g(x) + \frac{\rho}{2} \| \frac{p}{\rho} - x \|^2 \right\} \leq \frac{\|p\|^2}{2\rho} = \frac{\rho^2}{2\rho} = \frac{1}{2\rho} g \left( \frac{p}{\rho} \right) - \frac{\|p\|^2}{2\rho},
\]

which is a continuously differentiable function such that

\[
-\nabla g^*_\rho (-\cdot)(p) = -\frac{1}{\mu} \nabla \frac{1}{2\rho} g \left( \frac{p}{\mu} \right) - \frac{p}{\mu} = -\frac{1}{\mu} \left( \mu \left( -\frac{p}{\mu} - x_{\mu,p} \right) \right) - \frac{p}{\mu} = x_{\mu,p},
\]

and therefore,

\[\nabla g^*_\rho (-\cdot)(p) = -x_{\mu,p},\]

where

\[x_{\mu,p} = \text{Prox} \frac{1}{\rho} g \left( \frac{p}{\mu} \right) .\]

For \( p, q \in \mathbb{R}^m \), it holds

\[
\left\| \nabla g^*_\rho (-\cdot)(p) - \nabla g^*_\rho (-\cdot)(q) \right\| = \|x_{\rho,p} + x_{\mu,q}\| \leq \left\| \frac{p}{\mu} + \frac{q}{\mu} \right\| \leq \frac{1}{\mu} \|p + q\|,
\]

so that \( p \mapsto \nabla g^*_\rho (-\cdot)(p) \) is \( \frac{1}{\rho} \)-Lipschitz continuous.

**Remark 2.2** If \( f \) is \( \rho \)-strongly convex for \( \rho \in \mathbb{R}^{++} \), then there is no need to apply the first regularization for \( p \mapsto f^*(K^*p) \), as this function is already Fréchet differentiable with \( \frac{\|K\|^2}{\rho} \)-Lipschitz continuous gradient. Indeed, the \( \rho \)-strong convexity
of $f$ implies that $f^*$ is Fréchet differentiable with $\frac{1}{\rho}$-Lipschitz continuous gradient (see [11, Theorem 18.15]). Hence, for all $p,q \in \mathbb{R}^m$, we have
\[
\|\nabla(f^* \circ K^*)(p) - \nabla(f^* \circ K^*)(q)\| = \|K\nabla f^*(K^*p) - K\nabla f^*(K^*q)\| \\
\leq \frac{\|K\|}{\rho} \|K^*p - K^*q\| \leq \frac{\|K\|^2}{\rho} \|p - q\| .
\]
Taking
\[
x_{f,p} := \nabla f^*(K^*p),
\]
one has that $0 \in \partial(f - \langle K^*p, \cdot \rangle)(x_{f,p})$, which means that $x_{f,p}$ is the unique optimal solution (see [20, Lemma 2.33]) of the optimization problem
\[
\inf_{x \in \mathcal{H}} \{f(x) - \langle K^*p, x \rangle\}.
\]
The same applies for $p \mapsto g^*(-p)$. If $g$ is $\mu$-strongly convex with parameter $\mu \in \mathbb{R}_{++}$, the function $g^*(-\cdot)$ is known to be differentiable with $\frac{1}{\mu}$-Lipschitz continuous gradient.

The constants $D_f := \sup \{\|x\|^2_2 : x \in \text{dom} \ f\}$ and $D_g := \sup \{\|x\|^2_2 : x \in \text{dom} \ g\}$ will play an important role in the upcoming convergence schemes. Since $\text{dom} \ f$ and $\text{dom} \ g$ are bounded, $D_f$ and $D_g$ are real numbers.

**Proposition 2.3** For arbitrary $p \in \mathbb{R}^m$, it holds
\[
f^*_\rho(K^*p) \leq f^*(K^*p) \leq f^*_\rho(K^*p) + \rho D_f \quad \text{and} \quad g^*_\mu(-p) \leq g^*(-p) \leq g^*_\mu(-p) + \mu D_g.
\]
Proof. For $p \in \mathbb{R}^m$, one has
\[
f^*_\rho(K^*p) = \langle K^*p, x_{p,p} \rangle - f(x_{p,p}) - \frac{\rho}{2} \|x_{p,p}\|^2 \leq \langle K^*p, x_{p,p} \rangle - f(x_{p,p}) \leq f^*(K^*p) \\
\leq \sup_{x \in \text{dom} \ f} \left\{ \langle K^*p, x \rangle - f(x) - \frac{\rho}{2} \|x\|^2 \right\} + \sup_{x \in \text{dom} \ f} \left\{ \frac{\rho}{2} \|x\|^2 \right\} \\
= f^*_\rho(K^*p) + \rho D_f.
\]
The other estimates follow similarly. \hfill \blacksquare

For $\rho \in \mathbb{R}_{++}$ and $\mu \in \mathbb{R}_{++}$, let $\theta_{\rho,\mu} : \mathbb{R}^m \to \mathbb{R}$ be defined by $\theta_{\rho,\mu}(p) = f^*_\rho(K^*p) + g^*_\mu(-p)$. The function $\theta_{\rho,\mu}$ is differentiable with $L(\rho, \mu)$-Lipschitz continuous gradient
\[
\nabla \theta_{\rho,\mu}(p) = \nabla(f^*_\rho \circ K^*)(p) + \nabla g^*_\mu(-\cdot)(p) = Kx_{p,p} - x_{\mu,p},
\]
where $L(\rho, \mu) := \frac{\|K\|^2}{\rho} + \frac{1}{\mu}$.

Summing up the inequalities from Proposition 2.3, we get
\[
\theta_{\rho,\mu}(p) \leq \theta(p) \leq \theta_{\rho,\mu}(p) + \rho D_f + \mu D_g \quad \forall p \in \mathbb{R}^m. \tag{2.6}
\]
In the following, we let $p^* \in \mathbb{R}^m$ be an optimal solution to $(D)$. Further, for $p \in \mathbb{R}^m$, we have
\[
\theta_{\rho,\mu}(p) = f^*_\rho(K^*p) + g^*_\mu(-p) \\
= \langle p, Kx_{p,p} \rangle - f(x_{p,p}) - \frac{\rho}{2} \|x_{p,p}\|^2 - \langle p, x_{\mu,p} \rangle - g(x_{\mu,p}) - \frac{\mu}{2} \|x_{\mu,p}\|^2 ,
\]
and from here
\[ f(x_\rho,\mu) + g(x_\mu,\nu) + \theta(p^*) = \langle p, \nabla\theta_\rho,\mu(p) \rangle + (\theta(p^*) - \theta_\rho,\mu(p)) - \frac{\rho}{2} \|x_\rho,\mu\|^2 - \frac{\mu}{2} \|x_\mu,\nu\|^2. \]

This provides the estimate
\[ |f(x_\rho,\mu) + g(x_\mu,\nu) + \theta(p^*)| \leq |\langle p, \nabla\theta_\rho,\mu(p) \rangle| + |\theta_\rho,\mu(p) - \theta(p^*)| + \rho D_f + \mu D_g. \quad (2.7) \]

Since \(-\theta(p^*) = v(D) \leq v(P)\) (weak duality) and \(|\theta_\rho,\mu(p) + v(D)| \leq |\theta(p) + v(D)| + \rho D_f + \mu D_g\), we conclude that
\[ f(x_\rho,\mu) + g(x_\mu,\nu) - v(P) \leq |\langle p, \nabla\theta_\rho,\mu(p) \rangle| + |\theta(p) - \theta(p^*)| + 2\rho D_f + 2\mu D_g. \quad (2.8) \]

Following the ideas in [63], we further consider for the regularized optimization problem
\[ \inf_{p\in\mathbb{R}^m} \theta_\rho,\mu(p) \]
the following fast gradient scheme (see [100, scheme (3.11)):

**Init.:** Choose \(w_0 \in \mathbb{R}^m\) and set \(k := 0\).

For \(k \geq 0\):
1. Compute \(\theta_\rho,\mu(w_k)\) and \(\nabla\theta_\rho,\mu(w_k)\),
2. Find \(p_k = \arg\min_{w\in\mathbb{R}^m} \left\{ \langle \nabla\theta_\rho,\mu(w_k), w - w_k \rangle + \frac{L(\rho, \mu)}{2} \|w - w_k\|^2 \right\}, \)
3. Find \(z_k = \arg\min_{w\in\mathbb{R}^m} \left\{ L(\rho, \mu) \|w_0 - w\|^2 + \sum_{i=0}^{k+1} \frac{i+1}{2} |\theta_\rho,\mu(w_i) + \langle \nabla\theta_\rho,\mu(w_i), w - w_i \rangle| \right\}. \)
4. Set \(w_{k+1} := \frac{2}{k+3} z_k + \frac{k+1}{k+3} p_k. \)

Assuming that \(p^*_\mu \in \mathbb{R}^m\) is an optimal solution of (2.9), it follows that \(\nabla\theta_\rho,\mu(p^*_\mu) = 0\). Thus, due to the properties of the above convergence scheme provided in [100], we have
\[ \theta_\rho,\mu(p_k) - \theta_\rho,\mu(p^*_\mu) \leq \frac{4L(\rho, \mu) \|p_0 - p^*_\mu\|^2}{(k + 1)(k + 2)} \quad \forall k \geq 0. \quad (2.10) \]

From (2.6), we get \(\theta_\rho,\mu(p_k) \geq \theta(p_k) - \rho D_f - \mu D_g\) for all \(k \geq 0\) and \(\theta_\rho,\mu(p^*_\mu) \leq \theta(p^*)\). Hence, we obtain
\[ \theta_\rho,\mu(p_k) - \theta_\rho,\mu(p^*_\mu) \geq \theta(p_k) - \rho D_f - \mu D_g - \theta(p^*), \]
which further implies that
\[ \theta(p_k) - \theta(p^*) \leq \theta_\rho,\mu(p_k) - \theta_\rho,\mu(p^*_\mu) + \rho D_f + \mu D_g \leq \frac{4L(\rho, \mu) \|p_0 - p^*_\mu\|^2}{(k + 1)(k + 2)} + \rho D_f + \mu D_g \]
for all \(k \geq 0\). Now, in order to guarantee \(\theta(p_k) - \theta(p^*) \leq \varepsilon\), namely that \(p_k\) is a solution of the dual problem (\(D\)) with \(\varepsilon\)-accuracy, we can force all three terms in the above inequality to be less than or equal to \(\frac{\varepsilon}{3}\). By taking
\[ \rho := \rho(\varepsilon) = \frac{\varepsilon}{3D_f} \quad \text{and} \quad \mu := \mu(\varepsilon) = \frac{\varepsilon}{3D_g}, \]
this means that the amount of iterations \( k \) needed, in order to satisfy \( \varepsilon \)-optimality for the dual iterate, depends on the relation
\[
\frac{4L(\rho, \mu) \| p_0 - p_S^* \|^2}{(k + 1)(k + 2)} \leq \frac{\varepsilon}{3}.
\]
Since the Lipschitz constant \( L(\rho, \mu) = \frac{\| K \|^2}{\rho} + \frac{1}{\mu} \) is of order \( \frac{1}{\varepsilon} \), the rate of convergence for \( \theta(p_k) - \theta(p^*) \leq \varepsilon \) is \( O\left(\frac{1}{\varepsilon}\right) \).

Further, according to (2.8), in order to gain an accuracy for the primal optimization problem proportional to \( \varepsilon > 0 \), one only has to ensure that \( |\langle p_k, \nabla \theta_{\rho, \mu}(p_k) \rangle| \) is lower than or equal to \( O(\varepsilon) \). However, by [98, Theorem 2.1.5], we have
\[
\| \nabla \theta_{\rho, \mu}(p_k) \|^2 \leq \frac{2L(\rho, \mu) \| p_0 - p_S^* \|^2}{\sqrt{(k + 1)(k + 2)}} \quad \forall \ k \geq 0.
\]
This means that the norm of the gradient \( \nabla \theta_{\rho, \mu}(p_k) \) decreases with an order being \( O\left(\frac{1}{\varepsilon^2}\right) \). As a result, in order to achieve an accuracy for the primal optimization problem which is proportional to \( \varepsilon \) via the estimation (2.8), we need \( k = O\left(\frac{1}{\varepsilon^2}\right) \) iterations. This convergence, however, is generally seen to be slow and it is not better than the rate of convergence of the subgradient approach.

From another point of view, in order to get a feasible solution to the primal optimization problem \((P)\), it is necessary to investigate the distance between \( Kx_{\rho,p_k} \) and \( x_{\mu,p_k} \), since the functions \( f \) and \( g \circ K \) have to share the same argument (which would be \( x_{\rho,p_k} \), if \( \| \nabla \theta_{\rho,\mu}(p_k) \| = \| Kx_{\rho,p_k} - x_{\mu,p_k} \| = 0 \)). Therefore, the norm of the gradient \( \| \nabla \theta_{\rho,\mu}(p_k) \| \) is an indicator for an approximately feasible primal solution and necessarily has to tend to zero as well in an appropriate amount of time.

**Second smoothing**

In the following, a second regularization is applied to \( \theta_{\rho,\mu} \), as done in [63], in order to make it strongly convex, some fact which will allow us to use a fast gradient scheme with better convergence properties for the decrease of \( \| \nabla \theta_{\rho,\mu} \| \). Therefore, adding the strongly convex function \( \frac{\kappa}{2} \| \cdot \|^2 \) to \( \theta_{\rho,\mu} \) for some positive real number \( \kappa \) gives rise to the following regularization of the objective function
\[
\theta_{\rho,\mu,\kappa} : \mathbb{R}^m \to \mathbb{R}, \quad \theta_{\rho,\mu,\kappa}(p) := \theta_{\rho,\mu}(p) + \frac{\kappa}{2} \| p \|^2 = f^*_p(K^*p) + g^*_\mu(-p) + \frac{\kappa}{2} \| p \|^2,
\]
which is \( \kappa \)-strongly convex for \( \kappa \in \mathbb{R}^+ \) (cf. [82, Proposition B.1.1.2]). We further deal with the optimization problem
\[
\inf_{p \in \mathbb{R}^m} \theta_{\rho,\mu,\kappa}(p). \quad (2.11)
\]
By taking into account [18, Proposition A.8 and Proposition B.10], the optimization problem (2.11) has a unique minimizer. The function \( \theta_{\rho,\mu,\kappa} \) is differentiable and for
all $p \in \mathbb{R}^m$, it holds
\[
\nabla \theta_{p,\mu,\kappa}(p) = \nabla \left( \theta_{p,\mu}(\cdot) + \frac{\kappa}{2} \|\cdot\|^2 \right)(p) = K x_{p,\mu} - x_{\mu,\kappa} + \kappa p
\]
\[
= K \text{Prox}_{\frac{\kappa}{\rho}} \left( \frac{K^* \mu}{\rho} \right) - \text{Prox}_{\frac{\kappa}{\mu}} \left( - \frac{\rho}{\mu} \right) + \kappa p.
\]
This gradient is $L(\rho, \mu, \kappa)$-Lipschitz continuous with $L(\rho, \mu, \kappa) := \frac{\|K\|^2}{\rho} + \frac{1}{\mu} + \kappa$.

\textbf{Remark 2.4} If $\theta_{p,\mu}$ is $\kappa$-strongly convex, then there is no need to apply the second regularization, as this function is already endowed with the properties of $\theta_{p,\mu,\kappa}$.

\section*{2.1.3 An appropriate fast gradient method}

We denote by $p_{DS}^*$ the unique optimal solution to optimization problem (2.11). Further, let $p^* \in \mathbb{R}^m$ be an optimal solution to the dual optimization problem (2.2), and assume that the upper bound
\[
\|p^*\| \leq R
\]
is available for some $R \in \mathbb{R}_{++}$.

We now apply to the doubly regularized dual problem (2.11) the fast gradient method from [98, Algorithm 2.2.11].

\textbf{Algorithm 2.5} Let $w_0 = p_0 := 0 \in \mathbb{R}^m$, let $\rho, \mu, \kappa \in \mathbb{R}_{++}$, and set
\[
(\forall k \geq 0) \quad \begin{cases}
    p_{k+1} = w_k - \frac{1}{L(\rho, \mu, \kappa)} \nabla \theta_{p,\mu,\kappa}(w_k), \\
    w_{k+1} = p_{k+1} + \sqrt{\frac{L(\rho, \mu, \kappa)}{\rho}} (p_{k+1} - p_k).
\end{cases}
\]

By taking into account [98, Theorem 2.2.3], we obtain a sequence $(p_k)_{k \geq 0} \subseteq \mathbb{R}^m$ satisfying
\[
\theta_{p,\mu,\kappa}(p_k) - \theta_{p,\mu,\kappa}(p_{DS}^*) \leq \left( \theta_{p,\mu,\kappa}(0) - \theta_{p,\mu,\kappa}(p_{DS}^*) + \frac{\kappa}{2} \|p_{DS}^*\|^2 \right) \left( 1 - \sqrt{\frac{\kappa}{2L(\rho, \mu, \kappa)}} \right)^k \leq \left( \theta_{p,\mu,\kappa}(0) - \theta_{p,\mu,\kappa}(p_{DS}^*) + \frac{\kappa}{2} \|p_{DS}^*\|^2 \right) e^{-k \sqrt{\frac{\kappa}{2L(\rho, \mu, \kappa)}}} \forall k \geq 0.
\]

Since $p_{DS}^*$ is the unique optimal solution to (2.11), we have $\nabla \theta_{p,\mu,\kappa}(p_{DS}^*) = 0$, and therefore [98, Theorem 2.1.5] yields
\[
\|\nabla \theta_{p,\mu,\kappa}(p_k)\|^2 \leq 2L(\rho, \mu, \kappa) \left( \theta_{p,\mu,\kappa}(p_k) - \theta_{p,\mu,\kappa}(p_{DS}^*) \right) \leq 2L(\rho, \mu, \kappa) \left( \theta_{p,\mu,\kappa}(0) - \theta_{p,\mu,\kappa}(p_{DS}^*) \right) e^{-k \sqrt{\frac{\kappa}{2L(\rho, \mu, \kappa)}}} \forall k \geq 0.
\]

Due to the $\kappa$-strong convexity of $\theta_{p,\mu,\kappa}$, Theorem 2.1.8 in [98] states
\[
\|p_k - p_{DS}^*\|^2 \leq \frac{2}{\kappa} \left( \theta_{p,\mu,\kappa}(p_k) - \theta_{p,\mu,\kappa}(p_{DS}^*) \right) \leq \frac{2}{\kappa} \left( \theta_{p,\mu,\kappa}(0) - \theta_{p,\mu,\kappa}(p_{DS}^*) \right) e^{-k \sqrt{\frac{\kappa}{2L(\rho, \mu, \kappa)}}} \forall k \geq 0.
\]
In the following we will show that the rates of convergence for the decrease of \( \|\nabla \theta_{p,\mu}(p_k)\| \) and \( \theta(p_k) - \theta(p^*) \) are the same, namely equal to \( \mathcal{O}\left( \frac{1}{k} \ln \left( \frac{1}{k} \right) \right) \). This will allow us to efficiently recover approximately optimal solutions to the initial optimization problem \( (P) \).

### 2.1.4 Convergence of \( \theta(p_k) \) to \( \theta(p^*) \)

Using again \([98, \text{Theorem 2.1.8}]\), we obtain

\[
\| p_{DS}^* \|^2 \leq \frac{2}{\kappa} \left( \theta_{p,\mu,\kappa}(0) - \theta_{p,\mu,\kappa}(p_{DS}^*) \right) = \frac{2}{\kappa} \left( \theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*) - \frac{\kappa}{2} \| p_{DS}^* \|^2 \right),
\]

which implies that

\[
\| p_{DS}^* \|^2 \leq \frac{1}{\kappa} \left( \theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*) \right). \tag{2.18}
\]

On the other hand, in the light of \((2.15)\), it holds for all \( k \geq 0 \)

\[
\theta_{p,\mu}(p_k) - \theta_{p,\mu}(p_{DS}^*) \leq (\theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*)) \ e^{-k\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}} + \frac{\kappa}{2} \left( \| p_{DS}^* \|^2 - \| p_k \|^2 \right). \tag{2.19}
\]

Investigating the last term in the relation above and using \((2.17), (2.18)\), we get for all \( k \geq 0 \)

\[
\| p_{DS}^* \|^2 - \| p_k \|^2 = (\| p_{DS}^* \| - \| p_k \|) (\| p_{DS}^* \| + \| p_k \|) \\
\leq \| p_{DS}^* - p_k \| (\| p_{DS}^* \| + \| p_k \|) \\
\leq \| p_{DS}^* - p_k \| (2 \| p_{DS}^* \| + \| p_k - p_{DS}^* \|) \\
= \| p_{DS}^* - p_k \|^2 + 2 \| p_{DS}^* \| \| p_{DS}^* - p_k \| \\
\leq \frac{2}{\kappa} \left( \theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*) \right) \ e^{-k\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}} \\
+ 2\sqrt{\frac{2}{\kappa}} \left( \theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*) \right) \ e^{-\frac{k}{2}\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}} \\
\leq \frac{2}{\kappa} + 2\sqrt{2} \left( \theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*) \right) \ e^{-\frac{k}{2}\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}}.
\]

Inserting this estimate into \((2.19)\), we obtain for all \( k \geq 0 \)

\[
\theta_{p,\mu}(p_k) - \theta_{p,\mu}(p_{DS}^*) \leq (\theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*)) \left( e^{-k\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}} + (1 + \sqrt{2}) e^{-\frac{k}{2}\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}} \right) \\
\leq (2 + \sqrt{2}) (\theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*)) \ e^{-\frac{k}{2}\sqrt{\frac{\kappa}{\theta_{p,\mu,\kappa}}}}. \tag{2.20}
\]

Further, we have \( \theta_{p,\mu}(0) \leq \theta(0) \) and

\[
\theta_{p,\mu}(p_{DS}^*) \geq \theta(p_{DS}^*) - \rho D_f - \mu D_g \geq \theta(p^*) - \rho D_f - \mu D_g,
\]

and, from here,

\[
\theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS}^*) \leq \theta(0) - \theta(p^*) + \rho D_f + \mu D_g. \tag{2.21}
\]
Additionally, we conclude that

\[
\theta_{p,\mu}(p_{DS}) \leq \theta_{p,\mu}(p_D^*) + \frac{\kappa}{2} \|p_D^*\|^2 \leq \theta_{p,\mu}(p^*) + \frac{\kappa}{2} \|p^*\|^2 \leq \theta(p^*) + \frac{\kappa}{2} \|p^*\|^2,
\]

and, therefore, for all \(k \geq 0\)

\[
\theta_{p,\mu}(p_k) - \theta_{p,\mu}(p_{DS}) \geq \theta(p_k) - \rho D_f - \mu D_g - \theta(p^*) - \frac{\kappa}{2} \|p^*\|^2. \tag{2.26}
\]

In conclusion, we obtain for all \(k \geq 0\)

\[
\theta(p_k) - \theta(p^*) \leq \rho D_f + \mu D_g + \frac{\kappa}{2} \|p^*\|^2 + \theta_{p,\mu}(p_k) - \theta_{p,\mu}(p_{DS}) \leq (2.12), (2.20)
\]

\[
\text{\hspace{2cm}} \leq \rho D_f + \mu D_g + \frac{\kappa}{2} R^2 + (2 + \sqrt{2}) (\theta_{p,\mu}(0) - \theta_{p,\mu}(p_{DS})) e^{-\frac{k}{2} \sqrt{2 R^2}} \tag{2.21}
\]

\[
\text{\hspace{2cm}} + (2 + \sqrt{2}) (\theta(0) - \theta(p^*) + \rho D_f + \mu D_g) e^{-\frac{k}{2} \sqrt{2 R^2}}. \tag{2.23}
\]

Next, we fix \(\varepsilon > 0\). In order to get \(\theta(p_k) - \theta(p^*) \leq \varepsilon\) for a certain amount of iterations \(k\), we force all four terms in (2.23) to be less than or equal to \(\frac{\varepsilon}{4}\). Therefore, we choose

\[
\rho := \rho(\varepsilon) = \frac{\varepsilon}{4 D_f}, \quad \mu := \mu(\varepsilon) = \frac{\varepsilon}{4 D_g}, \quad \kappa := \kappa(\varepsilon) = \frac{\varepsilon}{2 R^2}. \tag{2.24}
\]

With these new parameters, we can simplify (2.23) to

\[
\theta(p_k) - \theta(p^*) \leq \frac{3 \varepsilon}{4} + (2 + \sqrt{2}) \left(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2}\right) e^{-\frac{k}{2} \sqrt{2 R^2}}.
\]

As one can see, the last term in the above estimate determines the number of iterations which is needed to obtain \(\varepsilon\)-accuracy for the dual objective function \(\theta\). Indeed, we obtain a worst-case estimate of

\[
\frac{\varepsilon}{4} \geq (2 + \sqrt{2}) \left(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2}\right) e^{-\frac{k}{2} \sqrt{2 R^2}} \iff e^{\frac{k}{2} \sqrt{2 R^2}} \geq \frac{4 (2 + \sqrt{2})}{\varepsilon} \left(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2}\right) \iff k \geq 2 \sqrt{\frac{\kappa}{L(\rho, \mu, \kappa)}} \ln \left(\frac{4 (2 + \sqrt{2})}{\varepsilon} \left(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2}\right)\right). \tag{2.25}
\]

A closer look on \(\frac{L(\rho, \mu, \kappa)}{\kappa}\) shows that

\[
\frac{L(\rho, \mu, \kappa)}{\kappa} = \frac{\|K\|^2}{\rho \kappa} + \frac{1}{\mu \kappa} + 1 = \frac{8 \|K\|^2 D_f R^2}{\varepsilon^2} + \frac{8 D_g R^2}{\varepsilon^2} + 1 = 1 + \frac{8 R^2}{\varepsilon^2} \left(\|K\|^2 D_f + D_g\right),
\]

hence, in order to obtain an approximately optimal solution to \((D)\), we need \(k = \mathcal{O}\left(\frac{1}{\varepsilon} \ln \left(\frac{1}{\varepsilon}\right)\right)\) iterations.
2.1.5 Convergence of $\|\nabla \theta_{\rho,\mu}(p_k)\|$ to 0

As it follows from (2.8), guaranteeing $\varepsilon$-optimality for the dual objective values is not sufficient for solving the initial primal optimization problem with a good convergence rate. In the following, we show that the fast gradient method in Algorithm 2.5 applied to the doubly regularized dual objective function $\theta_{\rho,\mu,\kappa}$ furnishes the desired properties for the decrease of $\|\nabla \theta_{\rho,\mu}(p_k)\|$.

It holds

$$
\|\nabla \theta_{\rho,\mu}(p_k)\| = \|\nabla \theta_{\rho,\mu,\kappa}(p_k) - \kappa p_k\| \leq \|\nabla \theta_{\rho,\mu,\kappa}(p_k)\| + \kappa \|p_k\| \quad \forall k \geq 0. \quad (2.26)
$$

Having a closer look on the first term in the right-hand side of the previous estimate, one can notice that

$$
\|\nabla \theta_{\rho,\mu,\kappa}(p_k)\| \leq \sqrt{2L(\rho, \mu, \kappa)} (\theta_{\rho,\mu}(0) - \theta_{\rho,\mu}(p_{DS}^*)) e^{-\frac{\kappa}{2} \sqrt{\frac{\kappa}{L(\rho, \mu, \kappa)}}} + \|p_{DS}^*\| \quad (2.21)
$$

On the other hand, the second term in the right-hand side of (2.26) can be estimated via

$$
\|p_k\| = \|p_k - p_{DS}^* + p_{DS}^*\| \leq \|p_k - p_{DS}^*\| + \|p_{DS}^*\| \quad (2.17)
$$

Furthermore, in order to gain an upper bound for the norm of $p_{DS}^*$, we notice that

$$
\theta(p^*) + \frac{\kappa}{2} \|p^*\|^2 \geq \theta_{\rho,\mu}(p^*) + \frac{\kappa}{2} \|p^*\|^2 \geq \theta_{\rho,\mu}(p_{DS}^*) + \frac{\kappa}{2} \|p_{DS}^*\|^2 \quad (2.6)
$$

$$
\geq \theta(p_{DS}^*) - \rho D_f - \mu D_g + \frac{\kappa}{2} \|p_{DS}^*\|^2 \\
\geq \theta(p^*) - \rho D_f - \mu D_g + \frac{\kappa}{2} \|p_{DS}^*\|^2,
$$

which implies

$$
\frac{\kappa}{2} \|p_{DS}^*\|^2 \leq \frac{\kappa}{2} \|p^*\|^2 + \rho D_f + \mu D_g,
$$

or, equivalently,

$$
\|p_{DS}^*\|^2 \leq \|p^*\|^2 + \frac{2\rho}{\kappa} D_f + \frac{2\mu}{\kappa} D_g.
$$

Hence,

$$
\|p_{DS}^*\| \leq \sqrt{\|p^*\|^2 + \frac{2\rho}{\kappa} D_f + \frac{2\mu}{\kappa} D_g} \quad (2.24)
= \sqrt{\|p^*\|^2 + \frac{\varepsilon}{2\kappa} \frac{\varepsilon}{2\kappa} \quad (2.24)
\leq \sqrt{\|p^*\|^2 + 2R^2} \quad (2.12)
\leq \sqrt{3R}, \quad (2.29)
$$
which, combined with (2.26), (2.27), and (2.28), provides the following estimate for the norm of the gradient of $\theta_{\rho,\mu}(p_k)$ for any $k \geq 0$, namely

$$\|\nabla \theta_{\rho,\mu}(p_k)\| \leq \left(\sqrt{L(\rho, \mu, \kappa)} + \sqrt{\kappa}\right) \sqrt{2 \left(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2}\right)e^{-\frac{2}{L(\rho, \mu, \kappa)}}} + \sqrt{3\kappa R} \tag{2.24}$$

where $L(\rho, \mu, \kappa) = L_{\rho,\mu}(\kappa)$. Resuming the achievements in the last two subsections, it follows that

$$k \geq 2\sqrt{\frac{\varepsilon^2 + 8R^2(\|K\|^2 D_f + D_g)}{\varepsilon \ln \left(\sqrt{L(\rho, \mu, \kappa)} + \sqrt{\kappa}\right) \sqrt{8R^2(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2})} \frac{2R}{(2 - \sqrt{3})\varepsilon}}}$$

$$k \geq 2\sqrt{\frac{\varepsilon^2 + 8R^2(\|K\|^2 D_f + D_g)}{\varepsilon \ln \left(\sqrt{4\|K\|^2 D_f + 4D_g + \frac{\varepsilon^2}{2R^2}} + \sqrt{\frac{\varepsilon^2}{2R^2}}\right) \sqrt{8R^2(\theta(0) - \theta(p^*) + \frac{\varepsilon}{2})} \frac{2R}{(2 - \sqrt{3})\varepsilon}}} \tag{2.31}$$

iterations of the fast gradient method in Algorithm 2.5. In the above estimate, we use that $L_{\rho,\mu}(\kappa) = 1 + \frac{8\kappa^2}{2R}$ and $L(\rho, \mu, \kappa) = \frac{4\|K\|^2 D_f + 4D_g + \frac{\varepsilon^2}{2R^2}}{2\kappa}$. Resuming the achievements in the last two subsections, it follows that

$$k = O\left(\frac{1}{\varepsilon} \ln \left(\frac{1}{\varepsilon}\right)\right)$$

iterations are needed to guarantee

$$\theta(p_k) - \theta(p^*) \leq \varepsilon \quad \text{and} \quad \|\nabla \theta_{\rho,\mu}(p_k)\| \leq \frac{\varepsilon}{R} \tag{2.32}$$

with a rate of convergence which is very similar except for constant factors.
2.1.6 How to construct an approximately primal optimal solution

Next, by making use of the approximate dual solution $p_k$, for $k \geq 0$, we construct an approximately primal optimal solution for the initial problem $(P)$ and investigate its accuracy. To this end, we will make use of the sequences $(x_{p,p_k})_{k \geq 0} \subseteq \text{dom } f$ and $(x_{\mu,p_k})_{k \geq 0} \subseteq \text{dom } g$ which are delivered by Algorithm 2.5. We will prove that, given a fixed accuracy $\varepsilon > 0$, we are able to reconstruct an approximately primal optimal solution such that, for $\rho$ and $\mu$ chosen as in (2.24), one gets

$$|f(x_{p,p_k}) + g(x_{\mu,p_k}) - v(D)| \leq 4\varepsilon, \quad (2.33)$$

$$\|Kx_{p,p_k} - x_{\mu,p_k}\| \leq \frac{\varepsilon}{R}, \quad (2.34)$$

in the same complexity of iterations as needed in order to satisfy (2.32). By means of weak duality, i.e., $v(D) \leq v(P)$, (2.33) would imply that $f(x_{p,p_k}) + g(x_{\mu,p_k}) \leq v(P) + 4\varepsilon$, which would further mean that $x_{p,p_k} \in \text{dom } f$ and $x_{\mu,p_k} \in \text{dom } g$ fulfilling (2.33) as well as (2.34) can be seen as approximately optimal and feasible solutions to the primal optimization problem $(P)$ with an accuracy which is proportional to $\varepsilon$. Let $k := k(\varepsilon)$ be the smallest index satisfying (2.25) and (2.31), thus guaranteeing (2.32).

Now let us prove the validity of the inequalities above. As $\nabla \theta_{\rho,\mu}(p_k) = Kx_{p,p_k} - x_{\mu,p_k}$, relation (2.34) follows directly from (2.32). Thus, we only have to prove that (2.33) is true. To this aim, we notice first that, since $\theta_{\rho,\mu}(p_k) - \theta(p^*) \leq \theta(p_k) - \theta(p^*) \leq \varepsilon$ and

$$\theta_{\rho,\mu}(p_k) - \theta(p^*) \overset{(2.6)}{=} \theta(p_k) - \rho D_f - \mu D_g - \theta(p^*) \overset{(2.24)}{=} \frac{\theta(p_k) - \theta(p^*)}{\geq 0} - \frac{\varepsilon}{2} \geq -\frac{\varepsilon}{2},$$

we have $|\theta_{\rho,\mu}(p_k) - \theta(p^*)| \leq \varepsilon$. From (2.7), it follows

$$|f(x_{p,p_k}) + g(x_{\mu,p_k}) + \theta(p^*)| \leq \|p_k\| \|\nabla \theta_{\rho,\mu}(p_k)\| + \varepsilon + \rho D_f + \mu D_g \overset{(2.24)}{\leq} \|p_k\| \|\nabla \theta_{\rho,\mu}(p_k)\| + 2\varepsilon \overset{(2.32)}{\leq} \frac{\varepsilon}{R} \|p_k\| + 2\varepsilon$$

In order to get an upper bound for $\|p_k\|$, we use (2.28) and (2.29), so that

$$\|p_k\| \overset{(2.24)}{=} \sqrt{\frac{2}{\kappa} \left( \theta(0) - \theta(p^*) + \frac{\varepsilon}{2} \right) e^{-\frac{\varepsilon}{2} \sqrt{\frac{\kappa}{\rho,\mu,\kappa}}} + \sqrt{3}R} \overset{(2.24)}{=} 2R \sqrt{\frac{1}{\varepsilon} \left( \theta(0) - \theta(p^*) + \frac{\varepsilon}{2} \right) e^{-\frac{\varepsilon}{2} \sqrt{\frac{\kappa}{\rho,\mu,\kappa}}} + \sqrt{3}R},$$

and, consequently, since $v(D) = -\theta(p^*)$, we obtain

$$|f(x_{p,p_k}) + g(x_{\mu,p_k}) - v(D)| \leq 2\varepsilon \left( \theta(0) - \theta(p^*) + \frac{\varepsilon}{2} \right) e^{-\frac{\varepsilon}{2} \sqrt{\frac{\kappa}{\rho,\mu,\kappa}}} + (\sqrt{3} + 2)\varepsilon.$$
Since $k = k(\varepsilon)$ was chosen in order to fulfill (2.31), it verifies

\[
k \geq 2 \sqrt{\frac{L(\rho, \mu, \kappa)}{\kappa} \ln \left( \frac{2 \sqrt{\varepsilon (\theta(0) - \theta(p^*) + \varepsilon)} + \varepsilon}{2 - \sqrt{3}\varepsilon} \right)}
\]

\[
\iff 4\varepsilon \geq 2 \sqrt{\varepsilon (\theta(0) - \theta(p^*) + \varepsilon)} e^{-\frac{\varepsilon}{2} \sqrt{\frac{\varepsilon}{4 (\rho, \mu, \kappa)}}} + (\sqrt{3} + 2)\varepsilon,
\]

thus (2.33) holds.

### 2.1.7 Existence of an optimal solution

In the following, we will study the convergence behavior of the primal sequences produced by the fast gradient method, i.e., we show their convergence to an optimal solution of $(P)$ when $\varepsilon \downarrow 0$. Let $(\varepsilon_n)_{n \geq 0} \subseteq \mathbb{R}^+$ be a decreasing sequence of positive scalars with $\lim_{n \to \infty} \varepsilon_n = 0$. For each $n \geq 0$, we can make $k = k(\varepsilon_n)$ iterations of Algorithm 2.5 with smoothing parameters $\rho_{\varepsilon_n}$, $\mu_{\varepsilon_n}$, and $\kappa_{\varepsilon_n}$ given by (2.24) in order to have (2.33) and (2.34) satisfied. For $n \geq 0$, we denote

\[
\overline{x}_n := x_{\rho_{\varepsilon_n}, \mu_{\varepsilon_n}, \kappa_{\varepsilon_n}} \in \text{dom } f \quad \text{and} \quad \overline{y}_n := x_{\rho_{\varepsilon_n}, \mu_{\varepsilon_n}, \kappa_{\varepsilon_n}} \in \text{dom } g.
\]

Due to the boundedness of $\text{dom } f$, its closure $\text{cl}(\text{dom } f)$ is weakly compact (see [11, Theorem 3.3]) and there exists a subsequence $(\overline{x}_{n_l})_{l \geq 0}$ and $\overline{x} \in \mathcal{H}$ such that $(\overline{x}_{n_l})$ weakly converges to $\overline{x} \in \text{cl}(\text{dom } f)$ when $l \to +\infty$. Since $K : \mathcal{H} \to \mathbb{R}^m$ is linear and bounded, the sequence $K\overline{x}_{n_l}$ will converge to $K\overline{x} \in \mathbb{R}^m$ when $l \to +\infty$. In view of relation (2.34), we get

\[
0 \leq \left\| K\overline{x}_{n_l} - \overline{y}_{n_l} \right\| \leq \frac{\varepsilon_{n_l}}{R} \forall l \geq 0. \tag{2.35}
\]

Now, since the sequence $(\overline{y}_{n_l})_{l \geq 0} \subseteq \text{dom } g$ is bounded, there exists a subsequence of it (still denoted by $(\overline{y}_{n_l})_{l \geq 0}$) and an element $\overline{y} \in \text{cl}(\text{dom } g)$ such that $\overline{y}_{n_l} \to \overline{y}$ when $l \to +\infty$. Taking $l \to +\infty$ in (2.35), it follows $K\overline{x} = \overline{y}$. Furthermore, due to (2.33), we have

\[
f(\overline{x}_{n_l}) + g(\overline{y}_{n_l}) \leq v(D) + 4\varepsilon_{n_l} \quad \forall l \geq 0,
\]

and, by using the lower semicontinuity of $f$ and $g$ and [11, Theorem 9.1], we obtain

\[
f(\overline{x}) + g(K\overline{x}) \leq \liminf_{l \to \infty} \left\{ f(\overline{x}_{n_l}) + g(\overline{y}_{n_l}) \right\} \leq \liminf_{l \to \infty} \{v(D) + 4\varepsilon_{n_l}\} = v(D) \leq v(P).
\]

Since $v(P) \in \mathbb{R}$, we have $\overline{x} \in \text{dom } f$ and $K\overline{x} \in \text{dom } g$, which yields that $\overline{x}$ is an optimal solution to $(P)$.

### 2.1.8 Improving the convergence rates

Our next aim is to investigate how additional assumptions on the functions $f$ and/or $g$ influence the implementation of the double smoothing approach and its complexity. Within this subsection, we work under the following standing assumptions:
Assumption 2.6 Let $f \in \Gamma(\mathcal{H})$ be a function with bounded effective domain, let $g \in \Gamma(\mathbb{R}^m)$ be $\mu$-strongly convex for some $\mu \in \mathbb{R}_{++}$, and let $K : \mathcal{H} \to \mathbb{R}^m$ be a linear and bounded operator fulfilling $K(\text{dom } f) \cap \text{dom } g \neq \emptyset$.

Different to the investigations made before, we strengthen here the convexity assumptions on $g$ (by asking the convexity of $g$ to be strong), but allow in counterpart dom $g$ to be unbounded.

By taking into account Remark 2.2, since $g$ is proper, $\mu$-strongly convex and lower semicontinuous, $g^*$ is differentiable and $\nabla g^*$ is $\frac{1}{\mu}$-Lipschitz continuous (cf. [11, Theorem 18.15]). Thus $(g^* \circ \text{Id})$ is Fréchet differentiable, too, and its gradient is $\frac{1}{\mu}$-Lipschitz continuous. By denoting

$$x_{g,p} := \nabla g^*(-p) = -\nabla (g^* \circ \text{Id})(p),$$

one has that $-p \in \partial g(x_{g,p})$, or, equivalently, $0 \in \partial (\langle p, \cdot \rangle + g)(x_{g,p})$, which means that $x_{g,p}$ is the unique optimal solution (see [20, Lemma 2.33]) of the optimization problem

$$\inf_{x \in \mathbb{R}^m} \{ \langle p, x \rangle + g(x) \}.$$ 

For $\rho \in \mathbb{R}_{++}$, let $\theta_\rho : \mathbb{R}^m \to \mathbb{R}$ be defined by $\theta_\rho(p) = f^*_\rho(K^*p) + g^*(-p)$. The function $\theta_\rho$ is differentiable with $L(\rho)$-Lipschitz continuous gradient

$$\nabla \theta_\rho(p) = \nabla (f^*_\rho \circ K^*)(p) + \nabla (g^* \circ \text{Id})(p) = Kx_{f,p} - x_{g,p} \forall p \in \mathbb{R}^m,$$

where $L(\rho) := \frac{\|K\|^2}{\rho} + \frac{1}{\mu}$.

In the light of Proposition 2.3, we get

$$\theta_\rho(p) \leq \theta(p) \leq \theta_\rho(p) + \rho D_f \quad \forall p \in \mathbb{R}^m.$$  \hfill (2.36)

For $\kappa \in \mathbb{R}_{++}$, we introduce $\theta_{\rho,\kappa} : \mathbb{R}^m \to \mathbb{R}$, $\theta_{\rho,\kappa}(p) = f^*_\rho(K^*p) + g^*(-p) + \frac{\kappa}{2}\|p\|^2$.

In [35], we have shown that when applying Algorithm 2.5 to the doubly regularized dual objective $\theta_{\rho,\kappa}$, the complexity for gaining inequalities in the sense of the one given in (2.32) is still $\mathcal{O} \left( \frac{1}{\varepsilon} \ln \left( \frac{1}{\varepsilon} \right) \right)$ whereas the hidden factors are noteworthy smaller.

In all three situations addressed in the sequel, the construction of the approximate primal solutions and the proof of the existence of an optimal solution to the primal problem can be made in analogy to Subsection 2.1.6 and Subsection 2.1.7, respectively.

It is worth to notice that the additional assumptions furnish an improvement of the complexity, which is motivated by the fact that constants of strong convexity and/or Lipschitz constants of the gradient are already available, thus they do not need to be constructed in the smoothing process to fulfill the $\varepsilon$-accuracy condition.

The case $f$ is strongly convex

Additionally to the standing assumptions, we assume first that the function $f : \mathcal{H} \to \mathbb{R}$ is $\rho$-strongly convex ($\rho > 0$), but remove the boundedness assumption on its domain. In this situation, the first smoothing as done in Subsection 2.1.2 can be omitted and the fast gradient method can be applied to the minimization problem

$$\inf_{p \in \mathbb{R}^m} \theta_\kappa(p),$$  \hfill (2.37)
where \( \theta_\kappa : \mathbb{R}^m \to \mathbb{R} \), \( \theta_\kappa := f^* (K^* p) + g^* (-p) + \frac{\kappa}{2} \| p \|^2 \), with \( \kappa \in \mathbb{R}_{++} \), is a \( \kappa \)-strongly convex and differentiable function with \( L(\kappa) \)-Lipschitz continuous gradient \( \nabla \theta_\kappa \), where \( L(\kappa) := \frac{\| K \|^2}{p} + \frac{1}{\mu} + \kappa \).

This gives rise to a sequence \( (p_k)_{k \geq 0} \) satisfying

\[
\theta_\kappa (p_k) - \theta_\kappa (p_{DS}^*) \leq \left( \theta_\kappa (0) - \theta_\kappa (p_{DS}^*) + \frac{\kappa}{2} \| p_{DS}^* \|^2 \right) e^{-k \sqrt{\frac{\kappa}{\mu}}} \tag{2.14}
\]

\[
= (\theta (0) - \theta (p_{DS}^*)) e^{-k \sqrt{\frac{\kappa}{\mu}}} \quad \forall k \geq 0, \tag{2.38}
\]

where \( p_{DS}^* \) denotes the unique optimal solution of the problem \( (2.37) \). Thus, from \( (2.39) \), it follows

\[
\| \nabla \theta_\kappa (p_k) \|^2 \leq 2L(\kappa) \left( \theta (0) - \theta (p_{DS}^*) \right) e^{-k \sqrt{\frac{\kappa}{\mu}}}, \tag{2.40}
\]

and

\[
\| p_k - p_{DS}^* \|^2 \leq \frac{2}{\kappa} \left( \theta_\kappa (p_k) - \theta_\kappa (p_{DS}^*) \right) \leq \frac{2}{\kappa} \left( \theta (0) - \theta (p_{DS}^*) \right) e^{-k \sqrt{\frac{\kappa}{\mu}}} \quad \forall k \geq 0. \tag{2.41}
\]

Additionally, in all iterations \( k \geq 0 \), we have

\[
\| p_{DS}^* \|^2 \leq \frac{1}{\kappa} \left( \theta (0) - \theta (p_{DS}^*) \right), \tag{2.42}
\]

and, by \( (2.41) \) and \( (2.42) \),

\[
\| p_{DS}^* \|^2 - \| p_k \|^2 \leq \| p_k - p_{DS}^* \| (2\| p_{DS}^* \| + \| p_k - p_{DS}^* \|) \leq \frac{2 + 2\sqrt{2}}{\kappa} \left( \theta (0) - \theta (p_{DS}^*) \right) e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\mu}}},
\]

thus

\[
\theta (p_k) - \theta (p_{DS}^*) \leq \left( \theta (0) - \theta (p_{DS}^*) \right) e^{-k \sqrt{\frac{\kappa}{\mu}}} + \frac{\kappa}{2} \left( \| p_{DS}^* \|^2 - \| p_k \|^2 \right) \leq \left( \theta (0) - \theta (p_{DS}^*) \right) \left( e^{-k \sqrt{\frac{\kappa}{\mu}}} + (1 + \sqrt{2}) e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\mu}}} \right) \leq (2 + \sqrt{2}) \left( \theta (0) - \theta (p_{DS}^*) \right) e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\mu}}} \quad \forall k \geq 0.
\]

We denote by \( p^* \in \mathbb{R}^m \) an optimal solution to the dual optimization problem \( (2.2) \) and assume that the upper bound \( \| p^* \| \leq R \) is available for some \( R \in \mathbb{R}_{++} \). Thus, since \( \theta (p_{DS}^*) \leq \theta_\kappa (p_{DS}^*) \leq \theta_\kappa (p^*) = \theta (p^*) + \frac{\kappa}{2} \| p^* \|^2 \), we obtain for all \( k \geq 0 \)

\[
\theta (p_k) - \theta (p^*) \leq \frac{\kappa}{2} \| p^* \|^2 + \theta (p_k) - \theta (p_{DS}^*) \leq \frac{\kappa}{2} R^2 + (2 + \sqrt{2}) \left( \theta (0) - \theta (p^*) \right) e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\mu}}}.
\]

Hence, when \( \varepsilon > 0 \), in order to guarantee \( \varepsilon \)-accuracy for the dual objective function, we can force both terms in the above estimate to be less than or equal to \( \frac{\varepsilon}{2} \). To this end, by taking

\[
\kappa := \kappa (\varepsilon) = \frac{\varepsilon}{R^2},
\]
we will need, in contrast to (2.25),
\[ k \geq 2\sqrt{\frac{L(\kappa)}{\kappa}} \ln \left( 2(2 + \sqrt{2}) \frac{(\theta(0) - \theta(p^*))}{\varepsilon} \right), \]
i.e., \( k = \mathcal{O} \left( \frac{1}{\sqrt{\varepsilon}} \ln \left( \frac{1}{\varepsilon} \right) \right) \) iterations. Further, using (2.40), we have
\[ \|\nabla \theta_\kappa(p_k)\| \leq \sqrt{2L(\kappa)(\theta(0) - \theta(p^*))} e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\varepsilon}}} \forall k \geq 0. \]

On the other hand, using
\[ \|p_k\| \leq \|p_k - p_{DS}^*\| + \|p_{DS}^*\| \leq \sqrt{\frac{2}{\kappa} (\theta(0) - \theta(p^*))} e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\varepsilon}}} + \|p_{DS}^*\|, \]and the relation \( \theta(p^*) + \frac{\varepsilon}{2} \|p_{DS}^*\|^2 \leq \theta_\kappa(p_{DS}^*) \leq \theta_\kappa(p^*) = \theta(p^*) + \frac{\varepsilon}{2} \|p^*\|^2 \), which yields
\[ \|p_{DS}^*\| \leq \|p^*\| \leq R, \]
we obtain
\[ \|\nabla \theta(p_k)\| \leq \|\nabla \theta_\kappa(p_k)\| + \kappa \|p_k\| \leq \left( \sqrt{L(\kappa) + \kappa} \right) \sqrt{2(\theta(0) - \theta(p^*))} e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\varepsilon}}} + \kappa R \]
\[ = \left( \sqrt{L(\kappa) + \kappa} \right) \sqrt{2(\theta(0) - \theta(p^*))} e^{-\frac{k}{2} \sqrt{\frac{\kappa}{\varepsilon}}} + \frac{\varepsilon}{R} \forall k \geq 0. \]

Therefore, in order to guarantee \( \|Kx_{f,p_k} - x_{p,p_k}\| = \|\nabla \theta(p_k)\| \leq \frac{2\varepsilon}{R} \), we need \( k = \mathcal{O} \left( \frac{1}{\sqrt{\varepsilon}} \ln \left( \frac{1}{\varepsilon} \right) \right) \) iterations, which coincides with the convergence rate for the dual objective values.

**The case \( g \) is everywhere differentiable with Lipschitz continuous gradient**

Throughout this subsection, additionally to the standing assumptions, we assume that \( g : \mathbb{R}^m \rightarrow \mathbb{R} \) has full domain and that it is differentiable with \( \frac{1}{\mu} \)-Lipschitz continuous gradient, for \( \kappa \in \mathbb{R}_{++} \). In this situation, the second smoothing as done in Subsection 2.1.2 can be omitted and Algorithm 2.5 can be applied to the minimization problem

\[ \inf_{p \in \mathbb{R}^m} \theta_\rho(p), \quad (2.43) \]
where \( \theta_\rho : \mathbb{R}^m \rightarrow \mathbb{R} \), \( \theta_\rho := f_\rho^* (K^* p) + g^* (-p) \), is \( \kappa \)-strongly convex due to [11, Theorem 18.15] and differentiable with \( L(\rho) \)-Lipschitz continuous gradient \( \nabla \theta_\rho \), where \( L(\rho) := \frac{\|K\|^2}{\rho} + \frac{1}{\mu} \).

This gives rise to a sequence \( (p_k)_{k \geq 0} \) satisfying
\[ \theta_\rho(p_k) - \theta_\rho(p_{DS}^*) \leq \left( \theta_\rho(0) - \theta_\rho(p_{DS}^*) + \frac{\kappa}{2} \|p_{DS}^*\|^2 \right) e^{-k \sqrt{\frac{\kappa}{\varepsilon}}} \] \[ \leq 2 \theta_\rho(0) - \theta_\rho(p_{DS}^*) e^{-k \sqrt{\frac{\kappa}{\varepsilon}}}, \]
and
\[ \|\nabla \theta_\rho(p_k)\|^2 \leq 4L(\rho)(\theta_\rho(0) - \theta_\rho(p_{DS}^*)) e^{-k \sqrt{\frac{\kappa}{\varepsilon}}} \forall k \geq 0, \]
where \( p_{DS}^* \) denotes the unique optimal solution of the problem (2.43). We denote by \( p^* \in \mathbb{R}^m \) the unique optimal solution of the dual optimization problem (2.2) and
would like to notice that in this context it is not necessary to know an upper bound of the norm of the dual optimal solution.

Since $\theta_\rho(0) \leq \theta(0)$ and $\theta_\rho(p_{DS}^*) \geq \theta(p_{DS}^*) - \rho D_f \geq \theta(p^*) - \rho D_f$, we obtain

$$\theta_\rho(0) - \theta_\rho(p_{DS}^*) \leq \theta(0) - \theta(p^*) + \rho D_f.$$  \hfill (2.47)

On the other hand, since $\theta_\rho(p_k) - \theta_\rho(p_{DS}^*) \geq \theta(p_k) - \rho D_f - \theta(p^*)$, it follows

$$\theta(p_k) - \theta(p^*) \leq \rho D_f + \theta(p_k) - \theta(p_{DS}^*) \leq \rho D_f + 2 \left( \theta(0) - \theta(p^*) + \rho D_f \right) e^{-k \sqrt{\frac{\rho}{\kappa}}} \forall k \geq 0.$$  

Hence, when $\varepsilon > 0$, in order to guarantee $\varepsilon$-optimality for the dual objective, we force both terms in the above estimate less than or equal to $\frac{\varepsilon}{2}$. By taking

$$\rho := \rho(\varepsilon) = \frac{\varepsilon}{2D_f},$$  \hfill (2.48)

in contrast to (2.25), we need

$$k \geq \sqrt{\frac{L(\rho)}{\kappa}} \ln \left( \frac{4 \left( \theta(0) - \theta(p^*) + \frac{\varepsilon}{2} \right)}{\varepsilon} \right),$$

i.e., $k = \mathcal{O}\left( \frac{1}{\varepsilon^2} \ln \left( \frac{1}{\varepsilon} \right) \right)$ iterations to obtain $\varepsilon$-accuracy for the dual objective values.

From (2.46), we obtain as well

$$\|\nabla \theta_\rho(p_k)\| \leq 2L(\rho) \left( \theta_{\rho}(0) - \theta_\rho(p_{DS}^*) \right) e^{-\frac{k}{2} \sqrt{\frac{\rho}{\kappa}}} \tag{2.47}$$

$$\leq 2L(\rho) \left( \theta(0) - \theta(p^*) + \rho D_f \right) e^{-\frac{k}{2} \sqrt{\frac{\rho}{\kappa}}} \tag{2.48}$$

$$\leq 2L(\rho) \left( \theta(0) - \theta(p^*) + \frac{\varepsilon}{2} \right) e^{-\frac{k}{2} \sqrt{\frac{\rho}{\kappa}}} \forall k \geq 0.$$  

Therefore, in order to guarantee $\|Kx_{f,p_k} - x_{g,p_k}\| = \|\nabla \theta_\rho(p_k)\| \leq \varepsilon$, we need $k = \mathcal{O}\left( \frac{1}{\varepsilon^2} \ln \left( \frac{1}{\varepsilon} \right) \right)$ iterations, which is the same convergence rate as for the dual objective values.

**The case $f$ is strongly convex and $g$ is everywhere differentiable with Lipschitz continuous gradient**

The third favorable situation which we address is when, additionally to the standing assumptions, the function $f : \mathcal{H} \to \mathbb{R}$ is $\rho$-strongly convex ($\rho \in \mathbb{R}_{++}$), however, without assuming anymore that dom $f$ is bounded, and the function $g : \mathbb{R}^m \to \mathbb{R}$ has full domain and it is differentiable with $\frac{1}{\kappa}$-Lipschitz continuous gradient ($\kappa \in \mathbb{R}_{++}$). In this case both the first and second smoothing can be omitted and Algorithm 2.5 can be applied to the minimization problem

$$\inf_{p \in \mathbb{R}^m} \theta(p),$$  \hfill (2.49)

where $\theta : \mathbb{R}^m \to \mathbb{R}$, $\theta := f^*(K^*p) + g^*(-p)$, is a $\kappa$-strongly convex and differentiable function with $L$-Lipschitz continuous gradient $\nabla \theta$, where $L := \frac{\|K\|^2}{\rho} + \frac{1}{\rho}$. We denote
by \( p^* \in \mathbb{R}^m \) the unique optimal solution of \((D)\), for which it is not necessary to know an upper bound of its norm.

This gives rise to a sequence \((p_k)_{k \geq 0}\) satisfying

\[
\theta(p_k) - \theta(p^*) \leq \left( \theta(0) - \theta(p^*) + \frac{\kappa}{2} \|p^*\|^2 \right) e^{-k\sqrt{\kappa}} \leq 2(\theta(0) - \theta(p^*)) e^{-k\sqrt{\kappa}},
\]

and

\[
\|\nabla \theta(p_k)\|^2 \leq 4L (\theta(0) - \theta(p^*)) e^{-k\sqrt{\kappa}} \quad \forall k \geq 0.
\]

From here, when \( \varepsilon > 0 \), we have

\[
2(\theta(0) - \theta(p^*)) e^{-k\sqrt{\kappa}} \leq \varepsilon \Leftrightarrow k \geq \sqrt{\frac{L}{\kappa}} \ln \left( \frac{2(\theta(0) - \theta(p^*))}{\varepsilon} \right),
\]

while

\[
2\sqrt{L(\theta(0) - \theta(p^*))} e^{-\frac{k}{2}\sqrt{\kappa}} \leq \varepsilon \Leftrightarrow k \geq 2\sqrt{\frac{L}{\kappa}} \ln \left( \frac{2\sqrt{L(\theta(0) - \theta(p^*))}}{\varepsilon} \right).
\]

In conclusion, in order to guarantee \( \varepsilon \)-accuracy for the dual objective values and for the decrease of \( \|\nabla \theta(\cdot)\| \) to 0, we need \( \mathcal{O} \left( \ln \left( \frac{1}{\varepsilon} \right) \right) \) iterations.

### 2.2 Variable Smoothing

The subject of this section, which is based on our article in [40], is to introduce and investigate the convergence properties of an algorithm for solving nondifferentiable optimization problems in the form of

\[
\inf_{x \in \mathcal{H}} \{ f(x) + g(Kx) \}, \quad (2.50)
\]

where \( \mathcal{H} \) and \( \mathcal{G} \) are real Hilbert spaces, \( f \in \Gamma(\mathcal{H}) \) and \( g \in \Gamma(\mathcal{G}) \) are convex and Lipschitz continuous functions, and the operator \( K : \mathcal{H} \to \mathcal{G} \) is linear and bounded. By replacing the functions \( f \) and \( g \) through their Moreau envelopes, an approach which can be seen as part of the family of smoothing techniques introduced in [99–101], we approximate \((2.50)\) by a convex optimization problem possessing some differentiable objective function with Lipschitz continuous gradient. This smoothing approach can be seen as the counterpart of the so-called double smoothing method investigated in Section 2.1, which assumes, in two steps, the smoothing of the Fenchel dual problem to \((2.50)\). In contrast to that approach, which asks for the boundedness of the effective domains of \( f \) and \( g \), determinant is here the boundedness of the effective domains of their conjugates \( f^* \) and \( g^* \), which is automatically guaranteed by the Lipschitz continuity of \( f \) and \( g \) (cf. [21]), respectively. In order to solve the resulting convex and differentiable optimization problem, we propose an extended version of the accelerated gradient method by Nesterov (cf. [97]) for convex minimization problems allowing a successive reduction of the smoothing parameters involved. This extension is strongly necessary since variable smoothing parameters lead to variable objective functions for the problem under investigation.
2.2 Variable Smoothing

The variable smoothing method which we propose in this section yields for the minimization of the primal objective function in (2.50) a rate of convergence of order \( O(\ln k/k) \). In the particular case when the smoothing parameters are not variable and therefore considered to be constant, the order of the rate of convergence becomes \( O(1/k) \). However, the use of variable smoothing parameters is favorable in view of a good approximation of the original problem, although the theoretical rate of convergence is not as good as when these are constant. In the first case the approach generates a sequence of iterates \((x_k)_{k \geq 1}\) such that \((f(x_k) + g(Kx_k))_{k \geq 1}\) converges to the optimal objective value of (2.50). In the case of constant smoothing variables, the approach provides a sequence of iterates which solves the problem in (2.50) with a given a priori accuracy, however, the sequence \((f(x_k) + g(Kx_k))_{k \geq 1}\) may not converge to the optimal primal objective value of this problem.

2.2.1 Problem description

The optimization problem that we investigate in this section is

\[
(P) \quad \inf_{x \in \mathcal{H}} \{ f(x) + g(Kx) \},
\]

where \( \mathcal{H} \) and \( \mathcal{G} \) are real Hilbert spaces, \( K : \mathcal{H} \to \mathcal{G} \) is a bounded linear operator, and \( f \in \Gamma(\mathcal{H}) \) and \( g \in \Gamma(\mathcal{G}) \) are functions such that \( f \) is \( L_f \)-Lipschitz continuous and \( g \) is \( L_g \)-Lipschitz continuous for some \( L_f, L_g \in \mathbb{R}^+ \). According to [21, Proposition 4.4.6], it holds that

\[
\text{dom } f^* \subseteq L_f B_{\mathcal{H}} \quad \text{and} \quad \text{dom } g^* \subseteq L_g B_{\mathcal{G}}, \tag{2.51}
\]

i.e., the effective domains of the conjugate functions \( f^* \) and \( g^* \) are bounded.

In the sequel, we show, on the one hand, that the two approaches of using variable as well as constant smoothing parameters can be designed to keep their convergence behavior in the case when \( f \) is differentiable with Lipschitz continuous gradient. On the other hand, we show that they can also be employed for solving the extended version of (2.50), i.e.,

\[
\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(K_ix) \right\}, \tag{2.52}
\]

where \( \mathcal{G}_i \) are real Hilbert spaces, \( g_i : \mathcal{G}_i \to \mathbb{R} \), are convex and Lipschitz continuous functions, and \( K_i : \mathcal{H} \to \mathcal{G}_i, \ i = 1, \ldots, m \), are bounded linear operators.

We would like to mention that variable smoothing parameters have been recently considered in the PRISMA algorithm (cf. [102]) for solving nonsmooth optimization problems having as objective the sum of three convex functions with different properties. However, our approach allows the consideration of compositions with linear bounded operators present in the objective function which is relevant in lots of practical applications. Beyond this, when comparing it to the popular augmented Lagrangian method (ALM) and the alternating direction method of multipliers (ADMM) (see [42]), our method has the advantage that the bounded linear operators (and their adjoints) are evaluated via explicit forward steps, while for the nondifferentiable functions separate proximal steps are performed. Thus, in view of the implementation, the variable smoothing method shares similarities with the recently introduced class of primal-dual algorithms (see, for instance, [37–39, 43, 48, 58, 122]).
2.2.2 The smoothing of the problem \((P)\)

The algorithms we would like to introduce and analyze from the point of view of their convergence properties assume in a first instance an appropriate smoothing of the problem \((P)\) which we are going to describe in the following.

For \(\rho \in \mathbb{R}^+\), we smooth \(f\) via its Moreau envelope of parameter \(\rho\), \(\rho f : \mathcal{H} \to \mathbb{R}\),

\[
\rho f(x) = \left( f \circ \frac{1}{2\rho} \cdot \| \cdot \|^2 \right)(x), \quad \text{for every } x \in \mathcal{H}.
\]

According to the Fenchel–Moreau Theorem and due to [11, Theorem 15.3], one has for \(x \in \mathcal{H}\)

\[
\rho f(x) = \left( f^{**} \circ \frac{1}{2\rho} \cdot \| \cdot \|^2 \right)(x) = \left( f^* + \frac{\rho}{2} \cdot \| \cdot \|^2 \right)^*(x) = \sup_{p \in \mathcal{H}} \{ (x, p) - f^*(p) - \frac{\rho}{2} \|p\|^2 \}.
\]

As already seen, \(\rho f\) is differentiable and its gradient (cf. (1.13) and (1.14))

\[
\nabla(\rho f) : \mathcal{H} \to \mathcal{H}, \quad \nabla(\rho f) = \frac{1}{\rho} (x - \text{Prox}_{\rho f}(x)) = \text{Prox}_{\frac{1}{\rho} f^*} \left( \frac{x}{\rho} \right) \quad \forall x \in \mathcal{H},
\]

is \(\frac{1}{\rho}\)-Lipschitz continuous.

For \(\mu \in \mathbb{R}^+\), we smooth \((g \circ K)\) via

\[(\mu g \circ K) : \mathcal{H} \to \mathbb{R}, \quad (\mu g \circ K)(x) = \left( g \circ \frac{1}{2\mu} \cdot \| \cdot \|^2 \right)(Kx)
\]

for every \(x \in \mathcal{H}\). According to the Fenchel–Moreau Theorem and due to [11, Theorem 15.3], one has

\[
(\mu g \circ K)(x) = \left( g^{**} \circ \frac{1}{2\mu} \cdot \| \cdot \|^2 \right)(Kx) = \left( g^* + \frac{\mu}{2} \cdot \| \cdot \|^2 \right)^*(Kx)
\]

\[
= \sup_{p \in \mathcal{G}} \left\{ \langle Kx, p \rangle - g^*(p) - \frac{\mu}{2} \|p\|^2 \right\} \quad \forall x \in \mathcal{H}.
\]

The function \((\mu g \circ K)\) is differentiable and its gradient \(\nabla(\mu g \circ K) : \mathcal{H} \to \mathcal{H}\) fulfills for all \(x \in \mathcal{H}\) (cf. (1.13) and (1.14))

\[
\nabla(\mu g \circ K)(x) = K^* \nabla(\mu g)(Kx) = \frac{1}{\mu} K^* (Kx - \text{Prox}_{\mu g}(Kx)) = K^* \text{Prox}_{\frac{1}{\mu} g^*} \left( \frac{Kx}{\mu} \right).
\]

Further, for every \(x, y \in \mathcal{H}\), it holds (see (1.11))

\[
\| \nabla(\mu g \circ K)(x) - \nabla(\mu g \circ K)(y) \| \leq \frac{1}{\mu} \|K\| \| (Kx - \text{Prox}_{\mu g}(Kx)) - (Ky - \text{Prox}_{\mu g}(Ky)) \|
\]

\[
\leq \frac{\|K\|^2}{\mu} \| x - y \|,
\]

which shows that \(\nabla(\mu g \circ K)\) is \(\frac{\|K\|^2}{\mu}\)-Lipschitz continuous.

Finally, we consider as smoothing function for \(f + g \circ K\) the function \(F^{\rho,\mu} : \mathcal{H} \to \mathbb{R}\),

\[
F^{\rho,\mu}(x) = \rho f(x) + \mu g \circ K(x), \quad \text{which is differentiable with } L(\rho, \mu)\text{-Lipschitz continuous gradient } \nabla F^{\rho,\mu} : \mathcal{H} \to \mathcal{H}\text{ given by}
\]

\[
\nabla F^{\rho,\mu}(x) = \text{Prox}_{\frac{1}{\rho} f^*} \left( \frac{x}{\rho} \right) + K^* \text{Prox}_{\frac{1}{\mu} g^*} \left( \frac{Kx}{\mu} \right) \quad \forall x \in \mathcal{H},
\]
where \( L(\rho, \mu) := \frac{1}{\rho} + \frac{\|K\|^2}{\mu} \).

For \( \rho_2 \geq \rho_1 > 0 \) and every \( x \in H \), it holds (cf. (2.51))

\[
\begin{align*}
\rho_1 f(x) &= \sup_{p \in \text{dom} f^*} \left\{ (x, p) - f^*(p) - \frac{\rho_1}{2} \|p\|^2 \right\} \\
&\leq \sup_{p \in \text{dom} f^*} \left\{ (x, p) - f^*(p) - \frac{\rho_2}{2} \|p\|^2 \right\} + \sup_{p \in \text{dom} f^*} \left\{ \frac{\rho_2 - \rho_1}{2} \|p\|^2 \right\} \\
&\leq \rho_2 f(x) + (\rho_2 - \rho_1) \frac{L_f^2}{2},
\end{align*}
\]

which yields, letting \( \rho_1 \downarrow 0 \) (cf. [11, Proposition 12.32]),

\[
\rho_2 f(x) \leq f(x) \leq \rho_2 f(x) + \rho_2 \frac{L_f^2}{2}.
\]

Similarly, for \( \mu_2 \geq \mu_1 > 0 \) and every \( y \in G \), it holds

\[
\mu_1 g(y) \leq \mu_2 g(y) + (\mu_2 - \mu_1) \frac{L_g^2}{2},
\]

and

\[
\mu_2 g(y) \leq g(y) \leq \mu_2 g(y) + \mu_2 \frac{L_g^2}{2}.
\]

Consequently, for \( \rho_2 \geq \rho_1 > 0, \mu_2 \geq \mu_1 > 0 \) and every \( x \in H \), we have

\[
F^{\rho_2, \mu_2}(x) \leq F^{\rho_1, \mu_1}(x) \leq F^{\rho_2, \mu_2}(x) + (\rho_2 - \rho_1) \frac{L_f^2}{2} + (\mu_2 - \mu_1) \frac{L_g^2}{2},
\]

and

\[
F^{\rho_2, \mu_2}(x) \leq F^{\rho_2, \mu_2}(x) + \rho_2 \frac{L_f^2}{2} + \mu_2 \frac{L_g^2}{2}.
\]

### 2.2.3 The variable and the constant smoothing algorithm

In the following we denote by \( F : H \to \mathbb{R}, F(x) = f(x) + g(Kx) \), the objective function of \( (P) \). The variable smoothing algorithm, which we present at the beginning of this subsection, can be seen as an extension of the accelerated gradient method of Nesterov (cf. [97]) by using variable smoothing parameters, which are updated in each iteration.

**Algorithm 2.7** Let \( y_1 = x_0 \in H, (\rho_k)_{k \geq 1}, (\mu_k)_{k \geq 1} \subseteq \mathbb{R}^+, \) let \( t_1 = 1 \), and set

\[
\begin{align*}
L_k &= \frac{1}{\rho_k} + \frac{\|K\|^2}{\mu_k}, \\
x_k &= y_k - \frac{1}{L_k} \left( \text{Prox}_{\frac{1}{\rho_k}} f^* \left( \frac{y_k}{\mu_k} \right) + K^* \text{Prox}_{\frac{1}{\rho_k}} \left( \frac{y_k}{\mu_k} \right) \right), \\
\rho_{k+1} &= \frac{1}{1 + \frac{4}{L_k} t_k} \\
y_{k+1} &= x_k + \frac{t_k - 1}{t_{k+1}} (x_k - x_{k-1}).
\end{align*}
\]

The convergence of Algorithm 2.7 is proved by the following theorem.
Theorem 2.8 Let \( f \in \Gamma(\mathcal{H}) \) be an \( L_f \)-Lipschitz continuous function, let \( g \in \Gamma(\mathcal{G}) \) be an \( L_g \)-Lipschitz continuous function, let \( K : \mathcal{H} \to \mathcal{G} \) be a bounded linear operator, and let \( x^* \in \mathcal{H} \) be an optimal solution to (P). Then, when choosing
\[
\rho_k = \frac{1}{a_k} \quad \text{and} \quad \mu_k = \frac{1}{b_k} \quad \forall k \geq 1,
\]
where \( a, b \in \mathbb{R}_{++} \), Algorithm 2.7 generates a sequence \((x_k)_{k \geq 1} \subseteq \mathcal{H}\) satisfying
\[
F(x_{k+1}) - F(x^*) \leq \frac{2(a + b \|K\|^2)}{k + 2} \|x_0 - x^*\|^2 + \frac{2(1 + \ln(k+1))}{k} \left( \frac{L_f^2}{a} + \frac{L_g^2}{b} \right)
\]
for all \( k \geq 1 \), thus yielding a rate of convergence for the objective of order \( \mathcal{O}\left(\frac{\ln k}{k}\right) \).

Proof. For any \( k \geq 1 \), we let \( F^k := F^{p_k, \mu_k} \), \( p_k := (t_k - 1)(x_{k-1} - x_k) \), and
\[
\xi_k := \nabla F^k(y_k) = \text{Prox}_{\frac{1}{\rho_k} f^*} \left( \frac{y_k}{\rho_k} \right) + K^* \text{Prox}_{\frac{1}{\mu_k} g^*} \left( \frac{Ky_k}{\mu_k} \right).
\]
For any \( k \geq 1 \), it holds
\[
p_{k+1} - x_{k+1} = (t_{k+1} - 1)(x_k - x_{k+1}) - x_{k+1}
= (t_{k+1} - 1)x_k - t_{k+1} \left( y_{k+1} - \frac{1}{L_{k+1}} \nabla F^{k+1}(y_{k+1}) \right)
= p_k - x_k + \frac{t_{k+1}}{L_{k+1}} \nabla F^{k+1}(y_{k+1}),
\]
and from here, it follows
\[
\|p_{k+1} - x_{k+1} + x^*\|^2
= \|p_k - x_k + x^*\|^2 + 2 \left( p_k - x_k + x^*, \frac{t_{k+1}}{L_{k+1}} \xi_{k+1} \right) + \left( \frac{t_{k+1}}{L_{k+1}} \right)^2 \|\xi_{k+1}\|^2
= \|p_k - x_k + x^*\|^2 + \frac{2t_{k+1}}{L_{k+1}} \langle p_k, \xi_{k+1} \rangle + \frac{2t_{k+1}}{L_{k+1}} \left( \frac{t_{k+1}}{L_{k+1}} \right)^2 \|\xi_{k+1}\|^2
= \|p_k - x_k + x^*\|^2 + \frac{2(t_{k+1} - 1)}{L_{k+1}} \langle p_k, \xi_{k+1} \rangle + \frac{2t_{k+1}}{L_{k+1}} \langle x^* - y_{k+1}, \xi_{k+1} \rangle + \left( \frac{t_{k+1}}{L_{k+1}} \right)^2 \|\xi_{k+1}\|^2.
\]
Further, using (1.16), since \( x_{k+1} = y_{k+1} - \frac{1}{L_{k+1}} \xi_{k+1} \), it follows
\[
F^{k+1}(x_{k+1}) \leq F^{k+1}(y_{k+1}) + \langle \xi_{k+1}, x_{k+1} - y_{k+1} \rangle + \frac{L_{k+1}}{2} \|x_{k+1} - y_{k+1}\|^2
= F^{k+1}(y_{k+1}) - \frac{1}{L_{k+1}} \|\xi_{k+1}\|^2 + \frac{1}{2L_{k+1}} \|\xi_{k+1}\|^2
= F^{k+1}(y_{k+1}) - \frac{1}{2L_{k+1}} \|\xi_{k+1}\|^2.
\]
and, from here, by making use of the convexity of $F^{k+1}$, we have

$$
\langle x^* - y_{k+1}, \xi_{k+1} \rangle \leq F^{k+1}(x^*) - F^{k+1}(y_{k+1})
$$

(2.57)

$$
\leq F^{k+1}(x^*) - F^{k+1}(x_{k+1}) - \frac{1}{2L_{k+1}} \| \xi_{k+1} \|^2 \quad \forall k \geq 1.
$$

(2.58)

On the other hand, since $F^{k+1}(x_k) - F^{k+1}(y_{k+1}) \geq \langle \xi_{k+1}, x_k - y_{k+1} \rangle$, we obtain

$$
\| \xi_{k+1} \|^2 \leq 2L_{k+1}(F^{k+1}(y_{k+1}) - F^{k+1}(x_{k+1}))
$$

(2.57)

$$
\leq 2L_{k+1} \left( F^{k+1}(x_k) - F^{k+1}(x_{k+1}) - \frac{1}{t_{k+1}} \langle \xi_{k+1}, p_k \rangle \right) \quad \forall k \geq 1.
$$

(2.59)

Thus, as $t_{k+1}^2 - t_{k+1} = t_k^2$, and by making use of (2.53), for any $k \geq 1$, it yields

$$
\| p_{k+1} - x_{k+1} + x^* \|^2 - \| p_k - x_k + x^* \|^2
$$

(2.58)

$$
\leq \frac{2(t_{k+1} - 1)}{L_{k+1}} \langle p_k, \xi_{k+1} \rangle + \frac{2t_{k+1}}{L_{k+1}} (F^{k+1}(x^*) - F^{k+1}(x_{k+1})) + \frac{t_{k+1}^2 - t_{k+1}}{L_{k+1}^2} \| \xi_{k+1} \|^2
$$

(2.59)

$$
\leq \frac{2t_{k+1}^2}{L_{k+1}} (F^{k+1}(x^*) - F^{k+1}(x_{k+1})) + \frac{2(t_{k+1}^2 - t_{k+1})}{L_{k+1}} (F^{k+1}(x_k) - F^{k+1}(x_{k+1}))
$$

(2.53)

$$
= \frac{2t_{k+1}^2}{L_{k+1}} (F^k(x_k) - F^k(x^*) + \rho_k \frac{L_f^2}{2} + \mu_k \frac{L_g^2}{2}) - \frac{2t_{k+1}^2}{L_{k+1}} (F^{k+1}(x_{k+1}) - F^{k+1}(x^*))
$$

$$
- \frac{2t_{k+1}^2}{L_{k+1}} \left( \rho_{k+1} \frac{L_f^2}{2} + \mu_{k+1} \frac{L_g^2}{2} \right).
$$

By using (2.54), it follows that for any $k \geq 1$

$$
F^k(x_k) - F^k(x^*) + \rho_k \frac{L_f^2}{2} + \mu_k \frac{L_g^2}{2} \geq F(x_k) - F^k(x^*) \geq F(x_k) - F(x^*) \geq 0,
$$

thus

$$
\| p_{k+1} - x_{k+1} + x^* \|^2 - \| p_k - x_k + x^* \|^2
$$

$$
\leq \frac{2t_{k+1}^2}{L_k} \left( F^k(x_k) - F^k(x^*) + \rho_k \frac{L_f^2}{2} + \mu_k \frac{L_g^2}{2} \right) - \frac{2t_{k+1}^2}{L_{k+1}} (F^{k+1}(x_{k+1}) - F^{k+1}(x^*))
$$

$$
- \frac{2t_{k+1}^2}{L_{k+1}} \left( \rho_{k+1} \frac{L_f^2}{2} + \mu_{k+1} \frac{L_g^2}{2} \right)
$$

$$
= \frac{2t_{k+1}^2}{L_k} \left( F^k(x_k) - F^k(x^*) + \rho_k \frac{L_f^2}{2} + \mu_k \frac{L_g^2}{2} \right) - \frac{2t_{k+1}^2}{L_{k+1}} (F^{k+1}(x_{k+1}) - F^{k+1}(x^*))
$$

$$
- \frac{2t_{k+1}^2}{L_{k+1}} \left( \rho_{k+1} \frac{L_f^2}{2} + \mu_{k+1} \frac{L_g^2}{2} \right) + \frac{2t_{k+1}}{L_{k+1}} \left( \rho_k \frac{L_f^2}{2} + \mu_k \frac{L_g^2}{2} \right),
$$
which implies that
\[
\|p_{k+1} - x_{k+1} + x^*\|^2 \leq \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) + \rho_{k+1} \frac{L_j^2}{2} + \mu_{k+1} \frac{L_g^2}{2} \right) + \|x_{k+1} - x^*\|^2
\]

Furthermore, since
\[
\|p_k - x_k + x^*\|^2 \leq \frac{2t_k^2}{L_k} \left( F^k(x_k) - F^k(x^*) + \rho_k \frac{L_j^2}{2} + \mu_k \frac{L_g^2}{2} \right) + \frac{2t_{k+1}}{L_{k+1}} \left( \rho_{k+1} \frac{L_j^2}{2} + \mu_{k+1} \frac{L_g^2}{2} \right).
\]

Making again use of (2.54), this further yields for any \( k \geq 1 \)
\[
\frac{2t_{k+1}^2}{L_{k+1}} (F(x_{k+1}) - F(x^*)) \\
\leq \frac{2t_k^2}{L_k} \left( F^k(x_k) - F^k(x^*) + \rho_k \frac{L_j^2}{2} + \mu_k \frac{L_g^2}{2} \right) + \|x_{k+1} - x^*\|^2 \\
\leq \frac{2t_k^2}{L_k} \left( F^1(x_1) - F^1(x^*) + \rho_1 \frac{L_j^2}{2} + \mu_1 \frac{L_g^2}{2} \right) + \|x_1 - x^*\|^2 \\
+ \frac{k}{L_{s+1}} \left( \rho_{s+1} \frac{L_j^2}{2} + \mu_{s+1} \frac{L_g^2}{2} \right). \tag{2.60}
\]

Since \( x_1 = y_1 - \frac{1}{L_1} \nabla F^1(y_1) \), and
\[
F^1(x^*) = F^1(y_1) + \left< \nabla F^1(y_1), x^* - y_1 \right>, \\
F^1(x_1) \leq F^1(y_1) + \left< \nabla F^1(y_1), x_1 - y_1 \right> + \frac{L_1}{2} \|x_1 - y_1\|^2,
\]
we get
\[
\frac{2t_k^2}{L_k} \left( F^1(x_1) - F^1(x^*) \right) + \|x_1 - x^*\|^2 \\
\leq 2 \langle x_1 - y_1, x^* - y_1 \rangle - \|x_1 - y_1\|^2 + \|x_1 - x^*\|^2 = \|y_1 - x^*\|^2 = \|x_0 - x^*\|^2;
\]
and this, together with (2.60), gives rise to the following estimate
\[
\frac{2t_{k+1}^2}{L_{k+1}} (F(x_{k+1}) - F(x^*)) \leq \|x_0 - x^*\|^2 + \sum_{s=1}^{k+1} \frac{t_s}{L_s} \left( \rho_s L_j^2 + \mu_s L_g^2 \right). \tag{2.61}
\]

Furthermore, since \( t_{k+1} \geq \frac{1}{2} + t_k \) for any \( k \geq 1 \), it follows that \( t_{k+1} \geq \frac{k+2}{2} \), which, along with the fact that \( L_k = \frac{1}{\rho_k} + \frac{\|K\|^2}{\mu_k} = (a + b \|K\|^2)k \), leads for any \( k \geq 1 \) to the following estimate
\[
F(x_{k+1}) - F(x^*) \\
\leq \frac{2(a + b \|K\|^2)(k + 1)}{(k + 2)^2} \left( \|x_0 - x^*\|^2 + L_j^2 \sum_{s=1}^{k+1} \frac{t_s \rho_s}{L_s} + L_g^2 \sum_{s=1}^{k+1} \frac{t_s \mu_s}{L_s} \right) \\
\leq \frac{2(a + b \|K\|^2)}{k + 2} \|x_0 - x^*\|^2 + \frac{2}{k + 2} \sum_{s=1}^{k+1} \frac{t_s}{s^2} \left( \frac{L_j^2}{a} + \frac{L_g^2}{b} \right).
\]
Using now that \( t_{k+1} \leq 1 + t_k \) for any \( k \geq 1 \), it yields that \( t_{k+1} \leq k+1 \) for any \( k \geq 0 \), thus

\[
\sum_{s=1}^{k} t_{s} s^2 \leq \sum_{s=1}^{k+1} s \leq 1 + \sum_{s=2}^{k+1} \int_{s-1}^{s} \frac{1}{x} \, dx = 1 + \int_{1}^{k+1} \frac{1}{x} \, dx = 1 + \ln(k+1).
\]

Finally, we obtain that

\[
F(x_{k+1}) - F(x^*) \leq \frac{2(a + b \|K\|^2)}{k+2} \|x_0 - x^*\|^2 + \frac{2(1 + \ln(k+1))}{k+2} \left( \frac{L_f^2}{a} + \frac{L_g^2}{b} \right)
\]

for all \( k \geq 1 \), which concludes the proof.

In the second part of this subsection, we propose a variant of Algorithm 2.7 formulated with constant smoothing parameters.

**Algorithm 2.9** Let \( y_1 = x_0 \in H, \rho, \mu \in \mathbb{R}^+ \), let \( t_1 = 1 \), \( L(\rho, \mu) = \frac{1}{\rho} + \frac{\|K\|^2}{\mu} \), and set

\[
(\forall k \geq 1) \quad \begin{cases}
 x_k = y_k - \frac{1}{L(\rho, \mu)} \left( \text{Prox}_{\frac{\rho}{\mu}} f^* \left( \frac{y_k}{\rho} \right) + K^* \text{Prox}_{\frac{\mu}{\rho}} g^* \left( \frac{K^* y_k}{\mu} \right) \right), \\
 t_{k+1} = \frac{1 + \sqrt{1 + 4k^2}}{2}, \\
 y_{k+1} = x_k + \frac{t_k - 1}{t_{k+1}} (x_k - x_{k-1}).
\end{cases}
\]

(2.62)

Constant smoothing parameters have also been used in Section 2.1 within the framework of double smoothing algorithms, which assume the regularization of the Fenchel dual problem to \((P)\) in two steps.

In consideration of the sequence \((x_k)_{k \geq 1} \subseteq H\) from Algorithm 2.9, the following statement establishes the convergence of the sequence of primal objective values to \(v(P)\).

**Theorem 2.10** Let \( f \in \Gamma(H) \) be an \( L_f\)-Lipschitz continuous function, let \( g \in \Gamma(G) \) be an \( L_g\)-Lipschitz continuous function, let \( K : H \to G \) be a bounded linear operator, and let \( x^* \in H \) be an optimal solution to \((P)\). Then, when choosing for \( \varepsilon > 0 \)

\[
\rho = \frac{2\varepsilon}{3L_f^2}, \text{ and } \mu = \frac{2\varepsilon}{3L_g^2},
\]

**Algorithm 2.9** generates a sequence \((x_k)_{k \geq 1} \subseteq H\) which provides an \( \varepsilon\)-optimal solution to \((P)\) with a rate of convergence for the objective of order \( \mathcal{O}(\frac{1}{k}) \).

Proof. In order to prove this statement, one only has to reproduce the first part of the proof of Theorem 2.8 when

\[
\rho_k = \rho, \mu_k = \mu, \text{ and } L_k = L(\rho, \mu) = \frac{1}{\rho} + \frac{\|K\|^2}{\mu} \quad \forall k \geq 1,
\]

some fact which leads to (2.61). This inequality reads in this particular situation

\[
F(x_{k+1}) - F(x^*) \leq \frac{L(\rho, \mu) \|x_0 - x^*\|^2}{2t_{k+1}^2} + \frac{\rho L_f^2}{2t_{k+1}^2} \sum_{s=1}^{k+1} t_s \quad \forall k \geq 1.
\]
Since \( t_{k+1}^2 = t_k^2 + t_{k+1} \) for any \( k \geq 1 \), one can inductively prove that \( t_{k+1}^2 = \sum_{s=1}^{k+1} t_s \), which, together with the fact that \( t_{k+1} \geq \frac{k+2}{2} \) for any \( k \geq 1 \), yields
\[
F(x_{k+1}) - F(x^*) \leq \frac{2L(\rho, \mu) \|x_0 - x^*\|^2}{(k+2)^2} + \frac{\rho L_f^2 + \mu L_g^2}{2} \forall k \geq 1.
\]
In order to obtain \( \varepsilon \)-optimality for the objective of the problem (P), where \( \varepsilon > 0 \) is a given level of accuracy, we choose \( \rho = \frac{2\varepsilon}{3L_f^2} \) and \( \mu = \frac{2\varepsilon}{3L_g^2} \), and, thus, we only have to force the first term in the right-hand side of the above estimate to be less than or equal to \( \frac{\varepsilon}{3} \). Taking also into account that in this situation \( L(\rho, \mu) = \frac{3L_f^2 + 3L_g^2 \|K\|^2}{2\varepsilon} \), it holds
\[
\frac{\varepsilon}{3} \geq \frac{2L(\rho, \mu) \|x_0 - x^*\|^2}{(k+2)^2} = \frac{3 \left(L_f^2 + L_g^2 \|K\|^2 \right) \|x_0 - x^*\|^2}{\varepsilon (k+2)^2}
\]
\[
\iff \frac{\varepsilon^2}{9} \geq \frac{\left(L_f^2 + L_g^2 \|K\|^2 \right) \|x_0 - x^*\|^2}{(k+2)^2}
\]
\[
\iff \frac{\varepsilon}{3} \geq \frac{\sqrt{L_f^2 + L_g^2 \|K\|^2} \|x_0 - x^*\|}{k+2},
\]
which shows that an \( \varepsilon \)-optimal solution to (P) can be provided with a rate of convergence for the objective of order \( \mathcal{O}(\frac{1}{k}) \).

The rate of convergence of Algorithm 2.7 may not be as good as the one proved for the algorithm with constant smoothing parameters depending on a fixed level of accuracy \( \varepsilon > 0 \). However, the main advantage of the variable smoothing method is given by the fact that the sequence of objective values \( (f(x_k) + g(Kx_k))_{k \geq 1} \) converges to the optimal objective value of (P), whereas, when generated by Algorithm 2.9, despite of the fact that it approximates the optimal objective value with a better convergence rate, this sequence may not converge to the optimal objective value. This behavior is illustrated in the following example.

**Example 2.11** Let us consider the following convex problem. Here we want to solve
\[
(P_{\|\|}) \quad \inf_{x \in \mathbb{R}} \left\{ \|x - 2\| + \frac{1}{2} |x| \right\},
\]
which fits into the framework considered in (2.50) by letting \( f : \mathbb{R} \to \mathbb{R}, f(x) = |x - 2|, g : \mathbb{R} \to \mathbb{R}, g(x) = \frac{1}{2} |x| \), and \( K = \text{Id} \). Our aim in this example is to show the lack of convergence with respect to the objective values when using the constant smoothing approach. The objective function together with its unique optimal solution \( x = 2 \) and its optimal objective value \( v(P_{\|\|}) = 1 \) is shown in Figure 2.1. Both \( f \) and \( g \) are nondifferentiable convex functions and therefore we apply the smoothing techniques introduced in Subsection 2.2.2 for some \( \rho, \mu \in \mathbb{R}_{+} \). Notice that \( f^* : \mathbb{R} \to \mathbb{R}, f^*(p) = \delta_{[-1,1]}(p) + 2p \), and that \( g^* : \mathbb{R} \to \mathbb{R}, g^*(p) = \delta_{[-\frac{1}{2}, \frac{1}{2}]}(p) \). In view of this, we obtain
\[
\rho f(x) = \left(f^* + \frac{\rho}{2} \|\cdot\|^2\right)^*(x) = \sup_{z \in \mathbb{R}} \left\{ xz - f^*(z) - \frac{\rho}{2} z^2 \right\} = \sup_{z \in [-1,1]} \left\{ xz - 2z - \frac{\rho}{2} z^2 \right\}
\]
\[
= \begin{cases} 
- x + 2 - \frac{\rho}{2}, & \text{if } x < 2 - \rho \\
\frac{(x-2)^2}{2\rho}, & \text{if } x \in [2 - \rho, 2 + \rho] \\
 x - 2 - \frac{\rho}{2}, & \text{if } x > 2 + \rho 
\end{cases}
\]
and
\[
\mu g(x) = \left(g^* + \frac{\mu}{2} \| \cdot \|^2 \right)^*(x) = \sup_{z \in \mathbb{R}} \left\{ xz - g^*(z) - \frac{\mu}{2} z^2 \right\} = \sup_{z \in [-\frac{1}{2}, \frac{1}{2}]} \left\{ xz - \frac{\mu}{2} z^2 \right\}
\]
\[
= \begin{cases} 
-\frac{x^2}{2} - \frac{\mu}{8}, & \text{if } x < -\frac{\mu}{2} \\
\frac{x^2}{2} - \frac{\mu}{8}, & \text{if } x \in [-\frac{\mu}{2}, \frac{\mu}{2}] \\
\frac{x^2}{2} - \frac{\mu}{8}, & \text{if } x > \frac{\mu}{2}
\end{cases}
\]

Now, by setting the gradient to zero and by letting the constant smoothing parameters be from the interval \((0, 1)\), the smoothed problem
\[
\inf_{x \in \mathbb{R}} \left\{ \rho f(x) + \mu g(x) \right\}
\]
shows to have the unique optimal solution \(x_{\rho, \mu} = 2 - \frac{\rho}{2}\). Therefore, due to Nesterov (cf. [97]), if we define \(F^{\rho, \mu}(x) := \rho f(x) + \mu g(x)\), Algorithm 2.9 generates a sequence \((x_k)_{k \geq 1}\), such that \(F^{\rho, \mu}(x_k) - F^{\rho, \mu}(x_{\rho, \mu}) \leq O\left(\frac{1}{k^2}\right)\). By making use of the coercivity of \(F^{\rho, \mu}\) and the uniqueness of \(x_{\rho, \mu}\), it follows that every subsequence of \((x_k)_{k \geq 1}\) is convergent and converges to \(x_{\rho, \mu}\), which therefore yields \(x_k \to x_{\rho, \mu}\). However, in consideration of the original objective function, we have \(f(x_k) + g(x_k) \to f(x_{\rho, \mu}) + g(x_{\rho, \mu})\), where
\[
f(x_{\rho, \mu}) + g(x_{\rho, \mu}) = f\left(2 - \frac{\rho}{2}\right) + g\left(2 - \frac{\rho}{2}\right) = \left| -\frac{\rho}{2} \right| + \frac{1}{2} \left| 2 - \frac{\rho}{2} \right| = 1 + \frac{\rho}{4} \neq v(P_{|\cdot|}).
\]

Therefore, we have shown that the sequence \((f(x_k) + g(x_k))_{k \geq 1}\) generated by the constant smoothing algorithm does not converge to the optimal objective value of \((P_{|\cdot|})\). Since \(\rho = O(\varepsilon)\) and \(\mu = O(\varepsilon)\), this algorithm only provides solutions with \(\varepsilon\)-accuracy.

### 2.2.4 The case when \(f\) is differentiable with Lipschitz continuous gradient

In this subsection, for solving the problem \((P)\), we show how Algorithm 2.7 and Algorithm 2.9 can be adapted to the situation when \(f\) is a differentiable function with Lipschitz continuous gradient. We provide iterative schemes with variable and constant smoothing variables and corresponding convergence statements. More
precisely, we deal with the optimization problem

\[
(P) \quad \inf_{x \in \mathcal{H}} \{ f(x) + g(Kx) \},
\]

where \( \mathcal{H} \) and \( \mathcal{G} \) are real Hilbert spaces, \( K : \mathcal{H} \to \mathcal{G} \) is a bounded linear operator, \( f \in \Gamma(\mathcal{H}) \) is a differentiable function with \( L_\nabla f \)-Lipschitz continuous gradient, and \( g \in \Gamma(\mathcal{G}) \) is an \( L_g \)-Lipschitz continuous function.

Algorithm 2.12 can be adapted to this framework as follows.

**Algorithm 2.12** Let \( y_1 = x_0 \in \mathcal{H} \), \((\mu_k)_{k \geq 1} \subseteq \mathbb{R}^+\), let \( t_1 = 1 \), and set

\[
(\forall k \geq 1) \quad \begin{cases}
L_k = L_\nabla f + \frac{\|K\|^2}{\mu_k}, \\
x_k = y_k - \frac{1}{L_k} \nabla f(y_k) + K^* \text{Prox}_{\frac{1}{\mu_k} g^*} \left( \frac{K y_k}{\mu_k} \right), \\
t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \\
y_{k+1} = x_k + \frac{t_{k+1}}{t_{k+1}} (x_k - x_{k-1}).
\end{cases}
\] (2.63)

The convergence of the sequence \((x_k)_{k \geq 1} \subseteq \mathcal{H}\) in Algorithm 2.12 with respect to the primal objective values is furnished by the following theorem.

**Theorem 2.13** Let \( f \in \Gamma(\mathcal{H}) \) be a differentiable function with \( L_\nabla f \)-Lipschitz continuous gradient, let \( g \in \Gamma(\mathcal{G}) \) be an \( L_g \)-Lipschitz continuous function, let \( K : \mathcal{H} \to \mathcal{G} \) be a bounded linear operator, and let \( x^* \in \mathcal{H} \) be an optimal solution to \((P)\). Then, when choosing

\[
\mu_k = \frac{1}{b_k} \quad \forall k \geq 1,
\]

where \( b \in \mathbb{R}^+ \), Algorithm 2.12 generates a sequence \((x_k)_{k \geq 1} \subseteq \mathcal{H}\) satisfying for any \( k \geq 1 \)

\[
F(x_{k+1}) - F(x^*) \leq \frac{2(L_\nabla f + b \|K\|^2)}{k + 2} \|x_0 - x^*\|^2 + \frac{2(1+\ln(k+1))}{k + 2} \frac{L_\nabla^2 f + b \|K\|^2}{b^2 \|K\|^2},
\] (2.64)

thus yielding a rate of convergence for the objective of order \( O\left( \frac{\ln k}{k} \right) \).

Proof. For any \( k \geq 1 \), we let \( F^k : \mathcal{H} \to \mathbb{R} \), \( F^k(x) = f(x) + \mu_k g(Kx) \). For any \( k \geq 1 \) and every \( x \in \mathcal{H} \), it holds \( \nabla F^k(x) = \nabla f(x) + K^* \text{Prox}_{\frac{1}{\mu_k} g^*} \left( \frac{K x}{\mu_k} \right) \) and \( \nabla F^k \) is \( L_k \)-Lipschitz continuous, where \( L_k = L_\nabla f + \frac{\|K\|^2}{\mu_k} \).

As in the proof of Theorem 2.8, by defining \( p_k := (t_k - 1)(x_{k-1} - x_k) \), we obtain for any \( k \geq 1 \)

\[
\|p_{k+1} - x_{k+1} + x^*\|^2 - \|p_k - x_k + x^*\|^2 \\
\leq \frac{2t_k^2}{L_{k+1}} \left( F^{k+1}(x_k) - F^{k+1}(x^*) \right) - \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) \\
\leq \frac{2t_k^2}{L_{k+1}} \left( F^k(x_k) - F^{k+1}(x^*) + (\mu_k - \mu_{k+1}) \frac{L_g^2}{2} \right) - \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) \\
\leq \frac{2t_k^2}{L_{k+1}} \left( F^k(x_k) - F^{k}(x^*) + \mu_k \frac{L_g^2}{2} \right) - \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) - \frac{t_k^2}{L_{k+1}} \mu_{k+1} L_g^2 \\
\leq \frac{2t_k^2}{L_{k+1}} \frac{L_g^2}{2} \left( F^k(x_k) - F^{k+1}(x^*) \right) - \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) - \frac{t_k^2}{L_{k+1}} \mu_{k+1} L_g^2.
\]
\[
\begin{align*}
&\leq \frac{2t_k^2}{L_k} \left( F^k(x_k) - F^k(x^*) + \mu_k \frac{L_g^2}{2} \right) - \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) - \frac{t_{k+1}^2}{L_{k+1}} \mu_{k+1} L_g^2 \\
&= \frac{2t_k^2}{L_k} \left( F^k(x_k) - F^k(x^*) + \mu_k \frac{L_g^2}{2} \right) - \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) \\
&- \frac{t_{k+1}^2 L_g^2}{L_{k+1}} \mu_{k+1} + \frac{t_{k+1} L_g^2}{L_{k+1}} \mu_{k+1},
\end{align*}
\]
and, consequently,
\[
\|p_{k+1} - x_{k+1} + x^*\|^2 \leq \|p_k - x_k + x^*\|^2 + \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) \leq \|p_k - x_k + x^*\|^2 + \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) + \frac{t_{k+1} L_g^2}{L_{k+1}} \mu_{k+1}.
\]

For any \( k \geq 1 \), it holds
\[
\frac{2t_{k+1}^2}{L_{k+1}} \left( F(x_{k+1}) - F(x^*) \right) \leq \frac{2t_{k+1}^2}{L_{k+1}} \left( F^{k+1}(x_{k+1}) - F^{k+1}(x^*) \right) + \|p_{k+1} - x_{k+1} + x^*\|^2
\]
\[
\leq \frac{2t_{k+1}^2}{L_{k+1}} \left( F(x_{k+1}) - F(x^*) \right) + \frac{t_{k+1} L_g^2}{L_{k+1}} \mu_{k+1},
\]
which yields
\[
\frac{2t_{k+1}^2}{L_{k+1}} \left( F(x_{k+1}) - F(x^*) \right) \leq \|x_0 - x^*\|^2 + \sum_{s=1}^{k+1} \frac{t_s L_g^2}{L_s} \mu_s.
\]
(2.65)

For any \( k \geq 1 \), since \( t_{k+1} \geq \frac{k+2}{2} \) and \( L_k = L\sqrt{f} + \frac{\|K\|^2}{\mu_k} = L\sqrt{f} + b \|K\|^2 k \), it follows
\[
F(x_{k+1}) - F(x^*) \leq \frac{2(L\sqrt{f} + b \|K\|^2 (k + 1))}{(k + 2)^2} \left( \|x_0 - x^*\|^2 + \sum_{s=1}^{k+1} \frac{t_s L_g^2}{L_s} (L\sqrt{f} + b \|K\|^2 s) b \right).
\]

Thus, for any \( k \geq 1 \), since \( t_k \leq k \), it yields
\[
F(x_{k+1}) - F(x^*) \leq \frac{2(L\sqrt{f} + b \|K\|^2 (k + 1))}{(k + 2)^2} \left( \|x_0 - x^*\|^2 + \sum_{s=1}^{k+1} \frac{L_g^2}{L_s} (L\sqrt{f} + b \|K\|^2 s) b \right)
\]
\[
\leq \frac{2(L\sqrt{f} + b \|K\|^2 (k + 1))}{(k + 2)^2} \left( \|x_0 - x^*\|^2 + \frac{k+1}{s=1} \frac{L_g^2}{b^2 \|K\|^2 s} (1 + \ln(k + 1)) \right)
\]
\[
\leq \frac{2(L\sqrt{f} + b \|K\|^2)}{k + 2} \left( \|x_0 - x^*\|^2 + \frac{2}{k + 2} \frac{2 + 2(k + 1) \ln(k + 1)}{b^2 \|K\|^2} \right).
\]
By adapting Algorithm 2.12 to the framework considered in this subsection, we obtain the following algorithm with constant smoothing variables.

**Algorithm 2.14** Let \( y_1 = x_0 \in \mathcal{H}, \mu \in \mathbb{R}_{++}, \) let \( t_1 = 1, \) \( L(\mu) = L\nabla f + \frac{\|K\|^2}{\mu}, \) and set

\[
(\forall k \geq 1) \quad \begin{align*}
x_k &= y_k - \frac{1}{L(\mu)} \left( \nabla f(y_k) + K^* \text{Prox}_{\mu \frac{1}{\mu}} \left( \frac{K y_k}{\mu} \right) \right), \\
t_{k+1} &= \frac{1}{\sqrt{1+4t^2}}, \\
y_{k+1} &= x_k + \frac{t_k-1}{t_{k+1}}(x_k - x_{k-1}).
\end{align*}
\]

(2.66)

The convergence of Algorithm 2.14 is stated by the following theorem, which can be proved in the lines of the proof of Theorem 2.13.

**Theorem 2.15** Let \( f \in \Gamma(\mathcal{H}) \) be a differentiable function with \( L\nabla f \)-Lipschitz continuous gradient, let \( g \in \Gamma(\mathcal{G}) \) be an \( L_g \)-Lipschitz continuous function, let \( K : \mathcal{H} \rightarrow \mathcal{G} \) be a bounded linear operator, and let \( x^* \in \mathcal{H} \) be an optimal solution to \((P)\). Then, when choosing for \( \varepsilon > 0 \)

\[
\mu = \frac{\varepsilon}{L^2_g}
\]

Algorithm 2.14 generates a sequence \((x_k)_{k \geq 1} \subseteq \mathcal{H}\) which provides an \( \varepsilon \)-optimal solution to \((P)\) with a rate of convergence for the objective of order \( \mathcal{O}(\frac{1}{k}) \).

### 2.2.5 The optimization problem with the sum of more than two functions in the objective

We close this section by discussing the employment of the algorithmic schemes presented in the previous two subsections to the optimization problem (2.52), i.e., to

\[
\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(K_i x) \right\},
\]

where \( \mathcal{H} \) and \( \mathcal{G}_i, i = 1, \ldots, m, \) are real Hilbert spaces, \( f : \mathcal{H} \rightarrow \mathbb{R} \) is a convex and either \( L_f \)-Lipschitz continuous or differentiable function with \( L\nabla f \)-continuous gradient, \( g_i : \mathcal{G}_i \rightarrow \mathbb{R} \) is convex and \( L_{g_i} \)-Lipschitz continuous, and \( K_i : \mathcal{H} \rightarrow \mathcal{G}_i \), is a bounded linear operator for each \( i = 1, \ldots, m \). We then endow the Hilbert space \( \mathcal{G} := \mathcal{G}_1 \times \ldots \times \mathcal{G}_m \) with the inner product defined as

\[
\langle y, z \rangle_{\mathcal{G}} = \sum_{i=1}^{m} \langle y_i, z_i \rangle_{\mathcal{G}_i}, \forall y, z \in \mathcal{G},
\]

and with the corresponding norm. Further, by defining \( g : \mathcal{G} \rightarrow \mathbb{R}, g(y_1, \ldots, y_m) = \sum_{i=1}^{m} g_i(y_i), \) and \( K : \mathcal{H} \rightarrow \mathcal{G}, K x = (K_1 x, \ldots, K_m x), \) problem (2.52) can equivalently be written as

\[
\inf_{x \in \mathcal{H}} \left\{ f(x) + g(K x) \right\},
\]

and, consequently, solved via one of the variable or constant smoothing algorithms introduced in Subsection 2.2.3 and Subsection 2.2.4, depending on the properties the function \( f \) is endowed with.
In the following, we determine the elements related to the above constructed function \( g \) which appear in these iterative schemes and in the corresponding convergence statements. Obviously, the function \( g \) is convex and, since for every \((y_1, \ldots, y_m), (z_1, \ldots, z_m) \in \mathcal{G}\), by making use of the Cauchy–Schwarz inequality,

\[
|g(y_1, \ldots, y_m) - g(z_1, \ldots, z_m)| \leq \sum_{i=1}^{m} L_{g_i} \|y_i - z_i\| \leq \left( \sum_{i=1}^{m} L_{g_i}^2 \right)^{\frac{1}{2}} \|(y_1, \ldots, y_m) - (z_1, \ldots, z_m)\|,
\]

it is \( \left( \sum_{i=1}^{m} L_{g_i}^2 \right)^{\frac{1}{2}} \)-Lipschitz continuous. On the other hand, for each \( \mu \in \mathbb{R}_{++} \) and \((y_1, \ldots, y_m) \in \mathcal{G}\), it holds

\[
\mu g(y_1, \ldots, y_m) = \sum_{i=1}^{m} \mu g_i(y_i),
\]

thus

\[
\nabla (\mu g)(y_1, \ldots, y_m) = (\nabla (\mu g_1)(y_1), \ldots, \nabla (\mu g_m)(y_m))
= \left( \text{Prox}_{\frac{1}{\mu} g_1^*} \left( \frac{y_1}{\mu} \right), \ldots, \text{Prox}_{\frac{1}{\mu} g_m^*} \left( \frac{y_m}{\mu} \right) \right).
\]

Since \( K^*(y_1, \ldots, y_m) = \sum_{i=1}^{m} K_i^* y_i \) for every \((y_1, \ldots, y_m) \in \mathcal{G}\), we have

\[
\nabla (\mu g \circ K)(x) = K^* \nabla (\mu g)(K_1 x, \ldots, K_m x) = \sum_{i=1}^{m} K_i^* \nabla (\mu g_i)(K_i x)
= \sum_{i=1}^{m} K_i^* \text{Prox}_{\frac{1}{\mu} g_i^*} \left( \frac{K_i x}{\mu} \right) \forall x \in \mathcal{H}.
\]

Finally, we notice that for arbitrary \( x, y \in \mathcal{H} \), one has

\[
\|\nabla (\mu g \circ K)(x) - \nabla (\mu g \circ K)(y)\| = \left\| \sum_{i=1}^{m} K_i^* \nabla (\mu g_i)(K_i x) - \sum_{i=1}^{m} K_i^* \nabla (\mu g_i)(K_i y) \right\|
\leq \sum_{i=1}^{m} K_i \| \nabla (\mu g_i)(K_i x) - \nabla (\mu g_i)(K_i y) \|
\leq \sum_{i=1}^{m} \frac{\|K_i\|}{\mu} \|K_i x - K_i y\| \leq \sum_{i=1}^{m} \frac{\|K_i\|^2}{\mu} \|x - y\|,
\]

which shows that \( \nabla (\mu g \circ K) \) is \( \sum_{i=1}^{m} \|K_i\|^2 \)-Lipschitz continuous.
Primal-dual algorithms for inclusion problems

In applied mathematics, a wide variety of convex optimization problems such as single- or multifacility location problems, support vector machine problems for classification and regression, problems in clustering and portfolio optimization as well as signal and image processing problems, all of them potentially possessing nonsmooth terms in their objectives, can be reduced to the solving of inclusion problems involving mixtures of monotone set-valued operators.

This chapter is devoted to the employment of so-called primal-dual algorithms for solving structured monotone inclusion problems in real Hilbert spaces. When the maximally monotone operators in the problem formulation correspond to subdifferentials of proper, convex, and lower semicontinuous functions, then one automatically solves a system of optimality conditions for some primal-dual pair of convex optimization problems. However, finding a primal-dual solution can only be guaranteed under the premise that appropriate qualification conditions are fulfilled, some fact which is given a special attention. The interested reader can find our survey paper in [28] for a short overview on primal-dual splitting philosophies.

The problems discussed in this chapter cover and extend a wide class of monotone inclusion problems in the literature (see, for instance, [25, 26, 54, 66, 88, 105, 121, 128]).

3.1 Convergence analysis of a forward-backward-forward method

In this section, and in view of our results given in [39], we investigate the convergence behavior of the primal-dual splitting method due to Combettes and Pesquet described and analyzed in [58] from two different points of view. To this end, in the particular case of solving convex minimization problems, we firstly derive convergence rate estimates for the primal-dual gap of function values restricted to some special bounded set. In the second part of this section, we propose two new schemes which accelerate the sequences of primal and/or dual iterates under the premise that certain operators occurring in the problem formulation are known to be strongly monotone.
When applied to convex optimization problems, the forward-backward-forward approach naturally suffers from its additional forward step which affects the computational performance. Nevertheless, the method is highly parallelizable since lots of its steps can be executed in parallel and it pursues the splitting philosophy which is essential for solving optimization problems having intricate formulations.

### 3.1.1 Problem description

In the following we describe the monotone inclusion problem which we aim to investigate throughout this section (see [58]).

**Problem 3.1** Consider the real Hilbert space $\mathcal{H}$, let $z \in \mathcal{H}$, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, and $C : \mathcal{H} \to \mathcal{H}$ be a monotone and $\mu$-Lipschitzian operator for some $\mu \in \mathbb{R}_{++}$. Furthermore, for every $i = 1, \ldots, m$, consider the real Hilbert space $\mathcal{G}_i$, let $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be a maximally monotone operator, let $D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be a monotone operator such that $D_i^{-1}$ is $\nu_i$-Lipschitzian for some $\nu_i \in \mathbb{R}_{++}$, and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator. The problem is to solve the primal inclusion

\[
\text{find } x \in \mathcal{H} \text{ such that } \quad z = A x + \sum_{i=1}^{m} L_i^* ( (B_i \square D_i)(L_i x - r_i) ) + C x, \quad (3.1)
\]

together with the dual inclusion

\[
\text{find } v_1 \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} 
z - \sum_{i=1}^{m} L_i^* v_i \in A x + C x, \\
v_i \in (B_i \square D_i)(L_i x - r_i), \quad i = 1, \ldots, m.
\end{cases} \quad (3.2)
\]

**Remark 3.2** For $\mathcal{G} = \mathcal{G}_1 \oplus \ldots \oplus \mathcal{G}_m$, we say that $(x, v_1, \ldots, v_m) \in \mathcal{H} \oplus \mathcal{G}$ is a primal-dual solution to Problem 3.1, if

\[
z - \sum_{i=1}^{m} L_i^* v_i \in A x + C x \text{ and } v_i \in (B_i \square D_i)(L_i x - r_i), \quad i = 1, \ldots, m. \quad (3.3)
\]

If $(x, v_1, \ldots, v_m) \in \mathcal{H} \oplus \mathcal{G}$ is a primal-dual solution to Problem 3.1, then $x$ is a solution to (3.1) and $(v_1, \ldots, v_m)$ is a solution to (3.2). Be aware that $x$ solves (3.1) $\iff$ $z - \sum_{i=1}^{m} L_i^* (B_i \square D_i)(L_i x - r_i) \in A x + C x$

$\iff \exists v_1 \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m$ such that

\[
\begin{cases} 
z - \sum_{i=1}^{m} L_i^* v_i \in A x + C x, \\
v_i \in (B_i \square D_i)(L_i x - r_i), \quad i = 1, \ldots, m.
\end{cases}
\]

Thus, if $x$ is a solution to (3.1), then there exists $(v_1, \ldots, v_m) \in \mathcal{G}$ such that $(x, v_1, \ldots, v_m)$ is a primal-dual solution to Problem 3.1, and, if $(v_1, \ldots, v_m)$ is a solution to (3.2), then there exists $x \in \mathcal{H}$ such that $(x, v_1, \ldots, v_m)$ is a primal-dual solution to Problem 3.1.

The next result provides the error-free variant of the primal-dual algorithm in [58] and some of the corresponding convergence statements as given in [58, Theorem 3.1].
Theorem 3.3 For Problem 3.1, suppose that

\[ z \in \text{ran} \left( A + \sum_{i=1}^{m} L_i^* (B_i \square D_i) (L_i \cdot -r_i) + C \right). \]

Let \( x_0 \in \mathcal{H} \) and \((v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}\), set

\[ \beta = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2}, \]

choose \( \varepsilon \in (0, \frac{1}{\beta+1}) \), let \((\gamma_n)_{n \geq 0} \) be a sequence in \( \left[ \varepsilon, \frac{1-\varepsilon}{\beta} \right] \), and set

\[
(\forall n \geq 0) \begin{cases} 
 p_{1,n} = J_{\gamma_n A} (x_n - \gamma_n (Cx_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z)), \\
 For i = 1, \ldots, m \\
 p_{2,i,n} = J_{\gamma_n B_i^{-1}} (v_{i,n} + \gamma_n (L_i x_n - D_i^{-1} v_{i,n} - r_i)), \\
 v_{i,n+1} = \gamma_n L_i (p_{1,n} - x_n) + \gamma_n (D_i^{-1} v_{i,n} - D_i^{-1} p_{2,i,n}) + p_{2,i,n}, \\
 x_{n+1} = \gamma_n \sum_{i=1}^{m} L_i^* (v_{i,n} - p_{2,i,n}) + \gamma_n (Cx_n - C p_{1,n}) + p_{1,n}.
\end{cases}
\]

Then the following statements are true:

(i) \( \sum_{n \in \mathbb{N}} \|x_n - p_{1,n}\|^2 < +\infty \) and \( \sum_{n \in \mathbb{N}} \|v_{i,n} - p_{2,i,n}\|^2 < +\infty \) for all \( i = 1, \ldots, m \),

(ii) There exists a primal-dual solution \((\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{H} \times \mathcal{G}\) to Problem 3.1 such that the following holds:

(a) \( x_n \to \bar{x} \) and \( p_{1,n} \to \bar{x} \),

(b) \( v_{i,n} \to \bar{v}_i \) and \( p_{2,i,n} \to \bar{v}_i \) for all \( i = 1, \ldots, m \).

The investigations we make in this section have as starting point the primal-dual algorithm for solving Problem 3.1 given in Theorem 3.3 above. Firstly, we consider Problem 3.1 in its particular formulation as a primal-dual pair of convex optimization problems, an approach which relies on the fact that the subdifferential of a proper, convex, and lower semicontinuous function is maximally monotone. By assuming that the sequence of step sizes in the algorithm in [58, Theorem 3.1] is nondecreasing and by making use of some ideas provided in [48], we prove that the convergence rate of the partial primal-dual gap function associated to the primal-dual pair of optimization problems at some primal-dual pair of explicitly generated iterates is of order \( O(\frac{1}{n}) \), where \( n \in \mathbb{N} \) is the number of passed iterations. From here we are able to derive under some appropriate assumptions the same rate of convergence for the sequence of primal objective function values on the iterates generated by the numerical scheme.

Further, in Subsection 3.1.3, we provide for the general monotone inclusion problem, as given in Problem 3.1, two new acceleration schemes which, under strong monotonicity assumptions, generate sequences of primal and/or dual iterates that converge with improved convergence properties. To this end, we use the fruitful idea of variable step sizes that have been first utilized in [128] and then shown in [48] to yield an accelerated algorithm in the case of convex optimization problems.
3.1.2 Convergence estimates for convex minimization problems

The aim of this subsection is to provide a rate of convergence for the sequence of objective function values at the iterates generated by a slight modification of the algorithm in [58, Theorem 3.1] when employed for the solving of a convex minimization problem and its conjugate dual. The primal-dual pair under investigation is described in the following.

Problem 3.4 Consider the real Hilbert space $\mathcal{H}$, let $z \in \mathcal{H}$, let $f \in \Gamma(\mathcal{H})$ and $h : \mathcal{H} \to \mathbb{R}$ be a convex and differentiable function with $\mu$-Lipschitzian gradient for some $\mu \in \mathbb{R}_{++}$. Furthermore, for every $i = 1, \ldots, m$, consider the real Hilbert space $\mathcal{G}_i$, let $r_i \in \mathcal{G}_i$, let $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that $l_i$ is $\nu_i^{-1}$-strongly convex for some $\nu_i \in \mathbb{R}_{++}$, and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator. We consider the convex minimization problem

$$
(P) \quad \inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \triangledown l_i)(L_i x - r_i) + h(x) - \langle x, z \rangle \right\} \tag{3.5}
$$

and its dual problem

$$
(D) \quad \sup_{(v_1, \ldots, v_m) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m} \left\{ -(f^* \triangledown h^*) \left( z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} (g_i^* (v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle) \right\}. \tag{3.6}
$$

The formulation in Problem 3.4 captures various types of convex optimization problems. One such particular instance is considered in the following.

Example 3.5 In Problem 3.4, set $m = 1$, $z = 0$ and $r_1 = 0$, let $\mathcal{G}_1 = \mathcal{G}$, $L_1 = L$, and $l_1 : \mathcal{G} \to \mathbb{R}$, $l_1 = \delta_{\{0\}}$. Then, the primal optimization problem (3.5) becomes

$$
\inf_{x \in \mathcal{H}} \{ f(x) + g(Lx) \},
$$

while the Fenchel dual problem (3.6) reads

$$
\sup_{v \in \mathcal{G}} \left\{ -f^*(-L^* v) - g^*(v) \right\}.
$$

For more primal-dual pairs of convex optimization problems which are particular instances of (3.5)–(3.6), we refer to [58, 122].

Generally, in order to investigate the primal-dual pair (3.5)–(3.6) in the context of Problem 3.1, one has to take

$$
A = \partial f, \ C = \nabla h, \text{ and } B_i = \partial g_i, \ D_i = \partial l_i \text{ for } i = 1, \ldots, m. \tag{3.7}
$$

Then $A$ and $B_i$, $i = 1, \ldots, m$ are maximal monotone, $C$ is monotone and $\mu$-Lipschitz continuous, by [11, Proposition 17.10], and $D_i^{-1} = \nabla l_i^*$ is monotone and $\nu_i$-Lipschitz continuous for $i = 1, \ldots, m$, according to [11, Proposition 17.10, Theorem 18.15 and Corollary 16.24]. One can easily see that (see, for instance, [58, Theorem 4.2]) whenever $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{H} \oplus \mathcal{G}$ is a primal-dual solution to Problem 3.1, with the above choice of the involved operators, $\bar{x}$ is an optimal solution to $(P)$, $(\bar{v}_1, \ldots, \bar{v}_m)$
3.1 Convergence analysis of a forward-backward-forward method

is an optimal solution to (D) and for (P)–(D) strong duality holds, thus the optimal objective values of the two problems coincide. On the other hand, when \( \mathbf{x} \) is an optimal solution to (P) and a qualification condition, like (see, for instance, [22, 58])

\[
\mathbf{E} = \left\{ (L_i x - y_i, \ldots, L_m x - y_m) : x \in \text{dom} f, y_i \in \text{dom} g_i + \text{dom} l_i, i = 1, \ldots, m \right\}
\]

\( (r_1, \ldots, r_m) \in \text{sqr} \mathbf{E} \)

is fulfilled, then there exists \( (\mathbf{v}_1, \ldots, \mathbf{v}_m) \), an optimal solution to (D), such that \( (\mathbf{x}, \mathbf{v}_1, \ldots, \mathbf{v}_m) \in \mathcal{H} \oplus \mathcal{G} \) is a primal-dual solution to Problem 3.1 in the particular formulation given by (3.7).

We also introduce the nonzero bounded linear operator \( \mathbf{L} : \mathcal{H} \to \mathcal{G} \), \( \mathbf{L}x = (L_1 x, \ldots, L_m x) \), its adjoint being \( \mathbf{L}^* : \mathcal{G} \to \mathcal{H} \), \( \mathbf{L}^* \mathbf{v} = \sum_{i=1}^m L_i^* v_i \).

In order to simplify the upcoming formulations and calculations, we introduce the following more compact notations. With respect to Problem 3.4, let \( F : \mathcal{H} \to \mathbb{R} \), \( F(x) = f(x) + h(x) - \langle x, z \rangle \). Then \( \text{dom} F = \text{dom} f \) and its conjugate \( F^* : \mathcal{H} \to \mathbb{R} \) is given by \( F^*(p) = (f + h)^*(z + p) = (f \circ h^*)(z + p) \), since \( \text{dom} h = \mathcal{H} \). Further, we set

\( \mathbf{v} = (v_1, \ldots, v_m) \), \( \mathbf{v} = (v_1, \ldots, v_m) \), \( \mathbf{p}_{2n} = (p_{2,1,n}, \ldots, p_{2,m,n}) \), \( \mathbf{r} = (r_1, \ldots, r_m) \).

We define the function \( G : \mathcal{G} \to \mathbb{R} \), \( G(y) = \sum_{i=1}^m (g_i \circ l_i)(y_i) \) and observe that its conjugate \( G^* : \mathcal{G} \to \mathbb{R} \) is given by \( G^*(\mathbf{v}) = \sum_{i=1}^m (g_i \circ l_i)^*(v_i) = \sum_{i=1}^m (g_i^* + l_i^*)(v_i) \).

Notice that, as \( l_i^*, i = 1, \ldots, m \), has full domain (cf. [11, Theorem 18.15]), we get

\[
\text{dom} G^* = \text{dom} g_1^* \times \ldots \times \text{dom} g_m^*. \quad (3.8)
\]

The primal and the dual optimization problems given in Problem 3.4 can be equivalently represented as

\[
(P) \quad \inf_{x \in \mathcal{H}} \{ F(x) + G(\mathbf{L}x - \mathbf{r}) \},
\]

and, respectively,

\[
(D) \quad \sup_{\mathbf{v} \in \mathcal{G}} \{ -F^*(\mathbf{L}^* \mathbf{v}) - G^*(\mathbf{v}) - \langle \mathbf{v}, \mathbf{r} \rangle \}.
\]

Then \( \mathbf{x} \in \mathcal{H} \) solves (P), \( \mathbf{v} \in \mathcal{G} \) solves (D) and for (P)–(D) strong duality holds if and only if (cf. [22, 30])

\[
-\mathbf{L}^* \mathbf{v} \in \partial F(\mathbf{x}) \quad \text{and} \quad \mathbf{L} \mathbf{x} - \mathbf{r} \in \partial G^*(\mathbf{v}). \quad (3.9)
\]

Let us also mention that for \( \mathbf{x} \in \mathcal{H} \) and \( \mathbf{v} \in \mathcal{G} \) fulfilling (3.9), it holds

\[
[(\mathbf{L} \mathbf{x} - \mathbf{r}, \mathbf{v}) + F(x) - G^*(\mathbf{v})] - [L(\mathbf{x} - \mathbf{r}, \mathbf{v}) + F(\mathbf{x}) - G^*(\mathbf{v})] \geq 0 \quad \forall x \in \mathcal{H} \quad \forall \mathbf{v} \in \mathcal{G}.
\]

For given sets \( B_1 \subseteq \mathcal{H} \) and \( B_2 \subseteq \mathcal{G} \), we introduce the so-called primal-dual gap function restricted to \( B_1 \times B_2 \)

\[
\mathcal{G}_{B_1 \times B_2}(x, \mathbf{v}) = \sup_{\mathbf{v} \in B_2} \{ \langle \mathbf{L}x - \mathbf{r}, \mathbf{v} \rangle + F(x) - G^*(\mathbf{v}) \}
\]- \inf_{\tilde{x} \in B_1} \{ \langle \mathbf{L} \tilde{x} - \mathbf{r}, \mathbf{v} \rangle + F(\tilde{x}) - G^*(\mathbf{v}) \}. \quad (3.10)
\]

We consider the following algorithm for solving (P)–(D), which differs from the primal-dual one given by Combettes and Pesquet in [58, Theorem 3.1] by the fact that we are asking the sequence \( (\gamma_n)_{n \geq 0} \subseteq \mathbb{R}^+ \) to be nondecreasing. 
Algorithm 3.6 Let $x_0 \in \mathcal{H}$ and $(v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}$, set
\[
\beta = \max \{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2},
\]
choose $\varepsilon \in \left(0, \frac{1}{\beta+1}\right)$ and $(\gamma_n)_{n \geq 0}$ a nondecreasing sequence in $\left[\varepsilon, \frac{1-\varepsilon}{\beta}\right]$, and set
\[
\begin{aligned}
(\forall n \geq 0) & \quad p_{1,n} = \text{Prox}_{\gamma_n f}(x_n - \gamma_n (\nabla h(x_n) + \sum_{i=1}^{m} L_i^* v_{i,n} - z)), \\
& \quad \text{For } i = 1, \ldots, m \\
& \quad p_{2,i,n} = \text{Prox}_{\delta g_i} v_{i,n} + \gamma_n (L_i x_n - \nabla l_i^*(v_{i,n} - r_i)), \\
& \quad v_{i,n+1} = \gamma_n L_i (p_{1,n} - x_n) + \gamma_n (\nabla l_i^*(v_{i,n} - \nabla l_i^*(p_{2,i,n})) + p_{2,i,n}, \\
& \quad x_{n+1} = \gamma_n \sum_{i=1}^{m} L_i^* (v_{i,n} - p_{2,i,n}) + \gamma_n (\nabla h(x_n) - \nabla h(p_{1,n})) + p_{1,n}.
\end{aligned}
\]
(3.11)

The following theorem is formulated in the spirit of [48, Theorem 1]. However, the techniques used in the proof are adjusted to the forward-backward-forward structure of Algorithm 3.6 and to the considerably more general problem description involving parallel sums and Lipschitzian operators.

Theorem 3.7 For Problem 3.4, suppose that
\[
z \in \text{ran} \left( \partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \square \partial l_i) (L_i \cdot -r_i) + \nabla h \right).
\]
Then there exists an optimal solution $\bar{x} \in \mathcal{H}$ to (P) and an optimal solution $(\bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{G}$ to (D), such that the following holds for the sequences generated by Algorithm 3.6:

(i) $z - \sum_{i=1}^{m} L_i \bar{v}_i \in \partial f(\bar{x}) + \nabla h(\bar{x})$ and $L_i \bar{x} - r_i \in \partial g_i(\bar{v}_i) + \nabla l_i^*(\bar{v}_i)$, $i = 1, \ldots, m$,
(ii) $x_n \rightharpoonup \bar{x}$, $p_{1,n} \rightharpoonup \bar{x}$ and $v_{i,n} \rightharpoonup \bar{v}_i$, $p_{2,i,n} \rightharpoonup \bar{v}_i$, $i = 1, \ldots, m$,
(iii) For $n \geq 0$, it holds
\[
\frac{\|x_n - \bar{x}\|^2}{2\gamma_n} + \sum_{i=1}^{m} \frac{\|v_{i,n} - \bar{v}_i\|^2}{2\gamma_n} \leq \frac{\|x_0 - \bar{x}\|^2}{2\gamma_0} + \sum_{i=1}^{m} \frac{\|v_{i,0} - \bar{v}_i\|^2}{2\gamma_0},
\]
(iv) If $B_1 \subseteq \mathcal{H}$ and $B_2 \subseteq \mathcal{G}$ are bounded, then for $x^N := \frac{1}{N} \sum_{n=0}^{N-1} p_{1,n}$ and $v_i^N := \frac{1}{N} \sum_{n=0}^{N-1} p_{2,i,n}$, $i = 1, \ldots, m$, the primal-dual gap has the upper bound
\[
\mathcal{G}_{B_1 \times B_2}(x^N, v_1^N, \ldots, v_m^N) \leq \frac{C(B_1, B_2)}{N},
\]
where
\[
C(B_1, B_2) = \sup_{(x,v_1,\ldots,v_m) \in B_1 \times B_2} \left\{ \frac{\|x_0 - x\|^2}{2\gamma_0} + \sum_{i=1}^{m} \frac{\|v_{i,0} - v_i\|^2}{2\gamma_0} \right\},
\]
(v) The sequence $(x^N, v_1^N, \ldots, v_m^N)$ converges weakly to $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m)$.  

3.1 Convergence analysis of a forward-backward-forward method

Proof. Theorem 4.2 in [58] guarantees the existence of an optimal solution \( \overline{x} \in \mathcal{H} \) to (3.5) and of an optimal solution \((\overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{G}\) to (3.6) such that strong duality holds, \( x_n \to \overline{x}, \ p_{1,n} \to \overline{x} \), as well as \( v_{i,n} \to \overline{v}_i \) and \( p_{2,i,n} \to \overline{v}_i \) for \( i = 1, \ldots, m \), when \( n \) converges to \(+\infty\). Hence (i) and (ii) are true. Thus, the solutions \( \overline{x} \) and \( \overline{v} = (\overline{v}_1, \ldots, \overline{v}_m) \) fulfill (3.9).

In view of the sequences \((p_{1,n})_{n \geq 0}\) and \((p_{2,i,n})_{n \geq 0}, \ i = 1, \ldots, m\), generated in Algorithm 3.6, we have for every \( n \geq 0 \)

\[
\frac{x_n - p_{1,n}}{\gamma_n} - \nabla h(x_n) - L^* v_n + z \in \partial f(p_{1,n}),
\]

and

\[
\frac{v_{i,n} - p_{2,i,n}}{\gamma_n} + L_i x_n - \nabla l_i^* (v_{i,n}) - r_i \in \partial g_i^* (p_{2,i,n}), \ i = 1, \ldots, m.
\]

In other words, it holds for every \( n \geq 0 \)

\[
f(x) \geq f(p_{1,n}) + \langle x - p_{1,n} \rangle \quad \forall x \in \mathcal{H}, \tag{3.13}
\]

and, for \( i = 1, \ldots, m \),

\[
g_i^* (v_i) \geq g_i^* (p_{2,i,n}) + \langle \nabla l_i^* (p_{2,i,n}), v_i - p_{2,i,n} \rangle \quad \forall v_i \in \mathcal{G}_i. \tag{3.14}
\]

Additionally, using that \( h \) and \( l_i^* \), \( i = 1, \ldots, m \), are convex and differentiable, it holds for every \( n \geq 0 \)

\[
h(x) \geq h(p_{1,n}) + \langle \nabla h(p_{1,n}), x - p_{1,n} \rangle \quad \forall x \in \mathcal{H}, \tag{3.15}
\]

and, for \( i = 1, \ldots, m \),

\[
l_i^* (v_i) \geq l_i^* (p_{2,i,n}) + \langle \nabla l_i^* (p_{2,i,n}), v_i - p_{2,i,n} \rangle \quad \forall v_i \in \mathcal{G}_i. \tag{3.16}
\]

Consider arbitrary \( x \in \mathcal{H} \) and \( \mathbf{v} = (v_1, \ldots, v_m) \in \mathcal{G} \). Since, for every \( i = 1, \ldots, m \),

\[
\langle x_n - p_{1,n}, x - p_{1,n} \rangle = \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x - p_{1,n}\|^2}{2\gamma_n} - \frac{\|x_n - x\|^2}{2\gamma_n},
\]

\[
\langle v_{i,n} - p_{2,i,n}, v_i - p_{2,i,n} \rangle = \frac{\|v_{i,n} - p_{2,i,n}\|^2}{2\gamma_n} + \frac{\|v_i - p_{2,i,n}\|^2}{2\gamma_n} - \frac{\|v_{i,n} - v_i\|^2}{2\gamma_n},
\]

we obtain for every \( n \geq 0 \), by using the more compact notation of the elements in \( \mathcal{G} \) and by summing up the inequalities (3.13)–(3.16),

\[
\begin{align*}
\frac{\|x_n - x\|^2}{2\gamma_n} + \frac{\|v_n - v\|^2}{2\gamma_n} & \geq \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x - p_{1,n}\|^2}{2\gamma_n} + \frac{\|v_n - p_{2,n}\|^2}{2\gamma_n} + \frac{\|v - p_{2,n}\|^2}{2\gamma_n} \\
& + \sum_{i=1}^m \langle L_i x_n + \nabla l_i^* (p_{2,i,n}) - \nabla l_i^* (v_{i,n}) - r_i, v_i - p_{2,i,n} \rangle - \sum_{i=1}^m (g_i^* + l_i^*) (v_i) + (f + h)(p_{1,n}) \\
& + \langle \nabla h(p_{1,n}) - \nabla h(x_n) - L^* v_n + z, x - p_{1,n} \rangle - \sum_{i=1}^m (g_i^* + l_i^*) (p_{2,i,n}) + (f + h)(x).
\end{align*}
\]
Further, using again the update rules in Algorithm 3.6 and the equations
\[
\left\langle p_{1,n} - x_{n+1}, x - p_{1,n} \right\rangle = \frac{\|x_{n+1} - x\|^2}{2\gamma_n} - \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|x - p_{1,n}\|^2}{2\gamma_n},
\]
and, for \( i = 1, \ldots, m \),
\[
\left\langle p_{2,i,n} - v_{i,n+1}, v_i - p_{2,i,n} \right\rangle = \frac{\|v_{i,n+1} - v_i\|^2}{2\gamma_n} - \frac{\|v_{i,n+1} - p_{2,i,n}\|^2}{2\gamma_n} - \frac{\|v_i - p_{2,i,n}\|^2}{2\gamma_n},
\]
we obtain for every \( n \geq 0 \)
\[
\frac{\|x_n - x\|^2}{2\gamma_n} + \frac{\|v_n - v\|^2}{2\gamma_n} \geq \frac{\|x_{n+1} - x\|^2}{2\gamma_n} + \frac{\|v_{n+1} - v\|^2}{2\gamma_n} + \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|v_n - p_{2,n}\|^2}{2\gamma_n}
- \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} - \frac{\|v_{n+1} - p_{2,n}\|^2}{2\gamma_n} + \left( \langle Lp_{1,n} - r, v \rangle - G^*(v) + F(p_{1,n}) \right)
- \left( \langle Lx - r, p_{2,n} \rangle - G^*(p_{2,n}) + F(x) \right). \tag{3.17}
\]
Further, we equip the Hilbert space \( \mathcal{H} = \mathcal{H} \oplus \mathcal{G} \) with the usual inner product
\[
\langle (y, p), (z, q) \rangle = \langle y, z \rangle + \langle p, q \rangle \quad \forall (y, p), (z, q) \in \mathcal{H} \oplus \mathcal{G}, \tag{3.18}
\]
and the associated norm \( \|(y, p)\| = \sqrt{\|y\|^2 + \|p\|^2} \) for every \( (y, p) \in \mathcal{H} \oplus \mathcal{G} \). For every \( n \geq 0 \), it holds
\[
\frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} + \frac{\|v_{n+1} - p_{2,n}\|^2}{2\gamma_n} = \frac{\|(x_{n+1}, v_{n+1}) - (p_{1,n}, p_{2,n})\|^2}{2\gamma_n},
\]
and, consequently, by making use of the Lipschitz continuity of \( \nabla h \) and \( \nabla l_i^*, i = 1, \ldots, m \), it shows that
\[
\|(x_{n+1}, v_{n+1}) - (p_{1,n}, p_{2,n})\| \leq \gamma_n \left( \sum_{i=1}^m \left( L_i^* (v_{i,n} - p_{2,i,n}) \right)^2 + \sum_{i=1}^m \|L_i (p_{1,n} - x_n)\|^2 \right)
+ \gamma_n \left( \|\nabla h(x_n) - \nabla h(p_{1,n})\|^2 + \sum_{i=1}^m \|\nabla l_i^* (v_{i,n}) - \nabla l_i^* (p_{2,i,n})\|^2 \right)
\leq \gamma_n \left( \sum_{i=1}^m \|L_i\|^2 \sum_{i=1}^m \|v_{i,n} - p_{2,i,n}\|^2 + \left( \sum_{i=1}^m \|L_i\|^2 \right) \|p_{1,n} - x_n\|^2 \right)
+ \gamma_n \left( \mu^2 \|x_n - p_{1,n}\|^2 + \sum_{i=1}^m \nu_i^2 \|v_{i,n} - p_{2,i,n}\|^2 \right)
\leq \gamma_n \left( \sum_{i=1}^m \|L_i\|^2 + \max \{\mu, \nu_1, \ldots, \nu_m\} \right) \|(x_n, v_n) - (p_{1,n}, p_{2,n})\|. \tag{3.19}
\]
3.1 Convergence analysis of a forward-backward-forward method

Hence, by taking into consideration the way in which \((\gamma_n)_{n \geq 0}\) is chosen, we have for every \(n \geq 0\)

\[
\frac{1}{2\gamma_n} \left[ \|x_n - p_{1,n}\|^2 + \|v_n - p_{2,n}\|^2 - \|x_{n+1} - p_{1,n}\|^2 - \|v_{n+1} - p_{2,n}\|^2 \right]
\]

\[
\geq \frac{1}{2\gamma_n} \left( 1 - \gamma_n^2 \left( \sum_{i=1}^m \|L_i\|^2 + \max\{\mu, \nu_1, \ldots, \nu_m\} \right) \right) \|v_n - (p_{1,n}, p_{2,n})\|^2 \geq 0,
\]

and, consequently, (3.17) reduces to

\[
\frac{\|x_n - x\|^2}{2\gamma_n} + \frac{\|v_n - v\|^2}{2\gamma_n} \geq \frac{\|x_{n+1} - x\|^2}{2\gamma_{n+1}} + \frac{\|v_{n+1} - v\|^2}{2\gamma_{n+1}} + \frac{\gamma_{n+1}}{\gamma_n}[\langle Lp_{1,n} - r, v \rangle - G^* (v) + F(p_{1,n})]
\]

\[
+ \frac{\gamma_{n+1}}{\gamma_n} \|v_{n+1} - v\|^2 - \left[ \langle Lx - r, p_{2,n} \rangle - G^* (p_{2,n}) + F(x) \right].
\]

Let \(N \geq 1\) be an arbitrary natural number. By summing up the above inequality from \(n = 0\) to \(N - 1\) and by using the fact that \((\gamma_n)_{n \geq 0}\) is nondecreasing, it follows that

\[
\frac{\|x_0 - x\|^2}{2\gamma_0} + \frac{\|v_0 - v\|^2}{2\gamma_0} \geq \frac{\|x_N - x\|^2}{2\gamma_N} + \sum_{n=0}^{N-1} \frac{\gamma_n}{\gamma_{n+1}} [\langle Lp_{1,n} - r, v \rangle - G^* (v) + F(p_{1,n})]
\]

\[
+ \frac{\|v_N - v\|^2}{2\gamma_N} - \sum_{n=0}^{N-1} \left[ \langle Lx - r, p_{2,n} \rangle - G^* (p_{2,n}) + F(x) \right].
\]

(3.20)

Replacing \(x = \bar{x}\) and \(v = \bar{v}\) in the above estimate, since they fulfill (3.9), we obtain

\[
\sum_{n=0}^{N-1} [\langle Lp_{1,n} - r, \bar{v} \rangle - G^* (\bar{v}) + F(p_{1,n})] - \sum_{n=0}^{N-1} \left[ \langle L\bar{x} - r, p_{2,n} \rangle - G^* (p_{2,n}) + F(\bar{x}) \right] \geq 0.
\]

Consequently,

\[
\frac{\|x_0 - \bar{x}\|^2}{2\gamma_0} + \frac{\|v_0 - \bar{v}\|^2}{2\gamma_0} \geq \frac{\|x_N - \bar{x}\|^2}{2\gamma_N} + \frac{\|v_N - \bar{v}\|^2}{2\gamma_N},
\]

and statement (iii) follows. On the other hand, dividing (3.20) by \(N\), using the convexity of \(F\) and \(G^*\), and denoting \(x^N := \frac{1}{N} \sum_{n=0}^{N-1} p_{1,n}\) and \(v^N := \frac{1}{N} \sum_{n=0}^{N-1} p_{2,i,n}\), \(i = 1, \ldots, m\), we obtain

\[
\frac{1}{N} \left( \frac{\|x_0 - x\|^2}{2\gamma_0} + \frac{\|v_0 - v\|^2}{2\gamma_0} \right) \geq \left[ \langle Lx^N - r, v \rangle - G^* (v) + F(x^N) \right]
\]

\[
- \left[ \langle Lx - r, v^N \rangle - G^* (v^N) + F(x) \right],
\]

which shows (3.12) when passing to the supremum over \(x \in B_1\) and \(v \in B_2\). In this way, statement (iv) is verified. The weak convergence of \((x^N, v^N)\) to \((\bar{x}, \bar{v})\) when \(N\) converges to \(+\infty\) is an easy consequence of the Stolz–Cesàro Theorem, a fact which shows (v).
In the light of Theorem 3.7, the following two remarks are concerned with the convergence of the primal (respectively dual) objective values when Algorithm 3.6 is applied to Problem 3.4.

**Remark 3.8** In the situation when the functions $g_i$ are Lipschitz continuous on $G_i$, $i = 1, \ldots, m$, inequality (3.12) provides for the sequence of primal objective values taken at $(x^N)_{N \geq 1}$ a convergence rate of $O\left(\frac{1}{N}\right)$, namely, it holds

$$F(x^N) + G(Lx^N - r) - F(\bar{x}) - G(L\bar{x} - r) \leq \frac{C(B_1, B_2)}{N} \quad \forall N \geq 1. \quad (3.21)$$

Indeed, due to statement (ii) of the previous theorem, the sequence $(p_{1,n})_{n \geq 0} \subseteq H$ is bounded and one can take $B_1 \subset H$ being a bounded, convex and closed set containing this sequence. Obviously, $\bar{x} \in B_1$. On the other hand, we take $B_2 = \text{dom } g_1^* \times \ldots \times \text{dom } g_m^*$, which is in this situation a bounded set. Then it holds, using the Fenchel–Moreau Theorem and the Young–Fenchel inequality, that

$$G_{B_1 \times B_2}(x^N, v^N) = F(x^N) + G(Lx^N - r) + G^*(v^N) - \inf_{\tilde{x} \in B_1} \left\{ \langle L\tilde{x} - r, v^N \rangle + F(\tilde{x}) \right\}$$

$$\geq F(x^N) + G(Lx^N - r) + G^*(v^N) - \langle L\bar{x} - r, v^N \rangle - F(\bar{x})$$

$$\geq F(x^N) + G(Lx^N - r) - F(\bar{x}) - G(L\bar{x} - r).$$

Hence, (3.21) follows by statement (iv) in Theorem 3.7.

In a similar way, one can show that, whenever $f$ is Lipschitz continuous, (3.12) provides for the sequence of dual objective values taken at $(v^N)_{N \geq 1}$ a convergence rate of $O\left(\frac{1}{N}\right)$.

**Remark 3.9** If $G_i$, $i = 1, \ldots, m$, are finite-dimensional real Hilbert spaces, then (3.21) is true, even under the weaker assumption that the convex functions $g_i$, $i = 1, \ldots, m$, have full domain, without necessarily being Lipschitz continuous. The set $B_1 \subset H$ can be chosen as in Remark 3.8, but this time we take $B_2 = \bigcup_{i=1}^m \cup_{n \geq 0} \partial g_i (L_ip_{1,n}) \subset G$, by also noticing that the functions $g_i$, $i = 1, \ldots, m$, are everywhere subdifferentiable.

The set $B_2$ is bounded, as for every $i = 1, \ldots, m$ the set $\cup_{n \geq 0} \partial g_i (L_ip_{1,n})$ is bounded. Let $i \in \{1, \ldots, m\}$ be fixed. Indeed, as $p_{1,n} \rightharpoonup \bar{x}$, it follows that $L_ip_{1,n} \rightharpoonup L_i\bar{x}$ for $i = 1, \ldots, m$. Using the fact that the subdifferential of $g_i$ is a locally bounded operator at $L_i\bar{x}$, the boundedness of $\cup_{n \geq 0} \partial g_i (L_ip_{1,n})$ follows automatically.

For this choice of the sets $B_1$ and $B_2$, by using the same arguments as in the previous remark, it follows that (3.21) is true.

### 3.1.3 Zeros of sums of monotone operators

In this subsection we turn our attention to the primal-dual monotone inclusion problems formulated in Problem 3.1 with the aim to provide accelerations of the iterative method proposed by Combettes and Pesquet in [58, Theorem 3.1] under additional strong monotonicity assumptions.

**The case when $A + C$ is strongly monotone**

For the beginning, we focus on the case when $A + C$ is $\rho$-strongly monotone for some $\rho \in \mathbb{R}_{++}$ and investigate the impact of this assumption on the convergence rate of
the sequence of primal iterates. The condition \( A + C \) being \( \rho \)-strongly monotone is fulfilled when either \( A : \mathcal{H} \to 2^{\mathcal{H}} \) or \( C : \mathcal{H} \to \mathcal{H} \) is \( \rho \)-strongly monotone. In case that \( A \) is \( \rho_1 \)-strongly monotone and \( C \) is \( \rho_2 \)-strongly monotone, the sum \( A + C \) is \( \rho \)-strongly monotone with \( \rho = \rho_1 + \rho_2 \).

**Remark 3.10** The situation when \( B_i^{-1} + D_i^{-1} \) is \( \tau_i \)-strongly monotone with \( \tau_i \in \mathbb{R}_{++} \) for \( i = 1, \ldots, m \), which improves the convergence rate of the sequence of dual iterates, can be handled with appropriate modifications.

Due to technical reasons, in the following we assume that the operators \( D_i^{-1} \) in Problem 3.1 are zero for \( i = 1, \ldots, m \), thus, we introduce condition (H1), which states

\[
\text{(H1)} \quad D_i(0) = \mathcal{G}_i \text{ and } D_i(x) = \emptyset \text{ for } x \neq 0, \forall i = 1, \ldots, m.
\]

In Remark 3.15, we show, by employing the product space approach, how the results given in this particular context can be employed when treating the primal-dual pair of monotone inclusions (3.1)–(3.2), however, under the assumption that \( D_i^{-1} \), \( i = 1, \ldots, m \), are cocoercive.

The subsequent algorithm represents an accelerated version of the one given in [58, Theorem 3.1] and relies on the fruitful idea of using a second sequence of variable step length parameters \((\sigma_n)_{n \geq 0} \subseteq \mathbb{R}_{++}\), which, together with the sequence of parameters \((\gamma_n)_{n \geq 0} \subseteq \mathbb{R}_{++}\), play an important role in the convergence analysis.

**Algorithm 3.11** Let \( x_0 \in \mathcal{H}, (v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G} \),

\[
\gamma_0 \in \left(0, \min \left\{ 1, \frac{\sqrt{1 + 4\rho}}{2(1 + 2\rho)} \right\} \right), \quad \text{and } \sigma_0 \in \left(0, \frac{1}{2\gamma_0(1 + 2\rho)\sum^m_{i=1} \|L_i\|^2} \right]
\]

Consider the following updates:

\[
(\forall n \geq 0) \quad \begin{aligned}
p_{1,n} &= J_{\gamma_n A} (x_n - \gamma_n (Cx_n + \sum^m_{i=1} L_i^* v_{i,n} - z)), \\
\text{For } i = 1, \ldots, m \quad \begin{aligned}
p_{2,i,n} &= J_{\sigma_n \rho_i^{-1}} (v_{i,n} + \sigma_n (L_i x_n - r_i)), \\
v_{i,n+1} &= \sigma_n L_i (p_{1,n} - x_n) + p_{2,i,n}, \\
x_{n+1} &= \gamma_n \sum^m_{i=1} L_i^* (v_{i,n} - p_{2,i,n}) + \gamma_n (Cx_n - C p_{1,n} + p_{1,n}), \\
\theta_n &= 1/\left(1 + 2\rho \gamma_n (1 - \gamma_n)\right), \quad \gamma_{n+1} = \theta_n \gamma_n, \quad \sigma_{n+1} = \sigma_n / \theta_n.
\end{aligned}
\end{aligned}
\] (3.22)

**Theorem 3.12** For Problem 3.1, suppose that \( A + C \) is \( \rho \)-strongly monotone with \( \rho \in \mathbb{R}_{++} \), suppose that (H1) is fulfilled, and let \((\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \oplus \mathcal{G}\) be a primal-dual solution to Problem 3.1. Then, for every \( n \geq 0 \), it holds

\[
\|x_n - \overline{x}\|^2 + \gamma_n \sum^m_{i=1} \|v_{i,n} - \overline{v}_i\|^2 \leq \gamma^2_n \left(\frac{\|x_0 - \overline{x}\|^2}{\gamma_0^2} + \sum^m_{i=1} \frac{\|v_{i,0} - \overline{v}_i\|^2}{\gamma_0 \sigma_0}\right),
\] (3.23)

where \( \gamma_n, \sigma_n \in \mathbb{R}_{++}, x_n \in \mathcal{H} \), and \((v_{1,n}, \ldots, v_{m,n}) \in \mathcal{G}\) are the iterates generated by Algorithm 3.11.
Proof. Taking into account the definitions of the resolvents occurring in Algorithm 3.11, we obtain
\[
\frac{x_n - p_{1,n}}{\gamma_n} - Cx_n - \sum_{i=1}^{m} L_i^* v_{i,n} + z \in Ap_{1,n},
\]
\[
\frac{v_{i,n} - p_{2,i,n}}{\sigma_n} + L_i x_n - r_i \in B_i^{-1} p_{2,i,n}, \quad i = 1, \ldots, m,
\]
which, in the light of the updating rules in (3.22), furnishes for every \(n \geq 0\)
\[
\frac{x_n - x_{n+1}}{\gamma_n} - \sum_{i=1}^{m} L_i^* p_{2,i,n} + z \in (A + C)p_{1,n},
\]
\[
\frac{v_{i,n} - v_{i,n+1}}{\sigma_n} + L_i p_{1,n} - r_i \in B_i^{-1} p_{2,i,n}, \quad i = 1, \ldots, m.
\]

(3.24)

In consideration of (H1), the primal-dual solution \((\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in H \oplus G\) to Problem 3.1 fulfills (see (3.3), where \(D_i^{-1}\) are taken to be zero for \(i = 1, \ldots, m\))
\[
z - \sum_{i=1}^{m} L_i^* v_i \in A\bar{x} + C\bar{x}, \quad \text{and} \quad v_i \in B_i(L_i \bar{x} - r_i), \quad i = 1, \ldots, m.
\]

Since the sum \(A + C\) is \(\rho\)-strongly monotone, we have for every \(n \geq 0\)
\[
\left\langle p_{1,n} - \bar{x}, \frac{x_n - x_{n+1}}{\gamma_n} - \sum_{i=1}^{m} L_i^* p_{2,i,n} + z - \left(z - \sum_{i=1}^{m} L_i^* v_i\right)\right\rangle \geq \rho \|p_{1,n} - \bar{x}\|^2, \quad (3.25)
\]
while, due to the monotonicity of \(B_i^{-1} : G_i \to 2^{G_i}\), we obtain for every \(n \geq 0\)
\[
\left\langle p_{2,i,n} - \bar{v}_i, \frac{v_{i,n} - v_{i,n+1}}{\sigma_n} + L_i p_{1,n} - r_i - (L_i \bar{x} - r_i)\right\rangle \geq 0, \quad i = 1, \ldots, m. \quad (3.26)
\]

Further, we set
\[
\vec{v} = (\bar{v}_1, \ldots, \bar{v}_m), \quad v_n = (v_{1,n}, \ldots, v_{m,n}), \quad p_{2,n} = (p_{2,1,n}, \ldots, p_{2,m,n}).
\]

Summing up the inequalities (3.25) and (3.26), it follows that
\[
\left\langle p_{1,n} - \bar{x}, \frac{x_n - x_{n+1}}{\gamma_n}\right\rangle + \left\langle p_{2,n} - \vec{v}, \frac{v_n - v_{n+1}}{\sigma_n}\right\rangle \geq \rho \|p_{1,n} - \bar{x}\|^2, \quad (3.27)
\]
and, from here,
\[
\left\langle p_{1,n} - \bar{x}, \frac{x_n - x_{n+1}}{\gamma_n}\right\rangle + \left\langle p_{2,n} - \vec{v}, \frac{v_n - v_{n+1}}{\sigma_n}\right\rangle \geq \rho \|p_{1,n} - \bar{x}\|^2 \quad \forall n \geq 0. \quad (3.28)
\]

In the light of the equations
\[
\left\langle p_{1,n} - \bar{x}, \frac{x_n - x_{n+1}}{\gamma_n}\right\rangle = \left\langle p_{1,n} - x_{n+1}, \frac{x_n - x_{n+1}}{\gamma_n}\right\rangle + \left\langle x_{n+1} - \bar{x}, \frac{x_n - x_{n+1}}{\gamma_n}\right\rangle
\]
\[
= \frac{\|x_{n+1} - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x_n - p_{1,n}\|^2}{2\gamma_n} + \frac{\|x_n - \bar{x}\|^2}{2\gamma_n} - \frac{\|x_{n+1} - \bar{x}\|^2}{2\gamma_n},
\]
and
\[
\left\langle \mathbf{p}_{2,n} - \mathbf{v}, \frac{\mathbf{v}_n - \mathbf{v}_{n+1}}{\sigma_n} \right\rangle = \left\langle \mathbf{p}_{2,n} - \mathbf{v}_{n+1}, \frac{\mathbf{v}_n - \mathbf{v}_{n+1}}{\sigma_n} \right\rangle + \left\langle \mathbf{v}_{n+1} - \mathbf{v}, \frac{\mathbf{v}_n - \mathbf{v}_{n+1}}{\sigma_n} \right\rangle
\]
\[
= \frac{\|\mathbf{v}_{n+1} - \mathbf{p}_{2,n}\|^2}{2\sigma_n} - \frac{\|\mathbf{v}_n - \mathbf{p}_{2,n}\|^2}{2\sigma_n} + \frac{\|\mathbf{v}_n - \mathbf{v}\|^2}{2\sigma_n} - \frac{\|\mathbf{v}_{n+1} - \mathbf{v}\|^2}{2\sigma_n},
\]

inequality (3.28) reads for every \(n \geq 0\)
\[
\frac{\|\mathbf{x}_n - \mathbf{v}\|^2}{2\gamma_n} + \frac{\|\mathbf{v}_n - \mathbf{v}\|^2}{2\sigma_n} \geq \rho \|\mathbf{p}_{1,n} - \mathbf{v}\|^2 + \frac{\|\mathbf{x}_{n+1} - \mathbf{v}\|^2}{2\gamma_n} + \frac{\|\mathbf{v}_{n+1} - \mathbf{v}\|^2}{2\gamma_n} + \frac{\|\mathbf{x}_n - \mathbf{p}_{1,n}\|^2}{2\gamma_n} + \frac{\|\mathbf{v}_n - \mathbf{p}_{1,n}\|^2}{2\gamma_n} - \frac{\|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\|^2}{2\gamma_n},
\]

(3.29)

Using that \(2ab \leq a \alpha^2 + b^2 \alpha^2\) for all \(a, b \in \mathbb{R}, \alpha \in \mathbb{R}_{++}\), we obtain for \(\alpha := \gamma_n\)
\[
\rho \|\mathbf{p}_{1,n} - \mathbf{v}\|^2 \geq \frac{\rho}{2\gamma_n} \|\mathbf{x}_{n+1} - \mathbf{v}\|^2 - 2\|x_{n+1} - \mathbf{v}\| \|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\| + \|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\|^2
\]
\[
\geq \frac{2\rho \gamma_n (1 - \gamma_n)}{2\gamma_n} \|\mathbf{x}_{n+1} - \mathbf{v}\|^2 - \frac{2\rho (1 - \gamma_n)}{2\gamma_n} \|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\|^2,
\]

which, in combination with (3.29), yields for every \(n \geq 0\)
\[
\frac{\|\mathbf{x}_n - \mathbf{v}\|^2}{2\gamma_n} + \frac{\|\mathbf{v}_n - \mathbf{v}\|^2}{2\sigma_n} \geq \frac{(1 + 2\rho \gamma_n (1 - \gamma_n)) \|\mathbf{x}_{n+1} - \mathbf{v}\|^2}{2\gamma_n} + \frac{\|\mathbf{x}_n - \mathbf{p}_{1,n}\|^2}{2\gamma_n} + \frac{\|\mathbf{v}_n - \mathbf{p}_{1,n}\|^2}{2\sigma_n} + \frac{\|\mathbf{v}_{n+1} - \mathbf{v}\|^2}{2\sigma_n} - \frac{\|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\|^2}{2\gamma_n},
\]

(3.30)

Investigating the last two terms in the right-hand side of the above estimate, it shows that for every \(n \geq 0\)
\[
- \frac{(1 + 2\rho (1 - \gamma_n)) \|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\|^2}{2\gamma_n}
\]
\[
\geq - \frac{(1 + 2\rho) \gamma_n}{2} \left( \sum_{i=1}^{m} L_i^2 \|v_i - p_{2,i,n}\| + (C \mathbf{x}_n - C \mathbf{p}_{1,n}) \right)
\]
\[
\geq - \frac{2(1 + 2\rho) \gamma_n}{2} \left( \sum_{i=1}^{m} \|L_i\|^2 \right) \|\mathbf{v}_n - \mathbf{p}_{2,n}\|^2 + \mu^2 \|\mathbf{x}_n - \mathbf{p}_{1,n}\|^2,
\]

and
\[
- \frac{\|\mathbf{v}_{n+1} - \mathbf{p}_{2,n}\|^2}{2\sigma_n} = - \frac{\sigma_n}{2} \left( \sum_{i=1}^{m} \|L_i(p_{1,n} - \mathbf{x}_n)\|^2 \right) \geq - \frac{\sigma_n}{2} \left( \sum_{i=1}^{m} \|L_i\|^2 \right) \|p_{1,n} - \mathbf{x}_n\|^2.
\]

Hence, for every \(n \geq 0\), it holds
\[
\frac{\|\mathbf{x}_n - \mathbf{p}_{1,n}\|^2}{2\gamma_n} + \frac{\|\mathbf{v}_n - \mathbf{p}_{2,n}\|^2}{2\sigma_n} \geq \frac{(1 + 2\rho (1 - \gamma_n)) \|\mathbf{x}_{n+1} - \mathbf{p}_{1,n}\|^2}{2\gamma_n} - \frac{\|\mathbf{v}_{n+1} - \mathbf{p}_{2,n}\|^2}{2\sigma_n}
\]
\[
\geq \frac{(1 - \gamma_n \sigma_n \sum_{i=1}^{m} \|L_i\|^2 - 2(1 + 2\rho) \gamma_n \mu^2)}{2\gamma_n} \|p_{1,n} - \mathbf{x}_n\|^2 + \frac{(1 - 2\gamma_n \sigma_n (1 + 2\rho) \sum_{i=1}^{m} \|L_i\|^2)}{2\sigma_n} \|\mathbf{v}_n - \mathbf{p}_{2,n}\|^2
\]
\[
\geq 0.
\]
The nonnegativity of the expression in the above relation follows since the sequence \((\gamma_n)_{n \geq 0}\) is nonincreasing, \(\gamma_n \sigma_n = \gamma_0 \sigma_0\) for every \(n \geq 0\), and
\[
\gamma_0 \in \left(0, \min \left\{1, \frac{\sqrt{1 + 4 \rho}}{2(1 + 2 \rho) \mu} \right\}\right), \text{ and } \sigma_0 \in \left(0, \frac{1}{2 \gamma_0 (1 + 2 \rho) \sum_{i=1}^m \|L_i\|^2} \right).
\]
Consequently, inequality (3.30) becomes for all \(n \geq 0\)
\[
\frac{\|x_n - \overline{x}\|^2}{2 \gamma_n} + \frac{\|v_n - \overline{v}\|^2}{2 \sigma_n} \geq \frac{(1 + 2 \rho \gamma_n (1 - \gamma_n)) \|x_{n+1} - \overline{x}\|^2}{2 \gamma_n} + \frac{\|v_{n+1} - \overline{v}\|^2}{2 \sigma_n}. \tag{3.31}
\]
Dividing (3.31) by \(\gamma_n\) and making use of
\[
\theta_n = \frac{1}{\sqrt{1 + 2 \rho \gamma_n (1 - \gamma_n)}}, \quad \gamma_{n+1} = \theta_n \gamma_n, \quad \sigma_{n+1} = \frac{\sigma_n}{\theta_n},
\]
we obtain
\[
\frac{\|x_n - \overline{x}\|^2}{2 \gamma_n^2} + \frac{\|v_n - \overline{v}\|^2}{2 \gamma_n \sigma_n} \geq \frac{\|x_{n+1} - \overline{x}\|^2}{2 \gamma_{n+1}^2} + \frac{\|v_{n+1} - \overline{v}\|^2}{2 \gamma_{n+1} \sigma_{n+1}} \quad \forall n \geq 0.
\]
Let \(N \geq 1\) be some arbitrary positive integer. Then, summing up these inequalities from \(n = 0\) to \(N - 1\), we finally get
\[
\frac{\|x_0 - \overline{x}\|^2}{2 \gamma_0^2} + \frac{\|v_0 - \overline{v}\|^2}{2 \gamma_0 \sigma_0} \geq \frac{\|x_N - \overline{x}\|^2}{2 \gamma_N^2} + \frac{\|v_N - \overline{v}\|^2}{2 \gamma_N \sigma_N}. \tag{3.32}
\]
In conclusion,
\[
\frac{\|x_0 - \overline{x}\|^2}{2} + \gamma_n \frac{\|v_0 - \overline{v}\|^2}{2 \sigma_n} \leq \gamma_n^2 \left(\frac{\|x_0 - \overline{x}\|^2}{2 \gamma_0^2} + \frac{\|v_0 - \overline{v}\|^2}{2 \gamma_0 \sigma_0}\right) \quad \forall n \geq 0, \tag{3.33}
\]
which completes the proof. \hfill \blacksquare

In the sequel, we show that \(\rho \gamma_n\) converges like \(\frac{1}{n}\) as \(n \to +\infty\).

**Proposition 3.13** Let \(\gamma_0 \in (0, 1)\) and consider the sequence \((\gamma_n)_{n \geq 0} \subseteq \mathbb{R}_{++}\), where
\[
\gamma_{n+1} = \frac{\gamma_n}{\sqrt{1 + 2 \rho \gamma_n (1 - \gamma_n)}} \quad \forall n \geq 0. \tag{3.34}
\]
Then \(\lim_{n \to +\infty} n \rho \gamma_n = 1\).

Proof. Since the sequence \((\gamma_n)_{n \geq 0} \subseteq (0, 1)\) is bounded and decreasing, it converges towards some \(l \in [0, 1]\) as \(n \to +\infty\). We let \(n \to +\infty\) in (3.34) and obtain
\[
l^2 (1 + 2 \rho (1 - l)) = l^2 \iff 2 \rho l^3 (1 - l) = 0,
\]
which shows that \(l = 0\), i.e., \(\gamma_n \to 0\) \((n \to +\infty)\). On the other hand, (3.34) implies that \(\gamma_n_{n \to +\infty} \to 1\) \((n \to +\infty)\). As \(\frac{1}{\gamma_n} \) is a strictly increasing and unbounded sequence, by applying the Stolz–Cesàro Theorem, it shows that
\[
\lim_{n \to +\infty} n \gamma_n = \lim_{n \to +\infty} \frac{n}{n} = \lim_{n \to +\infty} \frac{n + 1 - n}{\gamma_{n+1} - \gamma_n} = \lim_{n \to +\infty} \frac{\gamma_n \gamma_{n+1}}{\gamma_n - \gamma_{n+1}}
= \lim_{n \to +\infty} \frac{\gamma_n \gamma_{n+1} (\gamma_n + \gamma_{n+1})}{\gamma_n - \gamma_{n+1}} \tag{3.34} \Rightarrow \lim_{n \to +\infty} \frac{\gamma_n \gamma_{n+1} (\gamma_n + \gamma_{n+1})}{2 \rho \gamma_{n+1}^2 (1 - \gamma_n)}
= \lim_{n \to +\infty} \frac{\gamma_n + \gamma_{n+1}}{2 \rho (1 - \gamma_n)} = \lim_{n \to +\infty} \frac{\gamma_n + \gamma_{n+1}}{2 \rho (1 - \gamma_n)} = \frac{2}{2 \rho} = \frac{1}{\rho},
\]
which completes the proof. \hfill \blacksquare
The following result is a consequence of Theorem 3.12 and Proposition 3.13.

**Theorem 3.14** For Problem 3.1, suppose that $A + C$ is $\rho$-strongly monotone, suppose that (H1) is fulfilled, and let $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{H} \oplus \mathcal{G}$ be a primal-dual solution to Problem 3.1. Then, for any $\varepsilon > 0$, there exists some $n_0 \in \mathbb{N}$ (depending on $\varepsilon$ and $\rho \gamma_0$) such that for any $n \geq n_0$

$$
||x_n - \bar{x}||^2 \leq \frac{1 + \varepsilon}{n^2} \left( \frac{||x_0 - \bar{x}||^2}{\rho^2 \gamma_0^2} + \sum_{i=1}^{m} \frac{||v_{i,0} - \bar{v}_i||^2}{\rho^2 \gamma_0 \sigma_0} \right), \quad (3.35)
$$

where $\gamma_n, \sigma_n \in \mathbb{R}_{++}, x_n \in \mathcal{H}$, and $(v_{1,n}, \ldots, v_{m,n}) \in \mathcal{G}$ are the iterates generated by Algorithm 3.11.

**Remark 3.15** In Algorithm 3.11 and Theorem 3.14, we assumed that $D_i^{-1}(v) = 0$ for all $v \in \mathcal{G}_i$ and every $i = 1, \ldots, m$. However, similar statements can also be provided for Problem 3.1 under the additional assumption that the operators $D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ are such that $D_i^{-1}$ is $\nu_i^{-1}$-cocoercive with $\nu_i \in \mathbb{R}_{++}$ for $i = 1, \ldots, m$. This assumption is in general stronger than assuming that $D_i$ is monotone such that $D_i^{-1}$ is $\nu_i$-Lipschitzian for $i = 1, \ldots, m$. However, it guarantees that $D_i$ is $\nu_i^{-1}$-strongly monotone and maximally monotone for $i = 1, \ldots, m$ (see [11, Example 20.28, Proposition 20.22 and Example 22.6]). We introduce the Hilbert space $\mathcal{H} = \mathcal{H} \oplus \mathcal{G}$, the element $z = (z,0,\ldots,0) \in \mathcal{H}$ and the maximally monotone operator $A : \mathcal{H} \to 2^\mathcal{H}$, $A(x,y_1,\ldots,y_m) = (Ax,D_1y_1,\ldots,D_my_m)$ and the monotone and Lipschitzian operator $C : \mathcal{H} \to \mathcal{H}$, $C(x,y_1,\ldots,y_m) = (Cx,0,\ldots,0)$. Notice also that $A + C$ is strongly monotone on $\mathcal{H}$. Furthermore, we introduce the element $r = (r_1,\ldots,r_m) \in \mathcal{G}$, the maximally monotone operator $B : \mathcal{G} \to 2^\mathcal{G}$, $B(y_1,\ldots,y_m) = (B_1y_1,\ldots,B_my_m)$, and the bounded linear operator $L : \mathcal{H} \to \mathcal{G}$, $L(x,y_1,\ldots,y_m) = (L_1x-y_1,\ldots,L_mx-y_m)$, having as adjoint $L^* : \mathcal{G} \to \mathcal{H}$, $L^*(q_1,\ldots,q_m) = (\sum_{i=1}^{m} L_i^*q_i,-q_1,\ldots,-q_m)$. We consider the primal problem

$$
\text{find } \bar{x} = (\bar{x}, \bar{p}_1, \ldots, \bar{p}_m) \in \mathcal{H} \text{ such that } z \in A\bar{x} + L^*B(L\bar{x} - r) + C\bar{x}, \quad (3.36)
$$

together with the dual inclusion problem

$$
\text{find } \bar{v} \in \mathcal{G} \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - L^*\bar{v} \in Ax + Cx, \\ \bar{v} \in B(Lx - r). \end{cases} \quad (3.37)
$$

We notice that Algorithm 3.11 can be employed for solving this primal-dual pair of monotone inclusion problems by separately involving the resolvents of $A$, $B_i$, and $D_i$, $i = 1, \ldots, m$, as for $\gamma \in \mathbb{R}_{++}$

$$
J_{\gamma A}(x,y_1,\ldots,y_m) = (J_{\gamma A}x, J_{\gamma D_1}y_1,\ldots,J_{\gamma D_m}y_m) \forall (x,y_1,\ldots,y_m) \in \mathcal{H},
$$

$$
J_{\gamma B}(q_1,\ldots,q_m) = (J_{\gamma B_1}q_1,\ldots,J_{\gamma B_m}q_m) \forall (q_1,\ldots,q_m) \in \mathcal{G}.
$$

Having $(\bar{x}, \bar{v}) \in \mathcal{H} \oplus \mathcal{G}$ a primal-dual solution to the primal-dual pair of monotone inclusion problems (3.36)–(3.37), Algorithm 3.11 generates a sequence of primal iterates fulfilling (3.35) in $\mathcal{H}$. Moreover, $(\bar{x}, \bar{v})$ is a primal-dual solution to (3.36)–
\[ z - L^* v \in A x + C x \text{ and } v \in B (L x - r) \]

\[
\iff
z - \sum_{i=1}^{m} L_i^* v_i \in A x + C x \text{ and } v_i \in D_i \mathcal{P}_i, \forall i \in \{1, \ldots, m\}, i = 1, \ldots, m \\
\iff
z - \sum_{i=1}^{m} L_i^* v_i \in A x + C x \text{ and } v_i \in D_i \mathcal{P}_i, L_i x - r_i \in B_i^{-1} v_i + \mathcal{P}_i, i = 1, \ldots, m. 
\]

Thus, if \((\mathbf{x}, \mathbf{v})\) is a primal-dual solution to (3.36)–(3.37), then \((\mathbf{x}, \mathbf{v})\) is a primal-dual solution to (3.1)–(3.2). Vice versa, if \((\mathbf{x}, \mathbf{v})\) is a primal-dual solution to (3.1)–(3.2), then, choosing \(\mathcal{P}_i \in D_i^{-1} v_i, i = 1, \ldots, m, \) and \(\mathbf{x} = (\mathcal{P}_1 \ldots \mathcal{P}_m)\), it yields that \((\mathbf{x}, \mathbf{v})\) is a primal-dual solution to (3.36)–(3.37). In conclusion, the first component of every primal iterate in \(\mathcal{H}\) generated by Algorithm 3.11 for finding the primal-dual solution \((\mathbf{x}, \mathbf{v})\) to (3.36)–(3.37) will furnish a sequence of iterates verifying (3.35) in \(\mathcal{H}\) for the primal-dual solution \((\mathbf{x}, \mathbf{v})\) to (3.1)–(3.2).

**The case when \(A + C, \text{ and } B_i^{-1} + D_i^{-1}, i = 1, \ldots, m, \text{ are strongly monotone}**

Within this subsection we consider the case when \(A + C\) is \(\rho\)-strongly monotone with \(\rho \in \mathbb{R}_{++}\), and \(B_i^{-1} + D_i^{-1}\) is \(\tau_i\)-strongly monotone with \(\tau_i \in \mathbb{R}_{++}\) for \(i = 1, \ldots, m\). We provide an accelerated version of the algorithm in [58, Theorem 3.1] which generates sequences of primal and dual iterates that converge to the primal-dual solution to Problem 3.1 with an improved rate of convergence.

**Algorithm 3.16** Let \(x_0 \in \mathcal{H}, (v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}\), and \(\gamma \in (0, 1)\) such that

\[
\gamma \leq \frac{1}{\sqrt{1 + 2 \min \{\rho, \tau_1, \ldots, \tau_m\}} \left( \sqrt{\sum_{i=1}^{m} \|L_i\|^2} + \max \{\mu, \nu_1, \ldots, \nu_m\} \right)}.
\]

Consider the following updates:

\[
\forall n \geq 0, \begin{cases} 
    \mathcal{P}_1 = J_{\gamma A}(x_n - \gamma (C x_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z)), \\
    \text{For } i = 1, \ldots, m \\
    \mathcal{P}_{2,i,n} = J_{\gamma B_i^{-1}}(v_{i,n} + \gamma (L_i x_n - D_i^{-1} v_{i,n} - r_i)), \\
    v_{i,n+1} = \gamma L_i \mathcal{P}_{1,n} - x_n + \gamma (D_i^{-1} v_{i,n} - D_i^{-1} \mathcal{P}_{2,i,n}) + \mathcal{P}_{2,i,n}, \\
    x_{n+1} = \gamma \sum_{i=1}^{m} L_i^* (v_{i,n} - \mathcal{P}_{2,i,n}) + \gamma (C x_n - C \mathcal{P}_{1,n}) + \mathcal{P}_{1,n}.
\end{cases}
\]

**Theorem 3.17** For Problem 3.1, suppose that \(A + C\) is \(\rho\)-strongly monotone with \(\rho \in \mathbb{R}_{++}\), \(B_i^{-1} + D_i^{-1}\) is \(\tau_i\)-strongly monotone with \(\tau_i \in \mathbb{R}_{++}\) for \(i = 1, \ldots, m\), and let \((\mathbf{x}, \mathbf{v}_1, \ldots, \mathbf{v}_m) \in \mathcal{H} \oplus \mathcal{G}\) be the unique primal-dual solution to Problem 3.1. Then for every \(n \geq 0\), it holds

\[
\|x_n - x\|^2 + \sum_{i=1}^{m} \|v_{i,n} - \mathcal{P}_i\|^2 \leq \left( \frac{1}{1 + 2 \rho_{\min} \gamma (1 - \gamma)} \right)^n \left( \|x_0 - x\|^2 + \sum_{i=1}^{m} \|v_{i,0} - \mathcal{P}_i\|^2 \right),
\]

where \(\rho_{\min} = \min \{\rho, \tau_1, \ldots, \tau_m\}\), \(x_n \in \mathcal{H}\), and \((v_{1,n}, \ldots, v_{m,n}) \in \mathcal{G}\) are the iterates generated by Algorithm 3.16.
Proof. Taking into account the definitions of the resolvents occurring in Algorithm 3.16 and the fact that the primal-dual solution \((\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \oplus \mathcal{G}\) to Problem 3.1 is unique and fulfills (3.3), by the strong monotonicity of \(A + C\) and \(B_i^{-1} + D_i^{-1}, i = 1, \ldots, m\), we obtain for every \(n \geq 0\)

\[
\left\langle p_{1,n} - \overline{x}, \frac{x_n - x_{n+1}}{\gamma} \right\rangle - \sum_{i=1}^{m} L_i^* p_{2,i,n} + z - \left( z - \sum_{i=1}^{m} L_i^* \overline{v}_i \right) \geq \rho \|p_{1,n} - \overline{x}\|^2, \tag{3.39}
\]

and, respectively, for each \(i = 1, \ldots, m\),

\[
\left\langle p_{2,i,n} - \overline{v}_i, \frac{v_{i,n} - v_{i,n+1}}{\gamma} \right\rangle + L_i p_{1,n} - r_i - (L_i \overline{x} - r_i) \geq \tau_i \|p_{2,i,n} - \overline{v}_i\|^2. \tag{3.40}
\]

Consider again the Hilbert space \(\mathcal{H} = \mathcal{H} \oplus \mathcal{G}\), which is equipped with the inner product defined in (3.18) and associated norm, and set

\[
\overline{x} = (x_1, v_1, \ldots, v_m), \quad x_n = (x_n, v_{1,n}, \ldots, v_{m,n}), \quad p_n = (p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n}).
\]

Summing up the inequalities (3.39) and (3.40), and using

\[
\left\langle p_n - \overline{x}, \frac{x_n - x_{n+1}}{\gamma} \right\rangle = \frac{\|x_{n+1} - p_n\|^2}{2\gamma} - \frac{\|x_n - p_n\|^2}{2\gamma} + \frac{\|x_n - \overline{x}\|^2}{2\gamma} - \frac{\|x_{n+1} - \overline{x}\|^2}{2\gamma},
\]

we obtain for every \(n \geq 0\)

\[
\frac{\|x_n - \overline{x}\|^2}{2\gamma} \geq \rho_{\min} \|p_n - \overline{x}\|^2 + \frac{\|x_{n+1} - \overline{x}\|^2}{2\gamma} + \frac{\|x_n - p_n\|^2}{2\gamma} - \frac{\|x_{n+1} - p_n\|^2}{2\gamma}. \tag{3.41}
\]

Further, using the estimate \(2ab \leq \gamma a^2 + \frac{b^2}{\gamma}\) for all \(a, b \in \mathbb{R}\), we obtain

\[
\rho_{\min} \|p_n - \overline{x}\|^2 \geq \frac{2\rho_{\min} \gamma (1 - \gamma)}{2\gamma} \|x_{n+1} - \overline{x}\|^2 - \frac{2\rho_{\min} \gamma (1 - \gamma)}{2\gamma} \|x_{n+1} - p_n\|^2 \\
\geq \frac{2\rho_{\min} \gamma (1 - \gamma)}{2\gamma} \|x_{n+1} - \overline{x}\|^2 - \frac{2\rho_{\min}}{2\gamma} \|x_{n+1} - p_n\|^2 \quad \forall n \geq 0.
\]

Hence, (3.41) reduces to

\[
\frac{\|x_n - \overline{x}\|^2}{2\gamma} \geq \frac{(1 + 2\rho_{\min} \gamma (1 - \gamma)) \|x_{n+1} - \overline{x}\|^2}{2\gamma} + \frac{\|x_n - p_n\|^2}{2\gamma} - \frac{(1 + 2\rho_{\min}) \|x_{n+1} - p_n\|^2}{2\gamma} \quad \forall n \geq 0.
\]

Using the same arguments as in (3.19), it is easy to check that for every \(n \geq 0\)

\[
\frac{\|x_n - p_n\|^2}{2\gamma} - \frac{(1 + 2\rho_{\min}) \|x_{n+1} - p_n\|^2}{2\gamma} \\
\geq \left(1 - (1 + 2\rho_{\min})\gamma^2 \left(\sum_{i=1}^{m} \|L_i\|^2 + \max \{\mu, \nu_1, \ldots, \nu_m\}\right)\right) \frac{\|x_n - p_n\|^2}{2\gamma} \geq 0,
\]

3.1 Convergence analysis of a forward-backward-forward method
whereby the nonnegativity of this term is ensured by the assumption that

\[
\gamma \leq \frac{1}{\sqrt{1 + 2\rho_{\min}} \left( \sqrt{\sum_{i=1}^{m} \|L_i\|^2 + \max \{\mu, \nu_1, \ldots, \nu_m\}} \right)}.
\]

Therefore, we obtain

\[
\|x_n - \mathbf{f}\|^2 \geq (1 + 2\rho_{\min}\gamma(1 - \gamma))\|x_{n+1} - \mathbf{f}\|^2 \forall n \geq 0,
\]

which finally leads to

\[
\|x_n - \mathbf{f}\|^2 \leq \left( \frac{1}{1 + 2\rho_{\min}\gamma(1 - \gamma)} \right)^n \|x_0 - \mathbf{f}\|^2 \forall n \geq 0.
\]

\[\blacksquare\]

### 3.2 A Douglas–Rachford type primal-dual method

In this section, which is related to our article \[38\], we propose two different primal-dual error-tolerant methods for solving monotone inclusion problems with mixtures of composite and parallel sum type monotone operators. Both algorithms rely on the inexact Douglas–Rachford algorithm (cf. \[53, 54\]). Nonetheless, in view of their conceptual design, in many different ways they still differ clearly from each other.

An important feature of the two approaches, and, simultaneously, an advantage over many existing semismooth or nonsmooth methods from the literature is their capability of processing the set-valued operators separately via their resolvents, while the bounded linear operators existent in the problem description (resp. their adjoints) are evaluated via explicit forward steps. The resolvents of the maximally monotone operators are in general not available in closed form expressions, some fact which motivates the inexact versions of the algorithms, where implementation errors in the shape of summable sequences are allowed.

Modern research in nonsmooth convex optimization (see, for instance, \[25, 37, 39, 43, 58, 122\]) has shown that highly structured monotone inclusion problems can be efficiently solved via primal-dual splitting approaches. In the work of Combettes and Pesquet (cf. \[58\]), the problem involving sums of set-valued, linear composite, Lipschitzian, and parallel sum type monotone operators was decomposed and solved via an inexact forward-backward-forward algorithm. On the other hand, in the paper \[122\] by Vũ, instead of Lipschitzian operators, the author has assumed cocoercive operators and solved the resulting problem with an inexact forward-backward algorithm. To this end, our methods can be seen as natural extensions of these approaches, this time by employing the inexact Douglas–Rachford method. One further primal-dual method which relies on the same fundamental splitting algorithm is considered by Condat in \[60\] in terms of solving minimization problems having as objective the sum of two proper, convex, and lower semicontinuous functions, where one of them is composed with a bounded linear operator.

Due to the nature of Douglas–Rachford splitting (only backward evaluations appear in the iteration process), we will neither assume cocoercivity nor Lipschitz continuity for any of the operators occurring in the description of the monotone inclusion problem. This represents a drawback since we are not making distinctions between these specific operators and ordinary set-valued maximally monotone ones.
whenever they occur in corresponding applications. This, however, is compensated
by the advantage of allowing general maximal monotone operators in the parallel
sum constructions which relaxes the working hypotheses in [58,122] and is relevant
for several applications as demonstrated in Section 4.3.

3.2.1 Problem description

Within this section we provide two algorithms together with weak and strong
convergence results for the following primal-dual pair of monotone inclusion problems.

**Problem 3.18** Consider the real Hilbert space \( \mathcal{H} \), let \( z \in \mathcal{H} \), and let \( A : \mathcal{H} \to 2^\mathcal{H} \) be
a maximally monotone operator. Furthermore, for every \( i = 1, \ldots, m \), consider the
real Hilbert space \( \mathcal{G}_i \), let \( r_i \in \mathcal{G}_i \), let \( B_i : \mathcal{G}_i \to 2^{\mathcal{G}_i} \), and \( D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i} \) be maximally
monotone operators, and \( L_i : \mathcal{H} \to \mathcal{G}_i \) a nonzero bounded linear operator. The
problem is to solve the primal inclusion

\[
\text{find } \mathbf{v} \in \mathcal{H} \text{ such that } z \in A\mathbf{v} + \sum_{i=1}^{m} L_i^*(B_i \square D_i)(L_i\mathbf{v} - r_i)
\]

(3.42)

together with the dual inclusion

\[
\text{find } \mathbf{v}_1 \in \mathcal{G}_1, \ldots, \mathbf{v}_m \in \mathcal{G}_m \text{ such that } (\exists \mathbf{x} \in \mathcal{H}) \left\{ z - \sum_{i=1}^{m} L_i^*\mathbf{v}_i \in Ax, \right.
\]

\[
\left. \mathbf{v}_i \in (B_i \square D_i)(L_i\mathbf{x} - r_i), \; i = 1, \ldots, m. \right\}
\]

(3.43)

At this point we would like to remind the reader that the concept of primal-dual
solutions has been introduced in Remark 3.2, while a particular instance of this
problem was given in Example 3.5 in terms of solving convex optimization problems.

Before starting with our investigations, we consider the following proposition.

**Proposition 3.19** For each \( i = 1, \ldots, m \), let \( \phi_i : \mathbb{R}_+ \to [0, +\infty] \) be an increasing
function vanishing only at 0. Then, the function \( \phi : \mathbb{R}_+ \to [0, +\infty] \), defined as

\[
\phi(c) = \inf \left\{ \sum_{i=1}^{m} \phi_i(t_i) : \sum_{i=1}^{m} t_i^2 = c, \; t_i \in \mathbb{R}_+, \; i = 1, \ldots, m \right\},
\]

is increasing and vanishes only at 0.

Proof. We let \( c_1, c_2 \in \mathbb{R}_+ \), such that \( c_1 > c_2 \). Then, it shows that

\[
\phi(c_1) = \inf \left\{ \sum_{i=1}^{m} \phi_i(t_i) : \sum_{i=1}^{m} t_i^2 = c_1, \; t_i \in \mathbb{R}_+, \; i = 1, \ldots, m \right\}
\]

\[
\geq \inf \left\{ \sum_{i=1}^{m} \phi_i\left( \frac{c_2}{c_1} t_i \right) : \sum_{i=1}^{m} t_i^2 = c_1, \; t_i \in \mathbb{R}_+, \; i = 1, \ldots, m \right\}
\]

\[
= \inf \left\{ \sum_{i=1}^{m} \phi_i(\tilde{t}_i) : \sum_{i=1}^{m} \tilde{t}_i^2 = c_2, \; \tilde{t}_i \in \mathbb{R}_+, \; i = 1, \ldots, m \right\} = \phi(c_2),
\]
where the second equality holds, since, for every \( i = 1, \ldots, m \), we set
\[
\tilde{t}_i := \frac{c_2}{c_1} t_i, \quad \text{and therefore} \quad \sqrt{\sum_{i=1}^{m} \tilde{t}_i^2} = \sqrt{\sum_{i=1}^{m} \left( \frac{c_2}{c_1} t_i \right)^2} = \frac{c_2}{c_1} \sqrt{\sum_{i=1}^{m} t_i^2} = c_2.
\]

Now, we let \( c \in \mathbb{R}_+ \) and assume that \( \phi(c) = 0 \). Then, for each \( i = 1, \ldots, m \), there exists a sequence \( (t_{i,n})_{n \geq 0} \subseteq \mathbb{R}_+ \), such that \( \sqrt{\sum_{i=1}^{m} t_{i,n}^2} = c \) for all \( n \geq 0 \) and \( \sum_{i=1}^{m} \phi_i(t_{i,n}) \to 0 \ (n \to +\infty) \). Therefore, \( \phi_i(t_{i,n}) \to 0 \ (n \to +\infty) \) for each \( i = 1, \ldots, m \). In conclusion, since \( \phi_i \) vanishes only at 0, it follows that \( t_{i,n} \to 0 \) for each \( i = 1, \ldots, m \), which, finally, implies that \( c = 0 \).

\[\] 

### 3.2.2 A first primal-dual algorithm

The first iterative scheme we propose in this subsection has the particularity that it accesses the resolvents of \( A, B_i^{-1} \), and \( D_i^{-1} \), \( i = 1, \ldots, m \), and processes each operator \( L_i \) and its adjoint \( L_i^* \), \( i = 1, \ldots, m \), two times.

**Algorithm 3.20** Let \( x_0 \in \mathcal{H} \), \( (v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \times \cdots \times \mathcal{G}_m \), and \( \tau \) and \( \sigma_i \), \( i = 1, \ldots, m \), be strictly positive real numbers such that
\[
\tau \sum_{i=1}^{m} \sigma_i \| L_i \|^2 < 4.
\]
Furthermore, let \( (\lambda_n)_{n \geq 0} \) be a sequence in \( (0,2) \), \( (a_n)_{n \geq 0} \) a sequence in \( \mathcal{H} \), and \( (b_{i,n})_{n \geq 0} \) and \( (d_{i,n})_{n \geq 0} \) sequences in \( \mathcal{G}_i \) for all \( i = 1, \ldots, m \), and set
\[
\begin{align*}
p_{1,n} &= J_{rA} \left( x_n - \frac{\tau}{2} \sum_{i=1}^{m} L_i v_{i,n} + \tau z \right) + a_n, \\
w_{1,n} &= 2p_{1,n} - x_n, \\
&\quad \text{for } i = 1, \ldots, m \\
p_{2,i,n} &= J_{\sigma_i B_i^{-1}} \left( v_{i,n} + \frac{\sigma_i}{2} L_i w_{1,n} - \sigma_i t_i \right) + b_{i,n}, \\
w_{2,i,n} &= 2p_{2,i,n} - v_{i,n}, \\
z_{1,n} &= w_{1,n} - \frac{\tau}{2} \sum_{i=1}^{m} L_i^* w_{2,i,n}, \\
x_{n+1} &= x_n + \lambda_n (z_{1,n} - p_{1,n}), \\
&\quad \text{for } i = 1, \ldots, m \\
z_{2,i,n} &= J_{\sigma_i D_i^{-1}} \left( w_{2,i,n} + \frac{\sigma_i}{2} L_i (2z_{1,n} - w_{1,n}) \right) + d_{i,n}, \\
v_{i,n+1} &= v_{i,n} + \lambda_n (z_{2,i,n} - p_{2,i,n}).
\end{align*}
\]

**Theorem 3.21** For Problem 3.18, assume that
\[
z \in \text{ran} \left( A + \sum_{i=1}^{m} L_i^*(B_i \square D_i)(L_i \cdot r_i) \right), \tag{3.45}
\]
and consider the sequences generated by Algorithm 3.20.

(i) If
\[
\sum_{n=0}^{\infty} \lambda_n \| a_n \|_H < +\infty, \quad \sum_{n=0}^{\infty} \lambda_n(\| d_{i,n} \|_{\mathcal{G}_i} + \| b_{i,n} \|_{\mathcal{G}_i}) < +\infty, \quad i = 1, \ldots, m,
\]
and \( \sum_{n=0}^{\infty} \lambda_n (2 - \lambda_n) = +\infty \), then
(a) \((x_n, v_{1,n}, \ldots, v_{m,n})_{n \geq 0}\) converges weakly to an element \((\overline{v}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m\) such that, when setting
\[
\overline{p}_1 = J_{\mathcal{H}} \left( \overline{v} - \frac{\tau}{2} \sum_{i=1}^m L_i^* \overline{v}_i + \tau z \right),
\]
and \(\overline{p}_{2,i} = J_{\mathcal{G}_i} \left( \overline{v}_i + \frac{\sigma_i}{2} L_i (2 \overline{p}_1 - \overline{v}) - \sigma_i r_i \right), \quad i = 1, \ldots, m,
\]
the element \((\overline{p}_1, \overline{p}_{2,1}, \ldots, \overline{p}_{2,m})\) is a primal-dual solution to Problem 3.18.

(b) \(\lambda_n(z,_{1,n} - p_{1,n}) \to 0 (n \to +\infty)\) and \(\lambda_n(z,_{2,i,n} - p_{2,i,n}) \to 0 (n \to +\infty)\) for \(i = 1, \ldots, m,\)

(c) whenever \(\mathcal{H}\) and \(\mathcal{G}_i, i = 1, \ldots, m,\) are finite-dimensional Hilbert spaces, \(a_n \to 0 (n \to +\infty)\) and \(b_{i,n} \to 0 (n \to +\infty)\) for \(i = 1, \ldots, m,\) then \((p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n})_{n \geq 0}\) converges strongly to a primal-dual solution of Problem 3.18.

(ii) If
\[
\sum_{n=0}^{+\infty} \|a_n\|_H < +\infty, \quad \sum_{n=0}^{+\infty} (\|d_{i,n}\|_{\mathcal{G}_i} + \|b_{i,n}\|_{\mathcal{G}_i}) < +\infty, \quad i = 1, \ldots, m, \quad \inf_{n \geq 0} \lambda_n > 0,
\]
and \(A\) and \(B_i^{-1}, \quad i = 1, \ldots, m,\) are uniformly monotone,
then \((p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n})_{n \geq 0}\) converges strongly to the unique primal-dual solution of Problem 3.18.

Proof. We consider the Hilbert direct sum
\[
\mathcal{K} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \ldots \oplus \mathcal{G}_m,
\]
being endowed with scalar product and norm, respectively defined in (1.2). Consider the set-valued operator
\[
\mathcal{M} : \mathcal{K} \to 2^\mathcal{K}, \quad (x, v_1, \ldots, v_m) \mapsto (-z + Ax, r_1 + B_1^{-1}v_1, \ldots, r_m + B_m^{-1}v_m),
\]
which is maximally monotone, since \(A\) and \(B_i, \quad i = 1, \ldots, m,\) are maximally monotone (cf. [11, Propositions 20.22 and 20.23]) and the bounded linear operator
\[
\mathcal{S} : \mathcal{K} \to \mathcal{K}, \quad (x, v_1, \ldots, v_m) \mapsto \left( \sum_{i=1}^m L_i^* v_i, -L_1 x, \ldots, -L_m x \right),
\]
which proves to be \(skew\) (i.e., \(\mathcal{S}^* = -\mathcal{S}\)) and hence maximally monotone (cf. [11, Example 20.30]). Further, consider the set-valued operator
\[
\mathcal{Q} : \mathcal{K} \to 2^\mathcal{K}, \quad (x, v_1, \ldots, v_m) \mapsto \left( 0, D_1^{-1} v_1, \ldots, D_m^{-1} v_m \right),
\]
which is maximally monotone, as well, since \(D_i\) is maximally monotone for \(i = 1, \ldots, m.\) Therefore, since \(\text{dom} \mathcal{S} = \mathcal{K},\) both \(\frac{1}{2} \mathcal{S} + \mathcal{Q}\) and \(\frac{1}{2} \mathcal{S} + \mathcal{M}\) are maximally monotone (cf. [11, Corollary 24.4(i)]). On the other hand, according to [58, Eq.
(3.12), it holds that \((3.45) \iff \operatorname{zer}(M + S + Q) \neq \emptyset\), while [58, Eq. (3.21) and (3.22)] yield
\[
(x, v_1, \ldots, v_m) \in \operatorname{zer}(M + S + Q)
\]
\[
\Rightarrow (x, v_1, \ldots, v_m) \text{ is a primal-dual solution to Problem 3.18.} \tag{3.46}
\]

We additionally introduce the bounded linear operator
\[
V : K \to K, \quad (x, v_1, \ldots, v_m) \mapsto \left( \frac{x}{\tau} - \frac{1}{2} \sum_{i=1}^{m} L_i v_i, \frac{v_1}{\sigma_1} - \frac{1}{2} L_1 x, \ldots, \frac{v_m}{\sigma_m} - \frac{1}{2} L_m x \right).
\]

It is a simple calculation to prove that \(V\) is self-adjoint, i.e., \(V^* = V\). Furthermore, the operator \(V\) is \(\rho\)-strongly positive for
\[
\rho = \left(1 - \frac{1}{2} \sqrt{\frac{\tau}{\sum_{i=1}^{m} \sigma_i \|L_i\|^2}} \right) \min \left\{ \frac{1}{\tau}, \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_m} \right\},
\]
which is a positive real number due to the assumption
\[
\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 4 \tag{3.47}
\]
made in Algorithm 3.20. Indeed, using that \(2ab \leq a \alpha^2 + b^2\) for any \(a, b \in \mathbb{R}\) and any \(\alpha \in \mathbb{R}_{++}\), it yields for each \(i = 1, \ldots, m\)
\[
2 \|L_i\| \|x\|_{\mathcal{H}} \|v_i\|_{\mathcal{G}_i} \leq \frac{\sigma_i \|L_i\|^2}{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}} \|x\|_{\mathcal{H}}^2 + \frac{\sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}}{\sigma_i} \|v_i\|_{\mathcal{G}_i}^2, \tag{3.48}
\]
and, consequently, for each \(x = (x, v_1, \ldots, v_m) \in \mathcal{K}\), it follows that
\[
\langle x, Vx \rangle_{\mathcal{K}} = \frac{\|x\|_{\mathcal{H}}^2}{\tau} + \sum_{i=1}^{m} \frac{\|v_i\|_{\mathcal{G}_i}^2}{\sigma_i} - \sum_{i=1}^{m} \langle L_i x, v_i \rangle_{\mathcal{G}_i}
\]
\[
\geq \frac{\|x\|_{\mathcal{H}}^2}{\tau} + \sum_{i=1}^{m} \frac{\|v_i\|_{\mathcal{G}_i}^2}{\sigma_i} - \sum_{i=1}^{m} \|L_i\| \|x\|_{\mathcal{H}} \|v_i\|_{\mathcal{G}_i}
\]
\[
\geq \left(1 - \frac{1}{2} \sqrt{\frac{\tau}{\sum_{i=1}^{m} \sigma_i \|L_i\|^2}} \right) \left( \frac{\|x\|_{\mathcal{H}}^2}{\tau} + \sum_{i=1}^{m} \frac{\|v_i\|_{\mathcal{G}_i}^2}{\sigma_i} \right)
\]
\[
\geq \left(1 - \frac{1}{2} \sqrt{\frac{\tau}{\sum_{i=1}^{m} \sigma_i \|L_i\|^2}} \right) \min \left\{ \frac{1}{\tau}, \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_m} \right\} \|x\|^2_{\mathcal{K}}
\]
\[
= \rho \|x\|^2_{\mathcal{K}}. \tag{3.49}
\]

Since \(V\) is \(\rho\)-strongly positive, we have \(\overline{\operatorname{cl}(\operatorname{ran} V)} = \operatorname{ran} V\) (cf. [11, Fact 2.19]), \(\ker V = \{0\}\) and, as \((\operatorname{ran} V)^\perp = \ker V^* = \ker V = \{0\}\) (see, for instance, [11, Fact 2.18]), it holds that \(\operatorname{ran} V = \mathcal{K}\). Consequently, \(V^{-1}\) exists and \(\|V^{-1}\| \leq \frac{1}{\rho}\).
3.2 A Douglas–Rachford type primal-dual method

The algorithmic scheme (3.44) is equivalent to

\[
\begin{aligned}
\frac{x_{n+1} - p_{i,n}}{\tau} - \frac{1}{2} \sum_{i=1}^{m} L_i^* v_{i,n} &\in A(p_{1,n} - a_n) - z - \frac{a_n}{\sigma}, \\
 w_{1,n} &= 2p_{1,n} - x_n, \quad \text{For } i = 1, \ldots, m \\
\frac{v_{i,n} - p_{2,i,n}}{\sigma_i} - \frac{1}{2} L_i(x_n - p_{1,n}) &\in -\frac{1}{2} L_i p_{1,n} + B_i^{-1}(p_{2,i,n} - b_{i,n}) + r_i - \frac{b_{i,n}}{\sigma_i}, \\
 w_{2,i,n} &= 2p_{2,i,n} - v_{i,n}, \quad \text{For } i = 1, \ldots, m \\
 x_{n+1} &= x_n + \lambda_n(z_{1,n} - p_{1,n}), \\
 w_{1,n} - z_{1,n} &\in -\frac{1}{2} \sum_{i=1}^{m} L_i^* w_{2,i,n} = 0, \\
 v_{i,n+1} &= v_{i,n} + \lambda_n(z_{2,i,n} - p_{2,i,n}).
\end{aligned}
\] (3.50)

We introduce for every \( n \geq 0 \) the following notations:

\[
\begin{aligned}
x_n &= (x_n, v_{1,n}, \ldots, v_{m,n}), \\
y_n &= (p_{1,n}, p_{2,1,n}, \ldots, p_{2,m,n}), \\
w_n &= (w_{1,n}, w_{2,1,n}, \ldots, w_{2,m,n}), \\
z_n &= (z_{1,n}, z_{2,1,n}, \ldots, z_{2,m,n}), \\
d_n &= (0, d_{1,n}, \ldots, d_{m,n}), \\
d_n^\sigma &= (0, d_{1,n}^\sigma, \ldots, d_{m,n}^\sigma), \\
b_n &= (a_n, b_{1,n}, \ldots, b_{m,n}), \\
b_n^\sigma &= (a_n^\sigma, b_{1,n}^\sigma, \ldots, b_{m,n}^\sigma).
\end{aligned}
\] (3.51)

The scheme (3.50) can equivalently be written in the form

\[
\begin{aligned}
\forall n \geq 0 &\quad V(x_n - y_n) \in \left(\frac{1}{2} S + M\right)(y_n - b_n) + \frac{1}{2} S b_n - b_n^\sigma, \\
w_n &= 2y_n - x_n, \\
V(w_n - z_n) \in \left(\frac{1}{2} S + Q\right)(z_n - d_n) + \frac{1}{2} S d_n - d_n^\sigma, \\
x_{n+1} &= x_n + \lambda_n(z_n - y_n).
\end{aligned}
\] (3.52)

We set for every \( n \geq 0 \)

\[
\begin{aligned}
e_n^b &= V^{-1}\left(\left(\frac{1}{2} S + V\right)b_n - b_n^\sigma\right), \\
e_n^d &= V^{-1}\left(\left(\frac{1}{2} S + V\right)d_n - d_n^\sigma\right).
\end{aligned}
\] (3.53)

Next, we introduce the Hilbert space \( \mathcal{K}_V \) with inner product and norm respectively defined, for \( x, y \in \mathcal{K} \), as

\[
\langle x, y \rangle_{\mathcal{K}_V} = \langle x, V y \rangle_{\mathcal{K}} \quad \text{and} \quad \|x\|_{\mathcal{K}_V} = \sqrt{\langle x, V x \rangle_{\mathcal{K}}}. \] (3.54)

Since \( \frac{1}{2} S + M \) and \( \frac{1}{2} S + Q \) are maximally monotone on \( \mathcal{K} \), the operators

\[
B := V^{-1}\left(\frac{1}{2} S + M\right) \quad \text{and} \quad A := V^{-1}\left(\frac{1}{2} S + Q\right)\] (3.55)
are maximally monotone on \( \mathcal{K}_V \). Moreover, since \( V \) is self-adjoint and \( \rho \)-strongly positive, one can easily see that weak and strong convergence in \( \mathcal{K}_V \) are equivalent with weak and strong convergence in \( \mathcal{K} \), respectively.
Now, taking into account (3.52), for every \( n \geq 0 \), we have

\[
V(x_n - y_n) \in \left( \frac{1}{2} S + M \right) (y_n - b_n) + \frac{1}{2} S b_n - b_n^* \\
\iff V x_n \in \left( V + \frac{1}{2} S + M \right) (y_n - b_n) + \left( \frac{1}{2} S + V \right) b_n - b_n^* \\
\iff x_n \in \left( \text{Id} + V^{-1} \left( \frac{1}{2} S + M \right) \right) (y_n - b_n) + V^{-1} \left( \frac{1}{2} S + V \right) b_n - b_n^* \\
\iff y_n = \left( \text{Id} + V^{-1} \left( \frac{1}{2} S + M \right) \right)^{-1} (x_n - e_n^b) + b_n \\
\iff y_n = (\text{Id} + B)^{-1} (x_n - e_n^b) + b_n
\]

(3.56)

and

\[
V(w_n - z_n) \in \left( \frac{1}{2} S + Q \right) (z_n - d_n) + \frac{1}{2} S d_n - d_n^* \\
\iff z_n = \left( \text{Id} + V^{-1} \left( \frac{1}{2} S + Q \right) \right)^{-1} (w_n - e_n^d) + d_n \\
\iff z_n = (\text{Id} + A)^{-1} (w_n - e_n^d) + d_n.
\]

(3.57)

Thus, the iterative rules in (3.52) become

\[
(\forall n \geq 0) \quad \begin{cases} 
  y_n = J_B \left( x_n - e_n^b \right) + b_n, \\
  z_n = J_A \left( 2y_n - x_n - e_n^d \right) + d_n, \\
  x_{n+1} = x_n + \lambda_n (z_n - y_n).
\end{cases}
\]

(3.58)

In addition, we have

\[
\text{zer} \left( A + B \right) = \text{zer} \left( V^{-1} \left( M + S + Q \right) \right) = \text{zer} \left( M + S + Q \right).
\]

By defining for every \( n \geq 0 \)

\[
\beta_n = J_B \left( x_n - e_n^b \right) - J_B (x_n) + b_n \text{ and } \alpha_n = J_A \left( 2y_n - x_n - e_n^d \right) - J_A (2y_n - x_n) + d_n,
\]

the iterative scheme (3.58) becomes

\[
(\forall n \geq 0) \quad \begin{cases} 
  y_n = J_B (x_n) + \beta_n, \\
  z_n = J_A \left( 2y_n - x_n \right) + \alpha_n, \\
  x_{n+1} = x_n + \lambda_n (z_n - y_n).
\end{cases}
\]

(3.59)

Thus, it has the structure of an error-tolerant Douglas–Rachford algorithm (see [54]).

(i) The assumptions made on the error sequences yield

\[
\sum_{n=0}^{+\infty} \lambda_n \| d_n \|_K < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n \| d_n^* \|_K < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n \| b_n \|_K < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n \| b_n^* \|_K < +\infty,
\]

(3.60)

and, by the boundedness of \( V^{-1}, S, \) and \( V \), it follows

\[
\sum_{n=0}^{+\infty} \lambda_n \| e_n^b \|_K < +\infty \text{ and } \sum_{n=0}^{+\infty} \lambda_n \| e_n^d \|_K < +\infty.
\]

(3.61)
Further, by making use of the nonexpansiveness of the resolvents, the error sequences \((\alpha_n)_{n \geq 0}\) and \((\beta_n)_{n \geq 0}\) satisfy

\[
\sum_{n=0}^{+\infty} \lambda_n \|\alpha_n\|_K + \|\beta_n\|_K \leq \sum_{n=0}^{+\infty} \lambda_n \left[ \|J_A (2y_n - x_n - e_n^d) - JA (2y_n - x_n)\|_K + \|d_n\|_K + \|J_B (x_n - e_n^b) - JB (x_n)\|_K + \|b_n\|_K \right]
\]

\[
\leq \sum_{n=0}^{+\infty} \lambda_n \left[ \|e_n^d\|_K + \|d_n\|_K + \|e_n^b\|_K + \|b_n\|_K \right] < +\infty.
\]

By the linearity and boundedness of \(V\), it further follows that

\[
\sum_{n=0}^{+\infty} \lambda_n \|[\alpha_n]_K + \|\beta_n\|_K < +\infty.
\]

(i)(a) According to [54, Theorem 2.1(i)(a)], the sequence \((x_n)_{n \geq 0}\) converges weakly in \(K_V\) and, consequently, in \(K\) to an element \(x \in \text{Fix}(R_AR_B)\) with \(J_Bx \in \text{zer}(A + B)\). The claim follows by identifying \(J_Bx\) and by noting (3.46).

(i)(b) According to [54, Theorem 2.1(i)(b)], it follows that \((R_AR_Bx_n - x_n) \to 0\) \((n \to +\infty)\). From (3.59), it follows that for every \(n \geq 0\)

\[
\lambda_n(z_n - y_n) = \frac{\lambda_n}{2} (R_A(R_Bx_n + 2\beta_n) - x_n + 2\alpha_n),
\]

and thus, by taking into consideration the nonexpansiveness of the reflected resolvent and the boundedness of \((\lambda_n)_{n \geq 0}\), it yields

\[
\|\lambda_n(z_n - y_n)\|_K \leq \frac{\lambda_n}{2} \|R_AR_Bx_n - x_n\|_K + \frac{\lambda_n}{2} \|R_A(R_Bx_n + 2\beta_n) - R_A(R_Bx_n) + 2\alpha_n\|_K
\]

\[
\leq \|R_AR_Bx_n - x_n\|_K + \lambda_n \|[\alpha_n]_K + \|\beta_n\|_K \|K_V\|.
\]

The claim follows by taking into account that \(\lambda_n \|[\alpha_n]_K + \|\beta_n\|_K \|K_V\| \to 0\) \((n \to +\infty)\).

(i)(c) As shown in (a), we have that \(x_n \to x \in \text{Fix}(R_AR_B)\) \((n \to +\infty)\) with \(J_Bx \in \text{zer}(A + B) = \text{zer}(M + S + Q)\). Moreover, by the assumptions and (3.51), we have \(b_n \to 0\) \((n \to +\infty)\), hence by (3.53), it holds that \(e_n^b \to 0\) \((n \to +\infty)\) and therefore \(\beta_n \to 0\) \((n \to +\infty)\). In conclusion, by the continuity of \(J_B\) and (3.59), we have

\[
y_n = J_B(x_n) + \beta_n \to J_Bx \in \text{zer}(M + S + Q) \quad (n \to +\infty).
\]

(ii) The assumptions made on the error sequences yield

\[
\sum_{n=0}^{+\infty} \|d_n\|_K < +\infty, \quad \sum_{n=0}^{+\infty} \|d_n^r\|_K < +\infty, \quad \sum_{n=0}^{+\infty} \|b_n\|_K < +\infty, \quad \sum_{n=0}^{+\infty} \|b_n^r\|_K < +\infty,
\]

thus,

\[
\sum_{n=0}^{+\infty} \|e_n^b\|_K < +\infty \quad \text{and} \quad \sum_{n=0}^{+\infty} \|e_n^d\|_K < +\infty.
\]
This implies that
\[ \sum_{n=0}^{+\infty} [\|\alpha_n\|_K + \|\beta_n\|_K] < +\infty \]
which, due to the linearity and boundedness of \(V\), further yields
\[ \sum_{n=0}^{+\infty} [\|\alpha_n\|_{K_V} + \|\beta_n\|_{K_V}] < +\infty. \]

Since \(A\) and \(B_i^{-1}\), \(i = 1, \ldots, m\), are uniformly monotone, there exist increasing functions \(\phi_A : \mathbb{R}_+ \to [0, +\infty]\) and \(\phi_{B_i^{-1}} : \mathbb{R}_+ \to [0, +\infty]\), \(i = 1, \ldots, m\), vanishing only at 0, such that
\[ \langle x - y, u - z \rangle \geq \phi_A (\|x - y\|_H) \quad \forall (x, u), (y, z) \in \text{gra} A, \]
\[ \langle v - w, p - q \rangle \geq \phi_{B_i^{-1}} (\|v - w\|_{g_i}) \quad \forall (v, p), (w, q) \in \text{gra} B_i^{-1} \forall i = 1, \ldots, m. \] (3.62)

By Proposition 3.19, the function \(\phi_M : \mathbb{R}_+ \to [0, +\infty]\),
\[ \phi_M (c) = \inf \left\{ \phi_A (a) + \sum_{i=1}^{m} \phi_{B_i^{-1}} (b_i) : a^2 + \sum_{i=1}^{m} b_i^2 = c, a, b_i \in \mathbb{R}_+, i = 1, \ldots, m \right\}, \] (3.63)
is increasing and vanishes only at 0, and it fulfills for each \((x, u), (y, z) \in \text{gra} M\)
\[ \langle x - y, u - z \rangle_K \geq \phi_M (\|x - y\|_K). \] (3.64)

Thus, \(M\) is uniformly monotone on \(K\).

The function \(\phi_B : \mathbb{R}_+ \to [0, +\infty]\), \(\phi_B (t) = \phi_M \left( \frac{1}{\sqrt{\|V\|}} t \right)\), is increasing and vanishes only at 0. Let \((x, u), (y, z) \in \text{gra} B\). Then there exist \(v \in Mx\) and \(w \in My\) fulfilling \(Vu = \frac{1}{2} Sx + v\) and \(Vz = \frac{1}{2} Sy + w\), and it holds
\[ \langle x - y, u - z \rangle_{K_V} = \langle x - y, Vu - Vz \rangle_K \]
\[ = \langle x - y, \left( \frac{1}{2} Sx + v \right) - \left( \frac{1}{2} Sy + w \right) \rangle_K \]
\[ \geq \phi_M (\|x - y\|_K) \]
\[ \geq \phi_M \left( \frac{1}{\sqrt{\|V\|}} \|x - y\|_{K_V} \right) \]
\[ = \phi_B (\|x - y\|_{K_V}). \] (3.65)

Consequently, \(B\) is uniformly monotone on \(K_V\) and, according to [54, Theorem 2.1(ii)(b)], \((J_B x_n)_{n \geq 0}\) converges strongly to the unique element \(\bar{y} \in \text{zer} (A + B) = \text{zer} (M + S + Q)\). In the light of (3.59) and by using that \(\beta_n \to 0\) \((n \to +\infty)\), it follows that \(y_n \to \bar{y}\) \((n \to +\infty)\). \(\square\)

**Remark 3.22** In the sequel we summarize some facts concerning Algorithm 3.20 and Theorem 3.21.
Algorithm 3.20 is a fully decomposable iterative method, as each of the operators occurring in Problem 3.18 is processed individually. Moreover, a considerable number of steps in (3.44) can be executed in parallel.

(ii) The proof of Theorem 3.21, which states the convergence of Algorithm 3.20, relies on the reformulation of the iterative scheme as an inexact Douglas–Rachford method in a specific real Hilbert space. For the use of a similar technique in the context of a forward-backward-type method we refer to [122].

(iii) We would like to notice that the assumption \(\sum_{n=0}^{+\infty} \lambda_n \|a_n\|_H < +\infty\) does not necessarily imply that \((\|a_n\|_H)_{n\geq 0}\) is summable or that \((a_n)_{n\geq 0}\) (weakly or strongly) converges to 0 as \(n \to +\infty\). We refer to [54, Remark 2.2(iii)] for further considerations on the conditions imposed on the error sequences in Theorem 3.21.

Remark 3.23 In the following we emphasize the relations between the proposed algorithm and other existent primal-dual iterative schemes.

(i) Other iterative methods for solving the primal-dual monotone inclusion pair introduced in Problem 3.18 were given in [58] and [122] for \(D_i^{-1}, i = 1, \ldots, m\), monotone Lipschitzian and cocoercive operators, respectively. Different to the approach proposed in this subsection, there, the operators \(D_i^{-1}, i = 1, \ldots, m\), are processed within some forward steps.

(ii) When for every \(i = 1, \ldots, m\) one takes \(D_1(0) = \mathcal{G}_1\) and \(D_i(v) = \emptyset\) for all \(v \in \mathcal{G}_i \setminus \{0\}\), the algorithms proposed in [58, Theorem 3.1] (see, also, [43, Theorem 3.1] for the case \(m = 1\)) and [122, Theorem 3.1] applied to Problem 3.18 differ from Algorithm 3.20.

(iii) When solving the particular case of a primal-dual pair of convex optimization problems discussed in Example 3.5, one can, for example, make use of the iterative schemes provided in [60, Algorithm 3.1] and [48, Algorithm 1]. Let us notice that particularizing Algorithm 3.20 to this framework gives rise to a numerical scheme different to the ones in the mentioned literature.

3.2.3 A second primal-dual algorithm

In Algorithm 3.20, each operator \(L_i\) and its adjoint \(L_i^*\), \(i = 1, \ldots, m\), is processed two times. However, for largescale optimization problems these matrix-vector multiplications may be expensive compared with the computation of the resolvents of the operators \(A, B_i^{-1}\), and \(D_i^{-1}, i = 1, \ldots, m\).

The second primal-dual algorithm we propose for solving the monotone inclusions in Problem 3.18 has the particularity that it evaluates each operator \(L_i\) and its adjoint \(L_i^*\), \(i = 1, \ldots, m\), only once.

Algorithm 3.24 Let \(x_0 \in \mathcal{H}\), \((y_{1,0}, \ldots, y_{m,0}) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m\), \((v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m\), and \(\tau\) and \(\sigma_i, i = 1, \ldots, m\), be strictly positive real numbers such that
\[
\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < \frac{1}{4}.
\]

Furthermore, let \(\gamma_i \leq 2\sigma_i^{-1}\tau \sum_{j=1}^{m} \sigma_j \|L_j\|^2, i = 1, \ldots, m\), let \((\lambda_n)_{n\geq 0}\) be a sequence in \((0, 2)\), \((a_n)_{n\geq 0}\) a sequence in \(\mathcal{H}\), and \((b_{i,n})_{n\geq 0}\) and \((d_{i,n})_{n\geq 0}\) sequences in \(\mathcal{G}_i\) for all
\textit{Theorem 3.25} For Problem 3.18, suppose that

\[ z \in \text{ran} \left( A + \sum_{i=1}^{m} L_i^* (B_i \Box D_i) (L_i \cdot -r_i) \right), \]  

and consider the sequences generated by Algorithm 3.24.

\begin{enumerate}[i]
    \item If
    \begin{align*}
    \sum_{n=0}^{+\infty} \lambda_n \|a_n\|_H < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n (\|d_{i,n}\|_{\mathcal{G}_i} + \|b_{i,n}\|_{\mathcal{G}_i}) < +\infty, & \quad i = 1, \ldots, m, \\
    \end{align*}
    and \( \sum_{n=0}^{+\infty} \lambda_n (2 - \lambda_n) = +\infty \), then
    
    \begin{enumerate}[a]
        \item \( (x_n, y_{1,n}, \ldots, y_{m,n}, v_{1,n}, \ldots, v_{m,n})_{n \geq 0} \) converges weakly to an element \((\overline{x}, \overline{y}_1, \ldots, \overline{y}_m, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m \) such that \((\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \) is a primal-dual solution to Problem 3.18,
    \item \( \lambda_n (p_{1,n} - x_n) \to 0 \ (n \to +\infty), \lambda_n (p_{2,i,n} - y_{i,n}) \to 0 \ (n \to +\infty), \) and \( \lambda_n (p_{3,i,n} - v_{i,n}) \to 0 \ (n \to +\infty) \) for \( i = 1, \ldots, m, \)
    \item whenever \( \mathcal{H} \) and \( \mathcal{G}_i, \ i = 1, \ldots, m, \) are finite-dimensional Hilbert spaces, \( (x_n, v_{1,n}, \ldots, v_{m,n})_{n \geq 0} \) converges strongly to a primal-dual solution of Problem 3.18.
    \end{enumerate}

    \item If
    \begin{align*}
    \sum_{n=0}^{+\infty} \|a_n\|_H < +\infty, \quad \sum_{n=0}^{+\infty} (\|d_{i,n}\|_{\mathcal{G}_i} + \|b_{i,n}\|_{\mathcal{G}_i}) < +\infty, & \quad i = 1, \ldots, m, \\
    \inf_{n \geq 0} \lambda_n > 0, \\
    \end{align*}
    and \( A, B_i^{-1}, \) and \( D_i, i = 1, \ldots, m, \) are uniformly monotone,
    then \( (p_{1,n}, p_{3,1,n}, \ldots, p_{3,m,n})_{n \geq 0} \) converges strongly to the unique primal-dual solution of Problem 3.18.
\end{enumerate}

\textbf{Proof.} We let \( \mathcal{G} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m \) and consider the Hilbert direct sum

\[ \mathcal{K} = \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G} \]

being endowed with scalar product and norm, respectively defined in (1.2). In what follows, we set

\[ y = (y_1, \ldots, y_m), \quad v = (v_1, \ldots, v_m), \quad y = (\overline{y}_1, \ldots, \overline{y}_m), \quad v = (\overline{v}_1, \ldots, \overline{v}_m). \]
Consider the set-valued operator
\[ M : \mathcal{K} \to 2^{\mathcal{K}}, \quad (x, y, v) \mapsto (-z + Ax, D_1 y_1, \ldots, D_m y_m, r_1 + B_1^{-1} v_1, \ldots, r_m + B_m^{-1} v_m), \]
which is maximally monotone, since \( A, B_i, \) and \( D_i, i = 1, \ldots, m, \) are maximally monotone (cf. [11, Propositions 20.22 and 20.23]). Furthermore, consider the bounded linear operator
\[ S : \mathcal{K} \to \mathcal{K}, \quad (x, y, v) \mapsto \left( \sum_{i=1}^{m} L_i^* v_i, -v_1, \ldots, -v_m, -L_1 x + y_1, \ldots, -L_m x + y_m \right), \]
which proves to be skew (i.e., \( S^* = -S \)) and hence maximally monotone (cf. [11, Example 20.30]). Since \( \text{dom} S = \mathcal{K}, \) the sum \( M + S \) is maximally monotone, as well (cf. [11, Corollary 24.4(i)]). Moreover, we have

\[ (3.67) \iff (\exists \, x \in \mathcal{H}) \ z \in Ax + \sum_{i=1}^{m} L_i^* (B_i \square D_i) (L_i x - r_i) \]
\[ \iff (\exists (x, v) \in \mathcal{H} \times \mathcal{G}) \ \left\{ \begin{array}{l} z \in Ax + \sum_{i=1}^{m} L_i^* v_i \\ v_i \in (B_i \square D_i) (L_i x - r_i), \ i = 1, \ldots, m, \end{array} \right. \]
\[ \iff (\exists (x, v) \in \mathcal{H} \times \mathcal{G}) \ \left\{ \begin{array}{l} z \in Ax + \sum_{i=1}^{m} L_i^* v_i \\ L_i x - r_i \in B_i^{-1} v_i + D_i^{-1} v_i, \ i = 1, \ldots, m, \end{array} \right. \]
\[ \iff (\exists (x, y, v) \in \mathcal{K}) \ \left\{ \begin{array}{l} 0 \in -z + Ax + \sum_{i=1}^{m} L_i^* v_i \\ 0 \in D_i y_i - v_i, \ i = 1, \ldots, m \\ 0 \in r_i + B_i^{-1} v_i - L_i x + y_i, \ i = 1, \ldots, m, \end{array} \right. \]
\[ \iff (\exists (x, y, v) \in \mathcal{K}) \ (0, \ldots, 0) \in (M + S) (x, y, v) \]
\[ \iff \text{zer} (M + S) \neq \emptyset. \] (3.68)

From the above calculations, it follows that
\[ (x, y, v) \in \text{zer} (M + S) \Rightarrow \left\{ \begin{array}{l} z - \sum_{i=1}^{m} L_i^* v_i \in Ax, \\ v_i \in (B_i \square D_i) (L_i x - r_i), \ i = 1, \ldots, m, \end{array} \right. \]
\[ \iff (x, v_1, \ldots, v_m) \text{ is a primal-dual solution to Problem 3.18.} \] (3.69)

We also introduce the bounded linear operator
\[ V : \mathcal{K} \to \mathcal{K}, \]
\[ (x, y, v) \mapsto \left( \frac{x - \sum_{i=1}^{m} L_i^* v_i}{\gamma_1}, \frac{y_1}{\gamma_1} + v_1, \ldots, \frac{y_m}{\gamma_m} + v_m, \frac{v_1}{\gamma_1} - L_1 x + y_1, \ldots, \frac{v_m}{\gamma_m} - L_m x + y_m \right), \]
which is self-adjoint, i.e., \( V^* = V. \) In addition, the operator \( V \) is \( \rho \)-strongly positive for
\[ \rho = \left( 1 - 2 \sqrt{\frac{1}{\tau} \sum_{i=1}^{m} \sigma_i \| L_i \|^2} \right) \min \left\{ \frac{1}{\gamma_1}, \frac{1}{\gamma_2}, \ldots, \frac{1}{\gamma_m} \right\}, \]
which is a positive real number due to the assumption
\[ \tau \sum_{i=1}^{m} \sigma_i \| L_i \|^2 < \frac{1}{4} \] (3.70)
made in Algorithm 3.24. Indeed, for \( \gamma_i \leq 2\sigma_i^{-1}\tau \sum_{j=1}^{m} \sigma_j \|L_j\|^2 \), it yields for each \( i = 1, \ldots, m \),

\[
2 \langle L_ix - y_i, v_i \rangle_{G_i} \leq \frac{\sigma_i \|L_i\|^2 \|y_i\|_{\gamma_i}^2}{\sqrt{\tau \sum_{j=1}^{m} \sigma_j \|L_j\|^2}} + 2 \sqrt{\tau \sum_{j=1}^{m} \sigma_j \|L_j\|^2} \|y_i\|_{\gamma_i} + 2 \sqrt{\tau \sum_{j=1}^{m} \sigma_j \|L_j\|^2} \|v_i\|_{\gamma_i}^2
\]

and, consequently, for each \( x = (x, y, v) \in \mathcal{K} \), it follows that

\[
\langle x, Vx \rangle_{\mathcal{K}} = \frac{\|x\|^2}{\tau} + \sum_{i=1}^{m} \left[ \frac{\|y_i\|_{\gamma_i}^2}{\gamma_i} + \frac{\|v_i\|_{\gamma_i}^2}{\sigma_i} \right] - 2 \sum_{i=1}^{m} \langle L_ix - y_i, v_i \rangle_{G_i},
\]

\[
\geq 1 - 2 \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2} \min \left\{ \frac{1}{\tau}, \ldots, \frac{1}{\gamma_m}, \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_m} \right\} \|x\|^2_{\mathcal{K}}
\]

\[
= \rho \|x\|^2_{\mathcal{K}}.
\]

The algorithmic scheme (3.66) is equivalent to

\[
\begin{align*}
(x_n - p_{n,1}) - \sum_{i=1}^{m} L_i^*v_{i,n} &\in -z + A(p_{1,n} - a_n) - \frac{a_n}{\tau}, \\
x_{n+1} &\in x_n + \lambda_n(p_{1,n} - x_n),
\end{align*}
\]

\( (\forall n \geq 0) \)

\[
\begin{align*}
y_{i,n+1} &\in D_i(p_{2,i,n} - d_{i,n}) - \frac{d_{i,n}}{\gamma_i}, \\
y_{i,n} &\in y_{i,n} + \lambda_n(p_{2,i,n} - y_{i,n}), \\
v_{i,n+1} &\in v_{i,n} + \lambda_n(p_{3,i,n} - v_{i,n}),
\end{align*}
\]

\( (\forall n \geq 0) \)

We introduce for every \( n \geq 0 \) the following notations:

\[
\begin{align*}
\mathbf{x}_n &:= (x_n, y_{1,n}, \ldots, y_{m,n}, v_{1,n}, \ldots, v_{m,n}), \\
\mathbf{p}_n &:= (p_{1,n}, p_{1,n}, \ldots, p_{2,m,n}, p_{3,1,n}, \ldots, p_{3,m,n}), \\
\mathbf{a}_n &:= (a_n, d_{1,n}, \ldots, d_{m,n}, b_{1,n}, \ldots, b_{m,n}), \\
\mathbf{a}_\tau &:= (a_n, d_{1,n}, \ldots, d_{m,n}, b_{1,n}, \ldots, b_{m,n}).
\end{align*}
\]

By taking this into account, the scheme (3.72) can equivalently be written in the form

\[
\begin{align*}
\forall n \geq 0, \quad & V \mathbf{x}_n - \mathbf{p}_n \in (S + M)(\mathbf{p}_n - \mathbf{a}_n) + S\mathbf{a}_n - \mathbf{a}_\tau^n, \\
\mathbf{x}_{n+1} &\in \mathbf{x}_n + \lambda_n(\mathbf{p}_n - \mathbf{x}_n).
\end{align*}
\]

Considering again the Hilbert space \( \mathcal{K}_V \) with inner product and norm respectively defined as in (3.54), since \( V \) is self-adjoint and \( \rho \)-strongly positive, weak and strong convergence in \( \mathcal{K}_V \) are equivalent with weak and strong convergence in \( \mathcal{K} \), respectively. Moreover, \( A = V^{-1}(S + M) \) is maximally monotone on \( \mathcal{K}_V \). Thus, in the light of (3.56)–(3.57) and by denoting \( e_n = V^{-1}((S + V)\mathbf{a}_n - \mathbf{a}_\tau^n) \) for every \( n \geq 0 \), the iterative scheme (3.74) becomes

\[
\begin{align*}
\forall n \geq 0, \quad & \mathbf{p}_n = J_A(\mathbf{x}_n - \mathbf{e}_n) + \mathbf{a}_n, \\
\mathbf{x}_{n+1} &\in \mathbf{x}_n + \lambda_n(\mathbf{p}_n - \mathbf{x}_n).
\end{align*}
\]
Furthermore, introducing the maximal monotone operator $B : K \to 2^K, x \mapsto \{0\}$, and defining for every $n \geq 0$
\[
\alpha_n = J_A(x_n - e_n) - J_A(x_n) + a_n,
\]
the iterative scheme (3.75) reads (notice that $J_B = \text{Id}$)
\[
(\forall n \geq 0) \quad \begin{cases} 
  y_n = J_B(x_n), \\
  p_n = J_A(2y_n - x_n) + \alpha_n, \\
  x_{n+1} = x_n + \lambda_n(p_n - y_n),
\end{cases}
\] (3.76)
thus, it has the structure of the error-tolerant Douglas–Rachford algorithm from [54].

Obviously, $\text{zer}(A + B) = \text{zer}(M + S)$.

(i) The assumptions made on the error sequences yield
\[
\sum_{n=0}^{+\infty} \lambda_n\|a_n\|_K < +\infty \quad \text{and} \quad \sum_{n=0}^{+\infty} \lambda_n\|e_n\|_K < +\infty.
\]
Thus, by the nonexpansiveness of the resolvent of $A$,
\[
\sum_{n=0}^{+\infty} \lambda_n\|\alpha_n\|_K < +\infty
\]
and, consequently, by the linearity and boundedness of $V$,
\[
\sum_{n=0}^{+\infty} \lambda_n\|\alpha_n\|_{K_V} < +\infty
\]

(i)(a) Follows directly from [54, Theorem 2.1(i)(a)] by using that $J_B = \text{Id}$ and relation (3.69).

(i)(b) Follows in analogy to the proof of Theorem 3.21(i)(b).

(i)(c) Follows from Theorem 3.25(i)(a).

(ii) The iterative scheme (3.75) can also be formulated as
\[
(\forall n \geq 0) \quad \begin{cases} 
  p_n = J_A(x_n) + \alpha_n, \\
  y_n = J_B(2p_n - x_n), \\
  x_{n+1} = x_n + \lambda_n(y_n - p_n),
\end{cases}
\] (3.77)
with the error sequence fulfilling
\[
\sum_{n=0}^{+\infty} \|\alpha_n\|_{K_V} < +\infty.
\]
The statement follows from [54, Theorem 2.1(ii)(b)] by taking into consideration the uniform monotonicity of $A$ and relation (3.69).

Remark 3.26 When for every $i = 1, \ldots, m$, one takes $D_i(0) = G_i$ and $D_i(v) = \emptyset$ for all $v \in G_i \setminus \{0\}$, and $(d_{i,n})_{n \geq 0}$ as a sequence of zeros, one can show that the assertions made in Theorem 3.25 hold true for step length parameters satisfying
\[
\tau \sum_{i=1}^{m} \sigma_i \|L_i\| < 1,
\]
when choosing $(y_{1,0}, \ldots, y_{m,0}) = (0, \ldots, 0)$ in Algorithm 3.24, since the sequences $(y_{1,n}, \ldots, y_{m,n})_{n \geq 0}$ and $(v_{1,n}, \ldots, v_{m,n})_{n \geq 0}$ vanish in this particular situation.
**Remark 3.27** In the following we emphasize the relations between Algorithm 3.24 and other existent primal-dual iterative schemes.

(i) When for every $i = 1, \ldots, m$, one takes $D_i(0) = G_i$ and $D_i(v) = \emptyset$ for all $v \in G_i \setminus \{0\}$, and $(d_i,n)_{n \geq 0}$ as a sequence of zeros, Algorithm 3.24 with $(y_1,0, \ldots, y_m,0) = (0, \ldots, 0)$ as initial choice provides an iterative scheme which is identical to the one in [122, Eq. (3.3)], but differs from the one in [58, Theorem 3.1] (see, also, [43, Theorem 3.1] for the case $m = 1$) when the latter are applied to Problem 3.18.

(ii) When solving the particular case of a primal-dual pair of convex optimization problems as the one discussed in Example 3.5, and when considering as initial choice $y_1,0 = 0$, Algorithm 3.24 gives rise to an iterative scheme which is equivalent to [60, Algorithm 3.1]. In addition, under the assumption of exact implementations, the method in Algorithm 3.24 equals the one in [48, Algorithm 1], our choice of $(\lambda_n)_{n \geq 0}$ to be variable in the interval $(0, 2)$, however, relaxes the assumption in [48] that $(\lambda_n)_{n \geq 0}$ is a constant sequence in $(0, 1]$.

(iii) In general, the statements concerning weak convergence of the iterates in Theorem 3.25(i)(a) cannot be strengthened to strong convergence without considering additional hypotheses on the operators as those described in Theorem 3.25(ii). Indeed, by taking $\lambda_n = 1$ in (3.75) and ignoring the error sequences, the scheme reduces to the proximal point algorithm

$$\forall n \geq 0 \mid x_{n+1} = J_A(x_n),$$

which was shown to converge weakly but not strongly (cf. [9, 79]).

### 3.2.4 Application to convex minimization problems

In this subsection we particularize the two iterative schemes introduced and investigated in Subsection 3.2.2 and Subsection 3.2.3 in the context of solving a primal-dual pair of convex optimization problems. To this end we consider the following problem.

**Problem 3.28** Let $\mathcal{H}$ be a real Hilbert space, let $z \in \mathcal{H}$, and let $f \in \Gamma(\mathcal{H})$. For every $i = 1, \ldots, m$, suppose that $G_i$ is a real Hilbert space, let $r_i \in G_i$, let $g_i, l_i \in \Gamma(G_i)$, and let $L_i : \mathcal{H} \to G_i$ be a nonzero bounded linear operator. Consider the convex optimization problem

$$(P) \quad \inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \square l_i)(L_i x - r_i) - \langle x, z \rangle \right\} \quad (3.78)$$

and its conjugate dual problem

$$(D) \quad \sup_{(v_1, \ldots, v_m) \in G_1 \times \ldots \times G_m} \left\{ -f^* \left( z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} (g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle) \right\}. \quad (3.79)$$

By taking into account the maximal monotone operators

$$A = \partial f, \quad B_i = \partial g_i, \text{ and } D_i = \partial l_i, \quad i = 1, \ldots, m,$$
the monotone inclusion problem (3.42) reads

$$\text{find } x \in \mathcal{H} \text{ such that } z \in \partial f(x) + \sum_{i=1}^{m} L_i^*(\partial g_i \square \partial l_i)(L_i x - r_i),$$

(3.80) while the dual inclusion problem (3.43) reads

$$\text{find } v_i \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - \sum_{i=1}^{m} L_i^* v_i \in \partial f(x), \\
 v_i \in (\partial g_i \square \partial l_i)(L_i x - r_i), \quad i = 1, \ldots, m. \end{cases}$$

(3.81)

If $$(x, v_1, \ldots, v_m) \in \mathcal{H} \oplus \mathcal{G}_1 \ldots \oplus \mathcal{G}_m$$ is a primal-dual solution to (3.80)-(3.81), namely,

$$z - \sum_{i=1}^{m} L_i^* v_i \in \partial f(x) \quad \text{and} \quad v_i \in (\partial g_i \square \partial l_i)(L_i x - r_i), \quad i = 1, \ldots, m,$$

(3.82) then $x$ is an optimal solution to $$(P),$$ $$(v_1, \ldots, v_m)$$ is an optimal solution to $$(D)$$ and the optimal objective values of the two problems coincide (thus, strong duality holds).

Combining this statement with Algorithm 3.20 and Theorem 3.21 gives rise to the following iterative scheme and convergence results for the primal-dual pair of optimization problems $$(P)-(D),$$ respectively. We are also making use of the fact that the subdifferential of a uniformly convex function is uniformly monotone (cf. [11, Example 22.3(iii)]).

**Algorithm 3.29** Let $$x_0 \in \mathcal{H}, (v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \times \ldots \times \mathcal{G}_m,$$ and $$\tau$$ and $$\sigma_i, \quad i = 1, \ldots, m,$$ be strictly positive real numbers such that

$$\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 4.$$ 

Furthermore, let $$(\lambda_n)_{n \geq 0}$$ be a sequence in $$(0,2),$$ $$(a_n)_{n \geq 0}$$ a sequence in $$\mathcal{H},$$ and $$(b_{t,n})_{n \geq 0}$$ and $$(d_{i,n})_{n \geq 0}$$ sequences in $$\mathcal{G}_i$$ for all $$i = 1, \ldots, m,$$ and set

$$\begin{align*}
-p_{1,n} &= \text{Prox}_{\tau f}(x_n - \frac{\tau}{2} \sum_{i=1}^{m} L_i^* v_{i,n} + \tau z) + a_n, \\
 w_{1,n} &= 2p_{1,n} - x_n, \\
 &\text{For } i = 1, \ldots, m \quad \begin{cases} p_{2,i,n} = \text{Prox}_{\sigma_i g_i^*}(v_{i,n} + \frac{\sigma_i}{2} L_i w_{1,n} - \sigma_i r_i) + b_{i,n}, \\
 w_{2,i,n} = 2p_{2,i,n} - v_{i,n}, \\
 z_{1,n} = w_{1,n} - \frac{\tau}{2} \sum_{i=1}^{m} L_i^* w_{2,i,n}, \\
 x_{n+1} = x_n + \lambda_n (z_{1,n} - p_{1,n}), \\
 &\text{For } i = 1, \ldots, m \quad \begin{cases} z_{2,i,n} = \text{Prox}_{\sigma_i l_i^*}(w_{2,i,n} + \frac{\sigma_i}{2} L_i (2z_{1,n} - w_{1,n})) + d_{i,n}, \\
 v_{i,n+1} = w_{2,i,n} + \lambda_n (z_{2,i,n} - p_{2,i,n}). \end{cases} \end{cases} \end{align*}$$

(3.83)

**Theorem 3.30** For Problem 3.28, suppose that

$$z \in \text{ran} \left(\partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \square \partial l_i)(L_i \cdot -r_i)\right),$$

(3.84) and consider the sequences generated by Algorithm 3.29.
(i) If
\[
\sum_{n=0}^{+\infty} \lambda_n \|a_n\|_H < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n (\|d_{i,n}\|_{G_i} + \|b_{i,n}\|_{G_i}) < +\infty, \quad i = 1, \ldots, m,
\]
and \(\sum_{n=0}^{+\infty} \lambda_n (2 - \lambda_n) = +\infty\), then

(a) \((x_n, v_{1,n}, \ldots, v_{m,n})_{n \geq 0}\) converges weakly to an element \((\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in H \times G_1 \times \ldots \times G_m\) such that, when setting
\[
\begin{align*}
p_1 &= \text{Prox}_{r f} \left(\bar{x} - \frac{\tau}{2} \sum_{i=1}^{m} L_i^* \bar{v}_i + \tau z\right), \\
\text{and } p_{2,i} &= \text{Prox}_{\sigma_i b_i^*} \left(\bar{v}_i + \frac{\sigma_i}{2} L_i (2 \bar{v}_i - \bar{x}) - \sigma_i r_i\right) \quad i = 1, \ldots, m,
\end{align*}
\]

\(p_1\) is an optimal solution to (P), \((p_{2,1}, \ldots, p_{2,m})\) is an optimal solution to (D)

and \(v(P) = v(D)\),

(b) \(\lambda_n (z_{1,n} - p_{1,n}) \to 0\) \((n \to +\infty)\) and \(\lambda_n (z_{2,i,n} - p_{2,i,n}) \to 0\) \((n \to +\infty)\) for \(i = 1, \ldots, m\),

(c) whenever \(H\) and \(G_i, i = 1, \ldots, m\), are finite-dimensional Hilbert spaces, \(a_n \to 0\) \((n \to +\infty)\) and \(b_{i,n} \to 0\) \((n \to +\infty)\) for \(i = 1, \ldots, m\), then \((p_{1,n})_{n \geq 0}\) converges strongly to an optimal solution to (P) and \((p_{2,1,n}, \ldots, p_{2,m,n})_{n \geq 0}\) converges strongly to an optimal solution to (D).

(ii) If
\[
\sum_{n=0}^{+\infty} \|a_n\|_H < +\infty, \quad \sum_{n=0}^{+\infty} (\|d_{i,n}\|_{G_i} + \|b_{i,n}\|_{G_i}) < +\infty, \quad i = 1, \ldots, m, \quad \inf_{n \geq 0} \lambda_n > 0,
\]

and \(f\) and \(g_i^*\), \(i = 1, \ldots, m\), are uniformly convex,

then \(v(P) = v(D)\), the sequence \((p_{1,n})_{n \geq 0}\) converges strongly to an optimal solution to (P), and \((p_{2,1,n}, \ldots, p_{2,m,n})_{n \geq 0}\) converges strongly to an optimal solution to (D).

Algorithm 3.24 and Theorem 3.25 give rise to the following iterative scheme and corresponding convergence results for the primal-dual pair of optimization problems (P)–(D).

**Algorithm 3.31** Let \(x_0 \in H, (y_{1,0}, \ldots, y_{m,0}) \in G_1 \times \ldots \times G_m, (v_{1,0}, \ldots, v_{m,0}) \in G_1 \times \ldots \times G_m\), and \(\tau\) and \(\sigma_i, i = 1, \ldots, m\), be strictly positive real numbers such that
\[
\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < \frac{1}{4}.
\]

Furthermore, let \(\gamma_i \leq 2\sigma_i^{-1} \tau \sum_{j=1}^{m} \sigma_j \|L_j\|^2, i = 1, \ldots, m\), \((\lambda_n)_{n \geq 0}\) be a sequence in \((0, 2)\), \((a_n)_{n \geq 0}\) a sequence in \(H\), and \((b_{i,n})_{n \geq 0}\) and \((d_{i,n})_{n \geq 0}\) sequences in \(G_i\) for all
Theorem 3.32 For Problem 3.28, suppose that
\[ z \in \text{ran} \left( \partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \Box \partial l_i)(L_i \cdot -r_i) \right), \]  
and consider the sequences generated by Algorithm 3.31.

(i) If
\[ \sum_{n=0}^{+\infty} \lambda_n \|a_n\|_\mathcal{H} < +\infty, \quad \sum_{n=0}^{+\infty} \lambda_n (\|d_{i,n}\|_{\mathcal{G}_i} + \|b_{i,n}\|_{\mathcal{G}_i}) < +\infty, \quad i = 1, \ldots, m, \]
and \( \sum_{n=0}^{+\infty} \lambda_n (2 - \lambda_n) = +\infty \), then
(a) \((x_n, y_{1,n}, \ldots, y_{m,n}, v_{1,n}, \ldots, v_{m,n})_{n \geq 0}\) converges weakly to an element \((\overline{x}, \overline{y}_1, \ldots, \overline{y}_m, \overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m\) such that \(\overline{x}\) is an optimal solution to \((P)\), \((\overline{v}_1, \ldots, \overline{v}_m)\) is an optimal solution to \((D)\) and \(v(P) = v(D)\),
(b) \(\lambda_n (p_{1,n} - x_n) \to 0 \quad (n \to +\infty), \quad \lambda_n (p_{2,i,n} - y_{i,n}) \to 0 \quad (n \to +\infty), \quad \text{and} \quad \lambda_n (p_{3,i,n} - v_{i,n}) \to 0 \quad (n \to +\infty) \) for \(i = 1, \ldots, m\),
(c) whenever \(\mathcal{H}\) and \(\mathcal{G}_i, \quad i = 1, \ldots, m\), are finite-dimensional Hilbert spaces, \((x_n)_{n \geq 0}\) converges strongly to an optimal solution to \((P)\) and \((v_{1,n}, \ldots, v_{m,n})_{n \geq 0}\) converges strongly to an optimal solution to \((D)\).

(ii) If
\[ \sum_{n=0}^{+\infty} \|a_n\|_\mathcal{H} < +\infty, \quad \sum_{n=0}^{+\infty} (\|d_{i,n}\|_{\mathcal{G}_i} + \|b_{i,n}\|_{\mathcal{G}_i}) < +\infty, \quad i = 1, \ldots, m, \quad \inf_{n \geq 0} \lambda_n > 0, \]
and \(f, \ l_i, \) and \(g_i^*\), \(i = 1, \ldots, m\), are uniformly convex, then \(v(P) = v(D)\), the sequence \((p_{1,n})_{n \geq 0}\) converges strongly to the unique optimal solution to \((P)\), and \((p_{3,1,n}, \ldots, p_{3,m,n})_{n \geq 0}\) converges strongly to the unique optimal solution of \((D)\).

Remark 3.33 According to Remark 3.26, when \(l_i : \mathcal{G}_i \to \overline{\mathbb{R}}\), \(l_i = \delta_{\{0\}}\), and \((d_{i,n})_{n \geq 0}\) is chosen as a sequence of zeros for every \(i = 1, \ldots, m\), the assertions made in Theorem 3.32 hold true for step length parameters satisfying
\[ \tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 1 \]
when taking in Algorithm 3.31 as initial choice \((y_1,0,\ldots,y_m,0) = (0,\ldots,0)\). In this case the sequences \((y_{1,n},\ldots,y_{m,n})_{n \geq 0}\) and \((p_{2,1,n},\ldots,p_{2,m,n})_{n \geq 0}\) vanish and (3.85) reduces to

\[
    (\forall n \geq 0) \quad \begin{cases}
        p_{1,n} = \text{Prox}_{\tau f} (x_n - \tau (\sum_{i=1}^{m} L_i^* v_{i,n} - z)) + a_n, \\
        x_{n+1} = x_n + \lambda_n (p_{1,n} - x_n), \\
        p_{3,i,n} = \text{Prox}_{\sigma_i g_i^*} (v_{i,n} + \sigma_i (L_i(2p_{1,n} - x_n) - r_i)) + b_{i,n}, \\
        v_{i,n+1} = v_{i,n} + \lambda_n (p_{3,i,n} - v_{i,n}).
    \end{cases}
\]

(3.87)

**Remark 3.34** Condition (3.84) in Theorem 3.30 (respectively, condition (3.86) in Theorem 3.32) is fulfilled, if the primal optimization problem (3.78) has an optimal solution,

\[
    0 \in \text{sqri} (\text{dom } g_i^* - \text{dom } l_i^*), \quad i = 1,\ldots,m,
\]

and (see, also, [58, Proposition 4.3])

\[
    (r_1,\ldots,r_m) \in \text{sqri } E,
\]

where

\[
    E := \left\{ (L_1 x - y_1,\ldots,L_m x - y_m) : x \in \text{dom } f \text{ and } y_i \in \text{dom } g_i + \text{dom } l_i, \quad i = 1,\ldots,m \right\}.
\]

According to [11, Proposition 15.7], condition (3.88) guarantees that \(g_i \square l_i \in \Gamma(G_i), \quad i = 1,\ldots,m\). If one of the following two conditions

(i) for any \(i = 1,\ldots,m\) one of the functions \(g_i\) and \(l_i\) is real-valued,

(ii) \(\mathcal{H}\) and \(\mathcal{G}_i, \quad i = 1,\ldots,m\), are finite dimensional and there exists \(x \in \text{ri dom } f\) such that \(L_i x - r_i \in \text{ri dom } g_i + \text{ri dom } l_i, \quad i = 1,\ldots,m\),

is fulfilled, then condition (3.89) is obviously true. For (ii) one has to take into account that in finite dimensional spaces the strong quasi-relative interior of a convex set is nothing else than its relative interior and to use the properties of the latter.

### 3.3 Solving inclusions with parallel sums of linearly composed monotone operators

The subject of this section, which relies on our article in [37], is to develop primal-dual methods of different types for solving monotone inclusion problems involving parallel sums of linearly composed monotone operators. These specifically constructed parallel sums are inspired and motivated by a real-world application in imaging sciences as pointed out in the sequel.

As mentioned in the introduction, the beginnings of modern primal-dual algorithms arose from the proximal point algorithm (see [92,109]) for finding a zero of a maximally monotone operator and from the Douglas–Rachford splitting algorithm (see [64]) which generalizes the problem to the one of finding a zero in the sum of two maximally monotone operators by separately invoking their resolvents.
Dependent on the properties of the maximally monotone operators in the classical sum problem, three fundamental algorithms emerge from the field: the Douglas–Rachford method (cf. [64]), the forward-backward method (cf. [88, 103]), and the forward-backward-forward method (cf. [121]). However, in the last couple of years, motivated by different applications, the complexity of the monotone inclusion problems significantly increased by allowing in their formulation compositions of maximally monotone with bounded linear operators (see [43, 48]), (single-valued) Lipschitzian or cocoercive monotone operators and parallel sums of maximally monotone operators (see [17, 25, 38, 55, 58, 60, 122]).

In this section, we generalize the monotone inclusion problems even further by allowing linearly composed monotone operators in the parallel sum constructions. Our aim, therefore, is to find easily implementable schemes by providing fully decomposable methods which avoid the (most likely expensive and instable) inversion of bounded linear operators.

### 3.3.1 Problem description

Our problem formulation is inspired by a real-world application in imaging (cf. [47, 117]), where first- and second-order total variation functionals are linked via infimal convolutions in order to reduce staircasing effects in the reconstructed images. One such particular imaging problem is given in Subsection 4.1.2. The problem under investigation follows.

**Problem 3.35** Let \( \mathcal{H} \) be a real Hilbert space, \( z \in \mathcal{H} \), let \( A : \mathcal{H} \to 2^{\mathcal{H}} \) be a maximally monotone operator, and \( C : \mathcal{H} \to \mathcal{H} \) be a monotone \( \mu^{-1} \)-cocoercive operator for \( \mu \in \mathbb{R}_{++} \). Furthermore, for every \( i = 1, \ldots, m \), let \( G_i, X_i, Y_i \) be real Hilbert spaces, \( r_i \in G_i, B_i : X_i \to 2^{X_i} \), and \( D_i : Y_i \to 2^{Y_i} \) be maximally monotone operators, and consider the nonzero bounded linear operators \( L_i : \mathcal{H} \to G_i, K_i : G_i \to X_i \), and \( M_i : G_i \to Y_i \). The problem is to solve the primal inclusion

\[
\text{find } x \in \mathcal{H} \text{ such that } z \in A x + \sum_{i=1}^{m} L_i^* \left( K_i^* \circ B_i \circ K_i \right) \Box \left( M_i^* \circ D_i \circ M_i \right) \left( L_i x - r_i \right) + C x
\]

(3.90)

together with its dual inclusion

\[
\text{find } \begin{cases} p_i \in X_i, & i = 1, \ldots, m, \\ \eta_i \in Y_i, & i = 1, \ldots, m, \\ \psi_i \in G_i, & i = 1, \ldots, m, \end{cases} \quad \text{such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^* K_i^* \psi_i \in A x + C x, \\ K_i (L_i x - \eta_i - r_i) \in B_i^{-1} \psi_i, & i = 1, \ldots, m, \\ M_i \eta_i \in D_i^{-1} \eta_i, & i = 1, \ldots, m, \\ K_i^* \eta_i = M_i^* \eta_i, & i = 1, \ldots, m. \end{cases}
\]

(3.91)

We provide in this section iterative methods of forward-backward and forward-backward-forward type for solving this primal-dual pair of monotone inclusion problems and investigate their asymptotic behavior. A similar problem formulation was recently investigated in [17], however, the proposed iterative scheme there relies on the forward-backward-forward method and is different to the corresponding one which we derive. Nevertheless, by making a forward step less, the forward-backward
method is more attractive from the perspective of its numerical implementation which is supported by our experimental results reported in Subsection 4.1.2.

Having the sequences \((x_n)_{n \geq 0}\) and \((y_n)_{n \geq 0}\) in \(\mathcal{H}\), we mind errors in the implementation of the algorithms by using the following notation taken from [17]

\[
(x_n \approx y_n \ \forall n \geq 0) \iff \sum_{n \geq 0} \|x_n - y_n\| < +\infty.
\]

(3.92)

Within this section, we provide the asymptotic convergence of two different algorithms for solving the primal-dual inclusion introduced in Problem 3.35. In Subsection 3.3.3, however, the assumptions imposed on the monotone operator \(C : \mathcal{H} \to \mathcal{H}\) will be weakened in the sense that \(C\) is only assumed to be \(\mu\)-Lipschitz continuous for some \(\mu \in \mathbb{R}_{++}\).

In the following, we let

\[
\begin{align*}
\mathcal{X} &= \mathcal{X}_1 \oplus \ldots \oplus \mathcal{X}_m, \\
\mathcal{Y} &= \mathcal{Y}_1 \oplus \ldots \oplus \mathcal{Y}_m, \\
p &= (p_1, \ldots, p_m) \in \mathcal{X}, \\
q &= (q_1, \ldots, q_m) \in \mathcal{Y}, \\
\mathcal{G} &= \mathcal{G}_1 \oplus \ldots \oplus \mathcal{G}_m,
\end{align*}
\]

and

\[
\begin{align*}
y &\in (y_1, \ldots, y_m) \in \mathcal{G}.
\end{align*}
\]

We say that \((\pi, \underline{p}, \underline{q}, \underline{y}) \in \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G}\) is a primal-dual solution to Problem 3.35, if

\[
z - \sum_{i=1}^m L_i^* K_i^* \pi_i \in A\pi + C\pi
\]

(3.93)

and

\[
K_i (L_i \pi - y_i - r_i) \in B_i^{-1} \pi_i, \quad M_i \overline{y}_i \in D_i^{-1} \overline{q}_i, \quad K_i^* \pi_i = M_i^* \overline{q}_i, \quad i = 1, \ldots, m.
\]

If \((\pi, \underline{p}, \underline{q}, \underline{y}) \in \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G}\) is a primal-dual solution to Problem 3.35, then \(\pi\) is a solution to (3.90) and \((\overline{p}, \overline{q}, \overline{y})\) is a solution to (3.91). Notice also that

\[
\pi\text{ solves (3.90)} \iff z \in A\pi + \sum_{i=1}^m L_i^* \left( (K_i^* \circ B_i \circ K_i) \boxdot \left( M_i^* \circ D_i \circ M_i \right) \right) (L_i \pi - r_i) + C\pi
\]

\[
\iff \exists \underline{\pi} \in \mathcal{G} \text{ such that } \\
z - \sum_{i=1}^m L_i^* \pi_i \in A\pi + C\pi,
\]

\[
L_i \pi - r_i \in (K_i^* \circ B_i \circ K_i)^{-1} (\underline{\pi}_i) + (M_i^* \circ D_i \circ M_i)^{-1} (\underline{\pi}_i),
\]

\(i = 1, \ldots, m,
\]

\[
\iff \exists (\underline{\pi}, \underline{\overline{y}}) \in \mathcal{G} \oplus \mathcal{G} \text{ such that } \\
z - \sum_{i=1}^m L_i^* \pi_i \in A\pi + C\pi,
\]

\[
\underline{\pi}_i \in \left( K_i^* \circ B_i \circ K_i \right) (L_i \pi - \overline{y}_i - r_i), \quad i = 1, \ldots, m,
\]

\[
\underline{\overline{y}}_i \in \left( D_i \circ M_i \right) (\overline{y}_i), \quad i = 1, \ldots, m,
\]

\[
K_i^* \pi_i = M_i^* \overline{y}_i, \quad i = 1, \ldots, m,
\]

\[
\iff \exists (\underline{p}, \underline{q}, \underline{y}) \in \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G} \text{ such that } \\
z - \sum_{i=1}^m L_i^* K_i^* \pi_i \in A\pi + C\pi,
\]

\[
\underline{p}_i \in \left( B_i \circ K_i \right) (L_i \pi - \overline{y}_i - r_i), \quad i = 1, \ldots, m,
\]

\[
\overline{q}_i \in \left( D_i \circ M_i \right) (\overline{y}_i), \quad i = 1, \ldots, m,
\]

\[
K_i^* \pi_i = M_i^* \overline{q}_i, \quad i = 1, \ldots, m,
\]

(3.94)

Thus, if \(\pi\) is a solution to (3.90), then there exists \((\underline{p}, \underline{q}, \underline{y}) \in \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G}\) such that \((\pi, \underline{p}, \underline{q}, \underline{y})\) is a primal-dual solution to Problem 3.35 and, if \((\underline{p}, \underline{q}, \underline{y})\) is a solution to (3.91), then there exists \(\pi \in \mathcal{H}\) such that \((\pi, \underline{p}, \underline{q}, \underline{y})\) is a primal-dual solution to Problem 3.35.
Remark 3.36 The notations (3.92) have been introduced in order to allow errors in the implementation of the algorithm, without affecting the readability in the sequel. This is reasonable since errors preserve their summability under addition, scalar multiplication and bounded linear transformations.

3.3.2 An algorithm of forward-backward type

The algorithm we propose for solving Problem 3.35 follows. We will prove its convergence by showing that the iterative scheme reduces to an error-tolerant forward-backward algorithm.

Algorithm 3.37 Let \( x_0 \in \mathcal{H} \), and for every \( i = 1, \ldots, m \), let \( p_{i,0} \in \mathcal{X}_i \), \( q_{i,0} \in \mathcal{Y}_i \), and \( z_{i,0}, y_{i,0}, v_{i,0} \in \mathcal{G}_i \). For every \( i = 1, \ldots, m \), let \( \tau, \theta_1, \theta_2, \gamma_1, \gamma_2 \), and \( \sigma \) be strictly positive real numbers such that

\[
2\mu^{-1} (1 - \tau) \min_{i=1,\ldots,m} \left\{ \frac{1}{\tau} \frac{1}{\theta_1 i} \frac{1}{\theta_2 i} \frac{1}{\gamma_1 i} \frac{1}{\gamma_2 i} \frac{1}{\sigma} \right\} > 1,
\]

for

\[
\alpha = \max \left\{ \sum_{i=1}^{m} \sigma_i ||L_i||^2, \max_{j=1,\ldots,m} \left\{ \sqrt{\theta_1 j \gamma_1 j ||K_j||^2}, \sqrt{\theta_2 j \gamma_2 j ||M_j||^2} \right\} \right\}.
\]

Furthermore, let \( \varepsilon \in (0, 1) \), let \( (\lambda_n)_{n \geq 0} \) be a sequence in \([\varepsilon, 1]\), and set

\[
(\forall n \geq 0) \quad \begin{cases}
\bar{x}_n \approx J_{\tau A} (x_n - \tau (C x_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z)), \\
\text{For } i = 1, \ldots, m \\
\quad p_{i,n} \approx J_{\theta_1 i, B_i^{-1}} (p_{i,n} + \theta_1 i K_i z_{i,n}), \\
\quad q_{i,n} \approx J_{\theta_2 i, D_i^{-1}} (q_{i,n} + \theta_2 i M_i y_{i,n}), \\
\quad u_{1,i,n} \approx z_{1,i,n} + \gamma_1 i (K_i^* (p_{i,n} - 2\tilde{p}_{i,n}) + v_{i,n} + \sigma_i (L_i (2\tilde{x}_n - x_n) - r_i)), \\
\quad u_{2,i,n} \approx y_{i,n} + \gamma_2 i (M_i^* (q_{i,n} - 2\tilde{q}_{i,n}) + u_{1,i,n} + \sigma_i (L_i (2\tilde{x}_n - x_n) - r_i)), \\
\quad \tilde{z}_{i,n} \approx \frac{u_{1,i,n} - \sigma_i \gamma_1 i u_{2,i,n}}{1 + \sigma_i \gamma_1 i}, \\
\quad \tilde{y}_{i,n} \approx \frac{u_{1,i,n} - \sigma_i \gamma_2 i u_{2,i,n}}{1 + \sigma_i \gamma_2 i}, \\
\quad \tilde{v}_{i,n} \approx v_{i,n} + \sigma_i (L_i (2\tilde{x}_n - x_n) - r_i - \tilde{z}_{i,n} - \tilde{y}_{i,n}), \\
\quad \tilde{x}_{n+1} = x_n + \lambda_n (\tilde{x}_n - x_n), \\
\text{For } i = 1, \ldots, m \\
\quad p_{i,n+1} = p_{i,n} + \lambda_n (\tilde{p}_{i,n} - p_{i,n}), \\
\quad q_{i,n+1} = q_{i,n} + \lambda_n (\tilde{q}_{i,n} - q_{i,n}), \\
\quad z_{i,n+1} = z_{i,n} + \lambda_n (\tilde{z}_{i,n} - z_{i,n}), \\
\quad y_{i,n+1} = y_{i,n} + \lambda_n (\tilde{y}_{i,n} - y_{i,n}), \\
\quad v_{i,n+1} = v_{i,n} + \lambda_n (\tilde{v}_{i,n} - v_{i,n}).
\end{cases}
\]

Theorem 3.38 For Problem 3.35, suppose that

\[
z \in \text{ran} \left( A + \sum_{i=1}^{m} L_i^* \left( (K_i^* \circ B_i \circ K_i) \square (M_i^* \circ D_i \circ M_i) \right) (L_i \cdot -r_i) + C \right),
\]

and consider the sequences generated by Algorithm 3.37. Then there exists a primal-dual solution \((\bar{x}, p, q, y)\) to Problem 3.35 such that
We introduce the maximally monotone operators

\[ (\text{Propositions 20.22 and 20.23}), \]

and the bounded linear operator

\[ (3.97) \]

We introduce the real Hilbert space \( \mathcal{K} = \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G} \oplus \mathcal{G} \oplus \mathcal{G} \) and let

\[
\begin{align*}
 p &= (p_1, \ldots, p_m), \\
 q &= (q_1, \ldots, q_m), \quad \text{and} \\
 y &= (y_1, \ldots, y_m), \\
 z &= (z_1, \ldots, z_m), \\
 v &= (v_1, \ldots, v_m), \\
 r &= (r_1, \ldots, r_m).
\end{align*}
\]

We introduce the maximally monotone operators

\[ B : \mathcal{X} \to 2^\mathcal{X}, \quad p \mapsto B_1 p_1 \times \ldots \times B_m p_m \quad \text{and} \quad D : \mathcal{Y} \to 2^\mathcal{Y}, \quad q \mapsto D_1 q_1 \times \ldots \times D_m q_m. \]

Further, consider the set-valued operator

\[ (3.98) \]

which is maximally monotone, since \( A, B, \) and \( D \) are maximally monotone (cf. \[11, \) Propositions 20.22 and 20.23\]), and the bounded linear operator

\[ (x, p, q, y, z, v) \mapsto (0, 0, 0, -v, -v, z + y) \]

is skew and hence maximally monotone (cf. \[11, \) Example 20.30\]). Therefore, \( M \) can be written as the sum of two maximally monotone operators, one of them having full domain, a fact which leads to the maximality of \( M \) (see, for instance, \[11, \) Corollary 24.4(i)]). Furthermore, consider the bounded linear operators

\[ (3.99) \]

and

\[ (x, p, q, z, y, v) \mapsto \left( \sum_{i=1}^m L_i^* v_i, -K z, -M y, K^* p, M^* q, -L_1 x, \ldots, -L_m x \right). \]

The operator \( S \) is skew, as well, hence maximally monotone. As \( \text{dom} S = \mathcal{K} \), the sum \( M + S \) is maximally monotone (see \[11, \) Corollary 24.4(i)]).

Finally, we introduce the monotone operator

\[ (x, p, q, z, y, v) \mapsto (C x, 0, 0, 0, 0, 0) \]

which is, obviously, \( \mu^{-1} \)-cocoercive. By making use of \((3.94)\), we observe that

\[ (3.97) \iff \exists (x, p, q, y) \in \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G} : \]

\[ (3.98) \iff \exists (x, p, q) \in \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} : \]

\[ (3.99) \iff \exists (z, y, v) \in \mathcal{G} \oplus \mathcal{G} \oplus \mathcal{G} : \]

\[ (3.100) \iff \exists (z, y, v) \in \mathcal{G} \oplus \mathcal{G} \oplus \mathcal{G} : \]
\[ \Leftrightarrow \exists (x, p, q, z, y, v) \in \text{zer}(M + S + Q). \]

From here, it follows that

\[
(x, p, q, z, y, v) \in \text{zer}(M + S + Q) \quad \Rightarrow \quad \begin{cases}
z - \sum_{i=1}^{m} L_i^* K_i^p_i \in A x + C x, \\
K_i (L_i x - \bar{y}_i - r_i) \in B_i^{-1} q_i, \quad i = 1, \ldots, m, \\
M_i \bar{y}_i \in D_i^{-1} q_i, \quad i = 1, \ldots, m, \\
K_i^* p_i = M_i^* q_i, \quad i = 1, \ldots, m,
\end{cases}
\]

\[ \Leftrightarrow (x, p, q, y) \text{ is a primal-dual solution to Problem 3.35.} \quad (3.99) \]

Further, for positive real values \( \tau, \theta_{1,i}, \theta_{2,i}, \gamma_{1,i}, \gamma_{2,i}, \sigma_i \in \mathbb{R}_{++}, \ i = 1, \ldots, m, \) we introduce the notations

\[
\begin{align*}
 p_{\theta_1} &= \left( \frac{p_1}{\theta_{1,1}}, \ldots, \frac{p_m}{\theta_{1,m}} \right), \\
 q_{\theta_2} &= \left( \frac{q_1}{\theta_{2,1}}, \ldots, \frac{q_m}{\theta_{2,m}} \right), \\
 z_{\gamma_1} &= \left( \frac{z_1}{\gamma_{1,1}}, \ldots, \frac{z_m}{\gamma_{1,m}} \right), \\
 y_{\gamma_2} &= \left( \frac{y_1}{\gamma_{2,1}}, \ldots, \frac{y_m}{\gamma_{2,m}} \right), \\
 v_{\sigma} &= \left( \frac{v_1}{\sigma_1}, \ldots, \frac{v_m}{\sigma_m} \right),
\end{align*}
\]

and define the bounded linear operator

\[
V : \mathcal{K} \to \mathcal{K}, \quad (x, p, q, z, y, v) \mapsto \left( \frac{x}{\tau}, \frac{p}{\theta_1}, \frac{q}{\theta_2}, \frac{z}{\gamma_1}, \frac{y}{\gamma_2}, \frac{v}{\sigma} \right)
\]

\[ + \left( -\sum_{i=1}^{m} L_i^* v_i, \tilde{K} z, \tilde{M} y, \tilde{K}^* p, \tilde{M}^* q_i, -L_1 x, \ldots, -L_m x \right). \]

It is a simple calculation to prove that \( V \) is self-adjoint. Furthermore, the operator \( V \) is \( \rho \)-strongly positive with

\[
\rho = (1 - \overline{\alpha}) \min_{i=1,\ldots,m} \left\{ \frac{1}{\tau}, \frac{1}{\theta_{1,i}}, \frac{1}{\theta_{2,i}}, \frac{1}{\gamma_{1,i}}, \frac{1}{\gamma_{2,i}}, \frac{1}{\sigma_i} \right\} > 0,
\]

for

\[
\overline{\alpha} = \max \left\{ \sqrt{\frac{m}{\tau}} \sum_{i=1}^{m} \sigma_i ||L_i||^2, \max_{j=1,\ldots,m} \left\{ \sqrt{\theta_{1,j} \gamma_{1,j} ||K_j||^2}, \sqrt{\theta_{2,j} \gamma_{2,j} ||M_j||^2} \right\} \right\}.
\]

The fact that \( \rho \) is a positive real number then follows by the assumptions made in Algorithm 3.37. Indeed, using that \( 2ab \leq a \alpha^2 + \frac{b^2}{\alpha} \) for every \( a, b \in \mathbb{R} \) and every \( \alpha \in \mathbb{R}_{++} \), it yields for any \( i = 1, \ldots, m, \)

\[
2 ||L_i|| ||x|| ||v_i|| \leq \frac{\sigma_i ||L_i||^2}{\sqrt{\tau} \sum_{j=1}^{m} \sigma_j ||L_j||^2} ||x||^2_{\bar{b}_i} + \sqrt{\frac{\tau}{\sum_{j=1}^{m} \sigma_j ||L_j||^2}} ||v_i||^2_{\bar{g}_i},
\]

\[
2 ||K_i|| ||p_i|| ||z_i|| \leq \frac{\gamma_{1,i} ||K_i|| ||p_i||^2}{\sqrt{\theta_{1,i} \gamma_{1,i}}} ||x_i||^2_{\bar{b}_i} + \sqrt{\theta_{1,i} \gamma_{1,i} ||K_i||^2} ||z_i||^2_{\bar{g}_i}, \quad (3.100)
\]

\[
2 ||M_i|| ||q_i|| ||y_i|| \leq \frac{\gamma_{2,i} ||M_i|| ||q_i||^2}{\sqrt{\theta_{2,i} \gamma_{2,i}}} ||y_i||^2_{\bar{g}_i} + \sqrt{\theta_{2,i} \gamma_{2,i} ||M_i||^2} ||y_i||^2_{\bar{g}_i}.
\]
Consequently, for each \( x = (x, p, q, z, y, v) \in K \), using the Cauchy–Schwarz inequality and \((3.100)\), it follows that
\[
\langle x, Vx \rangle_K = \frac{\|x\|_K^2}{\tau} + \sum_{i=1}^{m} \left[ \frac{\|p_i\|_{\beta_i}}{\theta_1,i} + \frac{\|q_i\|_{\gamma_i}}{\theta_2,i} + \frac{\|z_i\|_{\gamma_1}}{\gamma_1,i} + \frac{\|y_i\|_{\gamma_2}}{\gamma_2,i} + \frac{\|v_i\|_{\sigma_i}}{\sigma_i} \right] \\
- 2 \sum_{i=1}^{m} \langle L_i x, v_i \rangle_{\theta_i} + 2 \sum_{i=1}^{m} \langle p_i, K_i z_i \rangle_{\chi_i} + 2 \sum_{i=1}^{m} \langle q_i, M_i y_i \rangle_{\gamma_i} \\
\geq (1 - \alpha) \min_{i=1,...,m} \left\{ \frac{1}{\tau}, \frac{1}{\theta_1,i}, \frac{1}{\theta_2,i}, \frac{1}{\gamma_1,i}, \frac{1}{\gamma_2,i}, \frac{1}{\sigma_i} \right\} \|x\|_K^2 \\
= \rho \|x\|_K^2. \tag{3.101}
\]
Since \( V \) is maximally monotone (cf. [11, Example 20.29]) and \( \rho \)-strongly positive, it is strongly monotone and therefore, by [11, Proposition 22.8], it holds that \( V \) is surjective. Consequently, \( V^{-1} \) exists and \( \|V^{-1}\| \leq \frac{1}{\rho} \).

In consideration of \((3.92)\), the algorithmic scheme \((3.96)\) can equivalently be written in the form
\[
\begin{aligned}
\forall n \geq 0 \\
\frac{x_{n+1} - \tilde{x}_n}{\tau} - \sum_{i=1}^{m} L_i^* (v_{i,n} - \tilde{v}_{i,n}) - C x_n \\
= -z + A(x_n - a_n) + \sum_{i=1}^{m} L_i^* \tilde{v}_{i,n} - \frac{a_n}{\tau},
\end{aligned}
\]
For \( i = 1, \ldots, m \)
\[
\begin{aligned}
\frac{p_{i,n} - \tilde{p}_{i,n}}{\theta_1,i} + K_i (\tilde{z}_{i,n} - \tilde{z}_{i,n}) &\in B_i^{-1}(\tilde{p}_{i,n} - b_{i,n}) - K_i \tilde{z}_{i,n} - \frac{b_{i,n}}{\theta_1,i}, \\
\frac{q_{i,n} - \tilde{q}_{i,n}}{\gamma_1,i} + M_i (\tilde{y}_{i,n} - \tilde{y}_{i,n}) &\in D_i^{-1}(\tilde{q}_{i,n} - d_{i,n}) - M_i \tilde{y}_{i,n} - \frac{d_{i,n}}{\gamma_1,i}, \\
\frac{z_{i,n} - \tilde{z}_{i,n}}{\gamma_2,i} + K_i^* (p_{i,n} - \tilde{p}_{i,n}) &\in -\tilde{v}_{i,n} + K_i^* \tilde{p}_{i,n} - e_{1,i,n}, \\
\frac{y_{i,n} - \tilde{y}_{i,n}}{\gamma_2,i} + M_i^* (q_{i,n} - \tilde{q}_{i,n}) &\in -\tilde{y}_{i,n} + M_i^* \tilde{q}_{i,n} - e_{2,i,n}, \\
\frac{v_{i,n} - \tilde{v}_{i,n}}{\sigma_i} - L_i(x_n - \tilde{x}_n) &\in r_i + \tilde{z}_{i,n} + \tilde{y}_{i,n} - L_i \tilde{x}_n - e_{3,i,n},
\end{aligned}
\]
\[
x_{n+1} = x_n + \lambda_n(x_n - x_n).
\tag{3.102}
\]

We further introduce the notations
\[
\begin{align*}
p_n &= (p_{1,n}, \ldots, p_{m,n}) \in \mathcal{X}, & \tilde{p}_n &= (\tilde{p}_{1,n}, \ldots, \tilde{p}_{m,n}) \in \mathcal{X}, \\
q_n &= (q_{1,n}, \ldots, q_{m,n}) \in \mathcal{Y}, & \tilde{q}_n &= (\tilde{q}_{1,n}, \ldots, \tilde{q}_{m,n}) \in \mathcal{Y}, \\
z_n &= (z_{1,n}, \ldots, z_{m,n}) \in \mathcal{G}, & \tilde{z}_n &= (\tilde{z}_{1,n}, \ldots, \tilde{z}_{m,n}) \in \mathcal{G}, \\
y_n &= (y_{1,n}, \ldots, y_{m,n}) \in \mathcal{G}, & \tilde{y}_n &= (\tilde{y}_{1,n}, \ldots, \tilde{y}_{m,n}) \in \mathcal{G}, \\
v_n &= (v_{1,n}, \ldots, v_{m,n}) \in \mathcal{G}, & \tilde{v}_n &= (\tilde{v}_{1,n}, \ldots, \tilde{v}_{m,n}) \in \mathcal{G},
\end{align*}
\]
which are combined to
\[
\begin{align*}
x_n &= (x_n, p_n, q_n, z_n, y_n, v_n) \in K, & \tilde{x}_n &= (\tilde{x}_n, \tilde{p}_n, \tilde{q}_n, \tilde{z}_n, \tilde{y}_n, \tilde{v}_n) \in K.
\end{align*}
\]

Also, for any \( n \geq 0 \), we consider sequences defined by
\[
\begin{align*}
a_n &\in \mathcal{H}, & e_{1,n} &= (e_{1,1,n}, \ldots, e_{1,m,n}) \in \mathcal{G}, \\
b_n &= (b_{1,n}, \ldots, b_{m,n}) \in \mathcal{X}, & e_{2,n} &= (e_{2,1,n}, \ldots, e_{2,m,n}) \in \mathcal{G}, \\
d_n &= (d_{1,n}, \ldots, d_{m,n}) \in \mathcal{Y}, & e_{3,n} &= (e_{3,1,n}, \ldots, e_{3,m,n}) \in \mathcal{G},
\end{align*}
\]
that are summable in the corresponding norm. Further, by denoting for any \( n \geq 0 \)
\[
\begin{align*}
e_n &= (a_n, b_n, d_n, 0, 0, 0) \in K, & e_n &= (\frac{a_n}{\tau}, \frac{b_n}{\theta_1}, \frac{d_n}{\gamma_1}, e_{1,n}, e_{2,n}, e_{3,n}) \in K,
\end{align*}
\]
which are also terms of summable sequences in the corresponding norm, it yields that the scheme in (3.102) is equivalent to

$$\forall n \geq 0 \quad \begin{cases} V(x_n - \tilde{x}_n) - Qx_n \in (M + S)(\tilde{x}_n - e_n) + Se_n - e_n^\tau, \\ x_{n+1} = x_n + \lambda_n (\tilde{x}_n - x_n). \end{cases} \quad (3.104)$$

We now introduce the notations

$$A_K := V^{-1}(M + S) \quad \text{and} \quad B_K := V^{-1}Q, \quad (3.105)$$

and the summable sequence with terms $e_n^V = V^{-1}((V + S)e_n - e_n^\tau)$ for any $n \geq 0$. Then, for any $n \geq 0$, we have

$$V(x_n - \tilde{x}_n) - Qx_n \in (M + S)(\tilde{x}_n - e_n) + Se_n - e_n^\tau$$

$$\iff Vx_n - Qx_n \in (V + M + S)(\tilde{x}_n - e_n) + (V + S)e_n - e_n^\tau$$

$$\iff x_n - V^{-1}Qx_n \in \left(\text{Id} + V^{-1}(M + S)\right)(\tilde{x}_n - e_n) + V^{-1}((V + S)e_n - e_n^\tau)$$

$$\iff \tilde{x}_n = \left(\text{Id} + V^{-1}(M + S)\right)^{-1}\left(x_n - V^{-1}Qx_n - e_n^V\right) + e_n$$

$$\iff \tilde{x}_n = (\text{Id} + A_K)^{-1}\left(x_n - B_Kx_n - e_n^V\right) + e_n. \quad (3.106)$$

By taking into account that the resolvent is 1-Lipschitz continuous, the sequence having as terms

$$e_n^{A_K} = J_{A_K}\left(x_n - B_Kx_n - e_n^V\right) - J_{A_K}\left(x_n - B_Kx_n\right) + e_n \ \forall n \geq 0,$$

is summable and we have

$$\tilde{x}_n = J_{A_K}\left(x_n - B_Kx_n\right) + e_n^{A_K} \ \forall n \geq 0.$$

Thus, the iterative scheme in (3.104) becomes

$$\forall n \geq 0 \quad \begin{cases} \tilde{x}_n \approx J_{A_K}\left(x_n - B_Kx_n\right), \\ x_{n+1} = x_n + \lambda_n (\tilde{x}_n - x_n). \end{cases} \quad (3.107)$$

which shows that the algorithm which we propose in this subsection has the structure of a forward-backward method.

In addition, let us observe that

$$\text{zer}(A_K + B_K) = \text{zer}(V^{-1}(M + S + Q)) = \text{zer}(M + S + Q).$$

We then introduce the Hilbert space $\mathcal{K}_V$ with inner product and norm respectively defined, for $x, y \in \mathcal{K}$, via

$$\langle x, y \rangle_{\mathcal{K}_V} = \langle x, V y \rangle_{\mathcal{K}} \quad \text{and} \quad \|x\|_{\mathcal{K}_V} = \sqrt{\langle x, V x \rangle_{\mathcal{K}}} \quad (3.108)$$

Since $M + S$ and $Q$ are maximally monotone on $\mathcal{K}$, the operators $A_K$ and $B_K$ are maximally monotone on $\mathcal{K}_V$. Moreover, since $V$ is self-adjoint and $\rho$-strongly positive, it once again holds that weak and strong convergence in $\mathcal{K}_V$ are equivalent with weak and strong convergence in $\mathcal{K}$, respectively. By making use of $\|V^{-1}\| \leq \frac{1}{\rho},$
one can show that $B_K$ is $(\mu^{-1}\rho)$-cocoercive on $K_V$. Indeed, we get for $x, y \in K_V$ that (see, also, [122, Eq. (3.35)])

$$
\langle x - y, B_K x - B_K y \rangle_{K_V} = \langle x - y, Q x - Q y \rangle_{K_V} \\
\geq \mu^{-1} \|Q x - Q y\|^2_{k_V} \\
\geq \frac{\mu^{-1}}{|V^{-1}|} \|V^{-1} Q x - V^{-1} Q y\|_{K_V} \|Q x - Q y\|_{K_V} \\
\geq \frac{\mu^{-1}}{|V^{-1}|} \|B_K x - B_K y\|^2_{K_V} \\
\geq \frac{\mu^{-1}}{|V^{-1}|} \|B_K x - B_K y\|^2_{K_V}.
$$

As our assumption imposes that $2\mu^{-1}\rho > 1$, we can use the statements given in [53, Corollary 6.5] in the context of an error tolerant forward-backward algorithm in order to establish the desired convergence results.

(i) By Corollary 6.5 in [53], the sequence $(x_n)_{n \geq 0}$ converges weakly in $K_V$ (and therefore in $K$) to some $\overline{x} = (\overline{\tau}, \overline{\rho}, \overline{q}, \overline{z}, \overline{y}, \overline{v}) \in \text{zer} (A_K + B_K) = \text{zer} (M + S + Q)$. By (3.99), it thus follows that $(\overline{\tau}, \overline{\rho}, \overline{q}, \overline{y})$ is a primal-dual solution with respect to Problem 3.35.

(ii) From [53, Remark 3.4], it follows

$$
\sum_{n \geq 0} \|B_K x_n - B_K \overline{x}\|^2_{K_V} < +\infty,
$$

and therefore we have $B_K x_n \to B_K \overline{x}$, or, equivalently, $Q x_n \to Q \overline{x}$ as $n \to +\infty$.

Considering the definition of $Q$, one can see that this implies $C x_n \to C \overline{x}$ as $n \to +\infty$.

As $C$ is uniformly monotone, there exists an increasing function $\phi_C : [0, +\infty) \to [0, +\infty]$ vanishing only at 0 such that

$$
\phi_C(\|x_n - \overline{x}\|) \leq \langle x_n - \overline{x}, C x_n - C \overline{x}\rangle \leq \|x_n - \overline{x}\|\|C x_n - C \overline{x}\| \quad \forall n \geq 0.
$$

The boundedness of $(x_n - \overline{x})_{n \geq 0}$ and the convergence $C x_n \to C \overline{x}$ further imply that $x_n \to \overline{x}$ as $n \to +\infty$.

**Remark 3.39** Suppose that $C : \mathcal{H} \to \mathcal{H}$, $x \mapsto \{0\}$ in Problem 3.35. Then condition (3.95) simplifies to

$$
\max \left\{ \sum_{i=1}^{m} \sigma_i \|L_i\|^2, \max_{1 \leq j \leq m} \left\{ \theta_{1,j} \gamma_{1,j} \|K_j\|^2, \theta_{2,j} \gamma_{2,j} \|M_j\|^2 \right\} \right\} < 1.
$$

In this situation, the scheme (3.107) reads

$$
(\forall n \geq 0) \quad x_{n+1} \approx x_n + \lambda_n (J_{A_K} x_n - x_n),
$$

and it can be shown to converge under the relaxed assumption that $(\lambda_n)_{n \geq 0} \subseteq [\varepsilon, 2 - \varepsilon]$, for $\varepsilon \in (0, 1)$ (see, for instance, [52, 53, 65]).

**Remark 3.40** (i) When implementing Algorithm 3.37, the term $L_i(2\overline{x}_n - x_n)$ should be stored in a separate variable for all $i = 1, \ldots, m$. Taking this into account, each bounded linear operator occurring in Problem 3.35 needs to be processed once via some forward evaluation and once via its adjoint.
(ii) The maximally monotone operators $A, B_i, \text{ and } D_i, \ i = 1, \ldots, m,$ in Problem 3.35 are accessed separately via their resolvents, hence we obtained a fully decomposable algorithm.

(iii) The possibility of performing a forward step for the cocoercive monotone operator $C$ is an important aspect, since forward steps are usually much easier to implement than resolvents (resp. proximity operators). Due to the Baillon–Haddad theorem (cf. [8, 10]), each $\mu$-Lipschitzian gradient with $\mu \in \mathbb{R}_{++}$ of a convex and Fréchet differentiable function $f : \mathcal{H} \to \mathbb{R}$ is $\mu^{-1}$-cocoercive.

### 3.3.3 An algorithm of forward-backward-forward type

Let us assume the setting as proposed in Problem 3.35 with the modification that the operator $C : \mathcal{H} \to \mathcal{H}$ is $\mu$-Lipschitz continuous for some $\mu \in \mathbb{R}_{++}$, but not $\mu^{-1}$-cocoercive.

**Algorithm 3.41** Let $x_0 \in \mathcal{H}$, and for every $i = 1, \ldots, m$, let $p_{i,0} \in \mathcal{X}_i$, $q_{i,0} \in \mathcal{Y}_i$, and $z_{i,0}, \ y_{i,0}, \ v_{i,0} \in \mathcal{G}_i$. Set

$$
\beta = \mu + \sqrt{\max \left\{ \sum_{i=1}^{m} \|L_i\|^2, \max_{1 \leq i \leq m} \left\{ \|K_i\|^2, \|M_i\|^2 \right\} \right\}, \quad (3.111)
$$

let $\varepsilon \in (0, 1/(\beta + 1))$, let $(\gamma_n)_{n \geq 0}$ be a sequence in $[\epsilon, (1-\varepsilon)/\beta]$, and set

$$
(\forall n \geq 0)
$$

\[
\begin{align*}
\tilde{x}_n &\approx J_{\gamma_n A} (x_n - \gamma_n (Cx_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z)) , \\
 &\quad \text{For } i = 1, \ldots, m \\
\tilde{p}_{i,n} &\approx J_{\gamma_n B_i^{-1}} (p_{i,n} + \gamma_n K_i z_{i,n}) , \\
q_{i,n} &\approx J_{\gamma_n D_i^{-1}} (q_{i,n} + \gamma_n M_i y_{i,n}) , \\
u_{1,i,n} &\approx z_{i,n} - \gamma_n (K_i^* p_{i,n} - v_{i,n} - \gamma_n (L_i x_n - r_i)) , \\
u_{2,i,n} &\approx y_{i,n} - \gamma_n (M_i^* q_{i,n} - v_{i,n} - \gamma_n (L_i x_n - r_i)) , \\
\hat{z}_{i,n} &\approx \frac{1}{1+2\alpha^2} (u_{1,i,n} - \frac{\gamma_n^2}{1+2\alpha^2} u_{2,i,n}) , \\
\hat{y}_{i,n} &\approx \frac{1}{1+2\alpha^2} (u_{2,i,n} - \frac{\gamma_n^2}{1+2\alpha^2} u_{2,i,n}) , \\
\hat{v}_{i,n} &\approx v_{i,n} + \gamma_n (L_i x_n - r_i - \hat{z}_{i,n} - \hat{y}_{i,n}) , \\
x_{n+1} &\approx \tilde{x}_n + \gamma_n (Cx_n - C\tilde{x}_n + \sum_{i=1}^{m} L_i^*(v_{i,n} - \hat{v}_{i,n})) , \\
&\quad \text{For } i = 1, \ldots, m \\
p_{i,n+1} &\approx \tilde{p}_{i,n} - \gamma_n (K_i (z_{i,n} - \hat{z}_{i,n})) , \\
q_{i,n+1} &\approx \hat{q}_{i,n} - \gamma_n (M_i (y_{i,n} - \hat{y}_{i,n})) , \\
z_{i,n+1} &\approx \hat{z}_{i,n} + \gamma_n (K_i^* (p_{i,n} - \hat{p}_{i,n})) , \\
y_{i,n+1} &\approx \hat{y}_{i,n} + \gamma_n (M_i^* (q_{i,n} - \hat{q}_{i,n})) , \\
v_{i,n+1} &\approx \hat{v}_{i,n} - \gamma_n (L_i (x_n - \tilde{x}_n)). \\
\end{align*}
\]

**Theorem 3.42** In Problem 3.35, let $C : \mathcal{H} \to \mathcal{H}$ only be $\mu$-Lipschitz continuous for $\mu \in \mathbb{R}_{++}$, suppose that

$$
z \in \text{ran} \left( A + \sum_{i=1}^{m} L_i^* \left( (K_i^* \circ B_i \circ K_i) \square (M_i^* \circ D_i \circ M_i) \right) (L_i \cdot -r_i) + C \right), \quad (3.113)
$$

and consider the sequences generated by Algorithm 3.41. Then there exists a primal-dual solution $(x, p, q, y)$ to Problem 3.35 such that
Therefore the sum \( \sum_{n \geq 0} \|x_n - \tilde{x}_n\|^2 < +\infty \), and for \( i = 1, \ldots, m \),

\[
\sum_{n \geq 0} \|p_{i,n} - \tilde{p}_{i,n}\|^2 < +\infty, \quad \sum_{n \geq 0} \|q_{i,n} - \tilde{q}_{i,n}\|^2 < +\infty, \quad \text{and} \quad \sum_{n \geq 0} \|y_{i,n} - \tilde{y}_{i,n}\|^2 < +\infty,
\]

(iii) \( x_n \to \overline{x}, \; \tilde{x}_n \to \overline{x} \), and for \( i = 1, \ldots, m \),

\[
\begin{aligned}
& \left\{ \begin{array}{l}
p_{i,n} \rightharpoonup \tilde{p}_{i,n}, \\
\tilde{p}_{i,n} \rightharpoonup \overline{p}_{i,n}, \\
q_{i,n} \rightharpoonup \tilde{q}_{i,n}, \\
\tilde{q}_{i,n} \rightharpoonup \overline{q}_{i,n}, \\
y_{i,n} \rightharpoonup \tilde{y}_{i,n}, \\
\tilde{y}_{i,n} \rightharpoonup \overline{y}_{i,n}.
\end{array} \right.
\end{aligned}
\]

Proof. Consider the notations introduced in (3.98) including the Hilbert direct sum

\[
\mathcal{K} = \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G} \oplus \mathcal{G} \oplus \mathcal{G}.
\]

Then, we let \( M : \mathcal{K} \to 2^\mathcal{K} \), \( S : \mathcal{K} \to \mathcal{K} \) and \( Q : \mathcal{K} \to \mathcal{K} \) be defined as in the proof of Theorem 3.38. The operator \( S + Q \) is monotone, Lipschitz continuous, hence maximally monotone (cf. [11, Corollary 20.25]), and fulfills \( \text{dom}(S + Q) = \mathcal{K} \). Therefore the sum \( M + S + Q \) is maximally monotone as well (see [11, Corollary 24.4(i)]). In the following we derive the Lipschitz constant of \( S + Q \). Let

\[
x = (x, p, q, z, y, v) \quad \text{and} \quad \tilde{x} = (\tilde{x}, \tilde{p}, \tilde{q}, \tilde{z}, \tilde{y}, \tilde{v}) \in \mathcal{K}
\]

be arbitrary, then, using the Cauchy–Schwarz inequality,

\[
\| (S + Q)x - (S + Q)\tilde{x} \| \leq \| Qx - Q\tilde{x} \| + \| Sx - S\tilde{x} \|
\]

\[
\leq \mu \| x - \tilde{x} \| + \left( \sum_{i=1}^{m} L_i^*(v_i - \tilde{v}_i), -\tilde{K}(z - \tilde{z}), -\tilde{M}(y - \tilde{y}), \tilde{K}^*(p - \tilde{p}), \tilde{M}^*(q - \tilde{q}), -L_1(x - \tilde{x}), \ldots, -L_m(x - \tilde{x}) \right)
\]

\[
= \mu \| x - \tilde{x} \| + \left( \sum_{i=1}^{m} L_i^*(v_i - \tilde{v}_i) \right)^2 + \sum_{i=1}^{m} \left[ \| K_i(y_i - \tilde{y}_i) \|^2 + \| M_i(y_i - \tilde{y}_i) \|^2 + \| K_i^*(p_i - \tilde{p}_i) \|^2 + \| M_i^*(q_i - \tilde{q}_i) \|^2 + \| L_i(x - \tilde{x}) \|^2 \right]^{1/2}
\]

\[
\leq \mu \| x - \tilde{x} \| + \left( \sum_{i=1}^{m} \| L_i \|^2 \right) \left( \| x - \tilde{x} \|^2 + \sum_{i=1}^{m} \| v_i - \tilde{v}_i \|^2 \right)^1/2 + \sum_{i=1}^{m} \left[ \| K_i \|^2 \| z_i - \tilde{z}_i \|^2 + \| M_i \|^2 \| q_i - \tilde{q}_i \|^2 \right]^{1/2}
\]

\[
\leq \left( \mu + \max_{1 \leq i \leq m} \left\{ \sum_{i=1}^{m} \| L_i \|^2, \| K_i \|^2, \| M_i \|^2 \right\} \right) \| x - \tilde{x} \|. \quad (3.114)
\]

In the following, we use the sequences in (3.103) for modeling summable errors in the implementation. In addition we let

\[
e_n = (a_n, b_n, d_n, 0, 0, 0) \quad \text{and} \quad \tilde{e}_n = (0, 0, e_{1,n}, e_{2,n}, e_{3,n})
\]
be summable sequences in $\mathcal{K}$. Note that (3.112) can equivalently be written as
\[
\forall n \geq 0 \quad \begin{cases}
\left[ x_n - \gamma_n (Cx_n + \sum_{i=1}^{m} L_i^* v_{i,n}) \right] \in \left( \text{Id} + \gamma_n (-z + A) \right) (\bar{x}_n - a_n), \\
\text{For } i = 1, \ldots, m \\
\left[ p_{i,n} + \gamma_n K_i z_{i,n} \right] \in (\text{Id} + \gamma_n B_i^{-1}) (\bar{p}_{i,n} - b_{i,n}), \\
\left[ q_{i,n} + \gamma_n M_i q_{i,n} \right] \in (\text{Id} + \gamma_n D_i^{-1}) (\bar{q}_{i,n} - d_{i,n}), \\
\left[ z_{i,n} - \gamma_n K_i^* p_{i,n} = \bar{z}_{i,n} - \gamma_n \bar{v}_{i,n} - e_{i,i,n} \\
\left[ y_{i,n} - \gamma_n M_i^* q_{i,n} = \bar{y}_{i,n} - \gamma_n \bar{v}_{i,n} - e_{2,i,n}, \\
\left[ v_{i,n} + \gamma_n L_i x_n = \bar{v}_{i,n} + \gamma_n (r_i + \bar{z}_{i,n} + \bar{y}_{i,n}) - e_{3,i,n} \right], \\
\end{cases}
\]
\]
\[
\forall n \geq 0 \quad \begin{cases}
x_{n+1} \approx \bar{x}_n + \gamma_n (S + Q)x_n - (S + Q)p_n \\
\end{cases}
\]
\[
(3.115)
\]
Therefore, (3.115) is nothing else than
\[
(\forall n \geq 0) \quad \begin{cases}
x_n - \gamma_n (S + Q)x_n \in (\text{Id} + \gamma_n M) (\bar{x}_n - e_n) - \bar{e}_n, \\
x_{n+1} \approx \bar{x}_n + \gamma_n ((S + Q)x_n - (S + Q)p_n).
\end{cases}
\]
\[
(3.116)
\]
We now introduce the notations
\[
A_\mathcal{K} := M \quad \text{and} \quad B_\mathcal{K} := S + Q.
\]
\[
(3.117)
\]
Then (3.116) is
\[
(\forall n \geq 0) \quad \begin{cases}
\bar{x}_n = J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n + \bar{e}_n) + e_n, \\
x_{n+1} \approx \bar{x}_n + \gamma_n (B_\mathcal{K} x_n - B_\mathcal{K} \bar{x}_n).
\end{cases}
\]
\[
(3.118)
\]
We observe that $\bar{x}_n = J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n) + e_n$ with
\[
\sum_{n \geq 0} \|e_n^\mathcal{K}\| = \sum_{n \geq 0} \|J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n + \bar{e}_n) - J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n) + e_n\| \\
\leq \sum_{n \geq 0} \left[ \|J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n + \bar{e}_n) - J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n) \| + \|e_n\| \right] \\
\leq \sum_{n \geq 0} (\|\bar{e}_n\| + \|e_n\|) < +\infty.
\]
To this end, (3.118) becomes
\[
(\forall n \geq 0) \quad \begin{cases}
\bar{x}_n \approx J_{\gamma_n A_\mathcal{K}} (x_n - \gamma_n B_\mathcal{K} x_n), \\
x_{n+1} \approx \bar{x}_n + \gamma_n (B_\mathcal{K} x_n - B_\mathcal{K} \bar{x}_n).
\end{cases}
\]
\[
(3.119)
\]
which is an error-tolerant forward-backward-forward method in $\mathcal{K}$ whose convergence has been investigated in [43]. Note that the exact version of this algorithm was proposed by Tseng in [121].
(i) By [43, Theorem 2.5(i)], we have
\[ \sum_{n \geq 0} \| x_n - \bar{x}_n \|^2 < +\infty, \]
which yields \( \sum_{n \geq 0} \| x_n - \bar{x}_n \|^2 < +\infty \), and for all \( i = 1, \ldots, m \),
\[ \sum_{n \geq 0} \| p_{i,n} - \tilde{p}_{i,n} \|^2 < +\infty, \sum_{n \geq 0} \| q_{i,n} - \tilde{q}_{i,n} \|^2 < +\infty, \text{ and } \sum_{n \geq 0} \| y_{i,n} - \tilde{y}_{i,n} \|^2 < +\infty. \]

(ii) Let \( \bar{x} = (\bar{x}, \bar{p}, \bar{q}, \bar{y}, \bar{v}) \in \text{zer}(M + S + Q) \). Using [43, Theorem 2.5(ii)], we obtain \( x_n \rightharpoonup \bar{x} \) and \( \bar{x}_n \rightharpoonup \bar{x} \). In consideration of (3.99), it thus follows that \( (\bar{x}, \bar{p}, \bar{q}, \bar{v}) \) is a primal-dual solution to Problem 3.35, \( x_n \rightharpoonup \bar{x}, \bar{x}_n \rightharpoonup \bar{x} \), and
\[
\begin{aligned}
&\{ p_{i,n} \rightharpoonup \bar{p}_{i,n}, \quad q_{i,n} \rightharpoonup \bar{q}_{i,n}, \quad \tilde{q}_{i,n} \rightharpoonup \bar{q}_{i,n}, \quad \text{and} \quad \{ y_{i,n} \rightharpoonup \bar{y}_{i,n}, \quad \tilde{y}_{i,n} \rightharpoonup \bar{y}_{i,n}, \quad \forall i = 1, \ldots, m. \\
\end{aligned}
\]

\[\blacksquare\]

**Remark 3.43**  
(i) In contrast to Algorithm 3.37, the iterative scheme in Algorithm 3.41 requires twice the amount of forward steps and is therefore more time-intensive. On the other hand, many steps in Algorithm 3.41 can be processed in parallel.  
(ii) In [17], a related problem with linearly composed parallel sums was investigated. The algorithm there, however, is different to the one given in Algorithm 3.41.

### 3.3.4 Application to convex minimization

In this subsection we employ the algorithm and its convergence statement discussed in the previous one in the context of solving primal-dual pairs of convex optimization problems. The problem under consideration is as follows.

**Problem 3.44** Let \( \mathcal{H} \) be a real Hilbert space, \( z \in \mathcal{H} \), and \( f, h \in \Gamma(\mathcal{H}) \) such that \( h \) is differentiable with \( \mu \)-Lipschitzian gradient for \( \mu \in \mathbb{R}_{++} \). Furthermore, for every \( i = 1, \ldots, m \), let \( \mathcal{G}_i, \mathcal{X}_i, \mathcal{Y}_i \) be real Hilbert spaces, \( r_i \in \mathcal{G}_i \), let \( g_i \in \Gamma(\mathcal{X}_i) \) and \( l_i \in \Gamma(\mathcal{Y}_i) \), and consider the nonzero bounded linear operators \( L_i : \mathcal{H} \rightarrow \mathcal{G}_i \), \( K_i : \mathcal{G}_i \rightarrow \mathcal{X}_i \), and \( M_i : \mathcal{G}_i \rightarrow \mathcal{Y}_i \). Then we solve the primal optimization problem
\[
\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} \left( (g_i \circ K_i) \Box (l_i \circ M_i) \right)(L_i x - r_i) + h(x) - \langle x, z \rangle \right\} \tag{3.120}
\]
together with its conjugate dual problem
\[
\sup_{(p,q) \in \mathcal{X} \oplus \mathcal{Y}} \left\{ - (f^* \Box h^*)(z - \sum_{i=1}^{m} L^*_i K^*_i p_i) - \sum_{i=1}^{m} \left[ g^*_i(p_i) + l^*_i(q_i) + \langle p_i, K_i r_i \rangle \right] \right\}. \tag{3.121}
\]

For every \( x \in \mathcal{H} \) and \( (p, q) \in \mathcal{X} \oplus \mathcal{Y} \) with \( K^*_i p_i = M^*_i q_i, i = 1, \ldots, m \), by the Young–Fenchel inequality, it holds
\[
f(x) + h(x) + (f^* \Box h^*)\left( z - \sum_{i=1}^{m} L^*_i K^*_i p_i \right) \geq \left\langle z - \sum_{i=1}^{m} L^*_i K^*_i p_i, x \right\rangle,
\]
and, for any \( i = 1, \ldots, m \), and \( y_i \in \mathcal{G} \),
\[
g_i(K_i(L_i x - r_i - y_i)) + g_i^*(p_i) \geq \langle p_i, K_i(L_i x - r_i - y_i) \rangle = \langle K_i^* p_i, L_i x - r_i - y_i \rangle,
\]
and
\[
l_i(M_i y_i) + l_i^*(q_i) \geq \langle q_i, M_i y_i \rangle = \langle M_i^* q_i, y_i \rangle.
\]
This yields
\[
\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} \left( (g_i \circ K_i) \Box (l_i \circ M_i) \right)(L_i x - r_i) + h(x) - \langle x, z \rangle \right\}
\]
\[
= \inf_{(x,y) \in \mathcal{H} \oplus \mathcal{G}} \left\{ f(x) + \sum_{i=1}^{m} \left( g_i(K_i(L_i x - r_i - y_i)) + l_i(M_i y_i) \right) + h(x) - \langle x, z \rangle \right\}
\]
\[
\geq \sup_{(p,q) \in \mathcal{X} \oplus \mathcal{Y}, K_i^* p_i = M_i^* q_i, i = 1, \ldots, m} \left\{ - (f^* \Box h^*) \left( z - \sum_{i=1}^{m} L_i^* K_i^* p_i - \sum_{i=1}^{m} \left[ g_i^*(p_i) + l_i^*(q_i) + \langle p_i, K_i r_i \rangle \right] \right) \right\},
\]
which means that for the primal-dual pair of optimization problems (3.120)–(3.121) weak duality is always given.

Considering \((\bar{x}, \bar{p}, \bar{q}, \bar{y}) \in \mathcal{H} \oplus \mathcal{X} \oplus \mathcal{Y} \oplus \mathcal{G} \) a solution of the primal-dual system of monotone inclusions
\[
z - \sum_{i=1}^{m} L_i^* K_i^* p_i \in \partial f(\bar{x}) + \nabla h(\bar{x}) \quad \text{and} \quad K_i(l_i x - \bar{y}_i - r_i) \in \partial g_i^*(\bar{p}_i), \ M_i \bar{y}_i \in \partial l_i^*(\bar{q}_i), \ K_i^* \bar{p}_i = M_i^* \bar{q}_i, \ i = 1, \ldots, m,
\]
it follows that \( \bar{x} \) is an optimal solution to (3.120) and that \((\bar{p}, \bar{q})\) is an optimal solution to (3.121). Indeed, as \( h \) is convex and everywhere differentiable, it holds
\[
z - \sum_{i=1}^{m} L_i^* K_i^* p_i \in \partial f(\bar{x}) + \nabla h(\bar{x}) \subseteq \partial (f + h)(\bar{x}),
\]
thus,
\[
f(\bar{x}) + h(\bar{x}) + (f^* \Box h^*) \left( z - \sum_{i=1}^{m} L_i^* K_i^* p_i \right) = \left\langle z - \sum_{i=1}^{m} L_i^* K_i^* p_i, \bar{x} \right\rangle.
\]
On the other hand, since \( g_i \in \Gamma(\mathcal{X}_i) \) and \( l_i \in \Gamma(\mathcal{Y}_i) \), we have for any \( i = 1, \ldots, m \),
\[
g_i(K_i(L_i x - \bar{y}_i - r_i)) + g_i^*(\bar{p}_i) = \langle K_i^* \bar{p}_i, L_i x - r_i - \bar{y}_i \rangle,
\]
and
\[
l_i(M_i \bar{y}_i) + l_i^*(\bar{q}_i) = \langle M_i^* \bar{q}_i, \bar{y}_i \rangle.
\]
By summing up these equations and using (3.123), it yields
\[
f(\bar{x}) + \sum_{i=1}^{m} (g_i \circ K_i) \Box (l_i \circ M_i)(L_i x - r_i) + h(\bar{x}) - \langle \bar{x}, z \rangle
\]
\[
\leq f(\bar{x}) + \sum_{i=1}^{m} \left( g_i(K_i(L_i x - r_i - \bar{y}_i)) + l_i(M_i \bar{y}_i) \right) + h(\bar{x}) - \langle \bar{x}, z \rangle
\]
\[
= - (f^* \Box h^*) \left( z - \sum_{i=1}^{m} L_i^* K_i^* p_i \right) - \sum_{i=1}^{m} \left[ g_i^*(\bar{p}_i) + l_i^*(\bar{q}_i) + \langle \bar{p}_i, K_i r_i \rangle \right].
\]
which, together with (3.122), leads to the desired conclusion.

In the following, by extending the result in [17, Proposition 4.2] to our setting, we provide sufficient conditions which guarantee the validity of (3.97) when applied to convex minimization problems.

**Proposition 3.45** Suppose that the primal problem (3.120) has an optimal solution, that

\[ 0 \in \text{sqr}(\text{dom}(g_i \circ K_i)^* - \text{dom}(l_i \circ M_i)^*), \quad i = 1, \ldots, m, \]  

(3.124)

and

\[ 0 \in \text{sqr} E, \]  

(3.125)

where

\[ E := \left\{ \bigwedge_{i=1}^{m} \left\{ K_i(L_i(\text{dom } f) - r_i - y_i) - \text{dom } g_i \right\} \times \bigwedge_{i=1}^{m} \left\{ M_i y_i - \text{dom } l_i \right\} : y_i \in G_i, i = 1, \ldots, m \right\}. \]

Then

\[ z \in \text{ran} \left( \partial f + \sum_{i=1}^{m} L_i^* \left( (K_i^* \circ \partial g_i \circ K_i) \Box (M_i^* \circ \partial l_i \circ M_i) \right) (L_i \cdot -r_i) + \nabla h \right). \]

Proof. Let \( \bar{x} \in \mathcal{H} \) be an optimal solution to (3.120). Since (3.125) holds, we have that \((g_i \circ K_i), (l_i \circ M_i) \in \Gamma(G_i), i = 1, \ldots, m \). Further, because of (3.124), [11, Proposition 15.7] guarantees for any \( i = 1, \ldots, m \), the existence of \( \bar{y}_i \in G_i \) such that

\[ \left( (g_i \circ K_i) \Box (l_i \circ M_i) \right)(\bar{x}) = (g_i \circ K_i)(\bar{x} - \bar{y}_i) + (l_i \circ M_i)(\bar{y}_i). \]

Hence, \((\bar{x}, \bar{y}) = (\bar{x}, \bar{y}_1, \ldots, \bar{y}_m)\) is an optimal solution to the convex optimization problem

\[ \inf_{(x,y) \in \mathcal{H} \oplus \mathcal{G}} \left\{ f(x) + h(x) - \langle x, z \rangle + \sum_{i=1}^{m} \left[ g_i(K_i(L_i x - r_i - y_i)) + l_i(M_i y_i) \right] \right\}. \]  

(3.126)

By denoting

\[ f : \mathcal{H} \oplus \mathcal{G} \to \overline{\mathbb{R}}, \quad f(x, y) = f(x) + h(x) - \langle x, z \rangle, \]

\[ g : \mathcal{X} \oplus \mathcal{Y} \to \overline{\mathbb{R}}, \quad g(x, y) = \sum_{i=1}^{m} \left[ g_i(x_i - K_i r_i) + l_i(y_i) \right], \]  

(3.127)

\[ L : \mathcal{H} \oplus \mathcal{G} \to \mathcal{X} \oplus \mathcal{Y}, \quad (x, y) \mapsto \bigwedge_{i=1}^{m} \left\{ K_i(L_i x - y_i) \right\} \times \bigwedge_{i=1}^{m} \left\{ M_i y_i \right\}, \]

the problem given in (3.126) can be equivalently written as

\[ \inf_{(x,y) \in \mathcal{H} \oplus \mathcal{G}} \left\{ f(x, y) + g(L(x, y)) \right\}. \]  

(3.128)

Thus, by Fermat’s rule, we have

\[ 0 \in \partial(f + g \circ L)(\bar{x}, \bar{y}). \]
Since \( E = L(\text{dom } f) - \text{dom } g \) and (3.125) is fulfilled, it holds (see, for instance, [11, 22, 30])
\[
0 \in \partial \left( f + g \circ L \right)(\overline{x}, \overline{y}) = \partial f(\overline{x}, \overline{y}) + \left( L^* \circ \partial g \circ L \right)(\overline{x}, \overline{y}),
\]
where
\[
L^* : \mathcal{X} \oplus \mathcal{Y} \to \mathcal{H} \oplus \mathcal{G}, \quad (p, q) \mapsto \left( \sum_{i=1}^m L_i^* K_i^* p_i - K_i^* p_i + M_i^* q_i, \ldots, -K_m^* p_m + M_m^* q_m \right).
\]
We obtain
\[
0 \in \partial f(\overline{x}, \overline{y}) + \left( L^* \circ \partial g \circ L \right)(\overline{x}, \overline{y})
\]
\[
\Leftrightarrow \quad \left\{
\begin{array}{l}
0 \in \partial f(\overline{x}) + \nabla h(\overline{x}) - z + \sum_{i=1}^m L_i^* K_i^* \partial g_i(\overline{x}) (L_i \overline{x} - r_i - \overline{y}_i), \\
0 \in - (K_i^* \circ \partial g_i \circ K_i)(L_i \overline{x} - r_i - \overline{y}_i) + (M_i^* \circ \partial l_i \circ M_i) \overline{y}_i, \quad i = 1, \ldots, m,
\end{array}\right.
\]
\[
\Leftrightarrow \exists \mathbf{v} \in \mathcal{G} : 
\begin{align*}
0 &\in \partial f(\overline{x}) + \nabla h(\overline{x}) - z + \sum_{i=1}^m L_i^* v_i, \\
v_i &\in \left( K_i^* \circ \partial g_i \circ K_i \right) (L_i \overline{x} - r_i - \overline{y}_i), \quad i = 1, \ldots, m, \\
v_i &\in \left( M_i^* \circ \partial l_i \circ M_i \right) \overline{y}_i, \quad i = 1, \ldots, m,
\end{align*}
\]
\[
\Leftrightarrow \exists \mathbf{v} \in \mathcal{G} : 
\begin{align*}
0 &\in \partial f(\overline{x}) + \nabla h(\overline{x}) - z + \sum_{i=1}^m L_i^* v_i, \\
\overline{y}_i &\in \left( M_i^* \circ \partial l_i \circ M_i \right)^{-1} v_i, \quad i = 1, \ldots, m,
\end{align*}
\]
\[
\Leftrightarrow \exists \mathbf{v} \in \mathcal{G} : 
\begin{align*}
0 &\in \partial f(\overline{x}) + \nabla h(\overline{x}) - z + \sum_{i=1}^m L_i^* v_i, \\
v_i &\in \left( K_i^* \circ \partial g_i \circ K_i \right) - \left( M_i^* \circ \partial l_i \circ M_i \right) (L_i \overline{x} - r_i), \quad i = 1, \ldots, m,
\end{align*}
\]
\[
\Leftrightarrow z \in \partial f(\overline{x}) + \sum_{i=1}^m L_i^* \left( \left( K_i^* \circ \partial g_i \circ K_i \right) - \left( M_i^* \circ \partial l_i \circ M_i \right) \right) (L_i \overline{x} - r_i),
\]
which completes the proof.

**Remark 3.46** If one of the following two conditions

- \( f \) is real-valued and the operators \( L_i, K_i, \) and \( M_i \) are surjective for any \( i = 1, \ldots, m, \)
- the functions \( g_i \) and \( l_i \) are real-valued for any \( i = 1, \ldots, m, \)

is fulfilled, then \( E = \mathcal{X} \oplus \mathcal{Y} \) and (3.125) is obviously true.

On the other hand, if \( \mathcal{H}, \mathcal{G}_i, \mathcal{X}_i, \) and \( \mathcal{Y}_i, i = 1, \ldots, m, \) are finite dimensional and

for any \( i = 1, \ldots, m, \) there exists \( y_i \in \mathcal{G}_i : \begin{cases} K_i y_i \in K_i (L_i (\text{ri dom } f) - r_i) - \text{ri dom } g_i, \\ M_i y_i \in \text{ri dom } l_i, \end{cases} \)

then (3.125) is also true. This follows by using that in finite dimensional spaces the strong quasi-relative interior of a convex set is nothing else than its relative interior and by taking into account the properties of the latter.
An algorithm of forward-backward type

When applied to \( (3.123) \), the iterative scheme introduced in \( (3.96) \) and the corresponding convergence statements read as follows.

**Algorithm 3.47** Let \( x_0 \in \mathcal{H} \) and for every \( i = 1, \ldots, m \), let \( p_{i,0} \in \mathcal{X}_i \), \( q_{i,0} \in \mathcal{Y}_i \), and \( y_{i,0}, z_{i,0}, v_{i,0} \in \mathcal{G}_i \). For every \( i = 1, \ldots, m \), let \( \tau, \theta_{1,i}, \theta_{2,i}, \gamma_{1,i}, \gamma_{2,i} \), and \( \sigma_i \) be strictly positive real numbers such that

\[
2\mu^{-1} (1 - \overline{\alpha}) \min_{i=1,\ldots,m} \left\{ \frac{1}{\tau}, \frac{1}{\theta_{1,i}}, \frac{1}{\theta_{2,i}}, \frac{1}{\gamma_{1,i}}, \frac{1}{\gamma_{2,i}}, \frac{1}{\sigma_i} \right\} > 1, \tag{3.129}
\]

for

\[
\overline{\alpha} = \max \left\{ \frac{\tau}{\tau}, \max_{j=1,\ldots,m} \left\{ \sqrt{\theta_{1,j}}, \gamma_{1,j}, \sqrt{\theta_{2,j}}, \gamma_{2,j} \right\} \right\}.
\]

Furthermore, let \( \varepsilon \in (0, 1) \), let \( (\lambda_n)_{n \geq 0} \) be a sequence in \( [\varepsilon, 1] \), and set

\[
\begin{align*}
\tilde{x}_n &\approx \text{Prox}_{\tau f} (x_n - \tau (C x_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z)) , \\
\text{For } i = 1, \ldots, m \quad &
\begin{cases}
\tilde{p}_{i,n} &\approx \text{Prox}_{\theta_{1,i} L_i^*} (p_{i,n} + \theta_{1,i} K_i z_{i,n}) , \\
\tilde{q}_{i,n} &\approx \text{Prox}_{\theta_{2,i} L_i^*} (q_{i,n} + \theta_{2,i} M_i y_{i,n}) , \\
u_{i,n} &\approx \gamma_{1,i} (p_{i,n} - 2\tilde{p}_{i,n}) + v_{i,n} + \sigma_i (L_i (2\tilde{x}_n - x_n) - r_i) , \\
u_{i,n} &\approx \gamma_{2,i} (q_{i,n} - 2\tilde{q}_{i,n}) + v_{i,n} + \sigma_i (L_i (2\tilde{x}_n - x_n) - r_i) , \\
\tilde{z}_{i,n} &\approx (1 + \sigma_i (\gamma_{1,i} + \gamma_{2,i})) (u_{i,n} - \sigma \gamma_{1,i} u_{2,i,n}) , \\
\tilde{y}_{i,n} &\approx (1 + \sigma_i (\gamma_{1,i} + \gamma_{2,i})) (u_{2,i,n} - \sigma_i \gamma_{2,i} \tilde{z}_{i,n}) , \\
\tilde{v}_{i,n} &\approx v_{i,n} + \sigma_i (L_i (2\tilde{x}_n - x_n) - r_i - \tilde{z}_{i,n} - \tilde{y}_{i,n}) , \\
x_{n+1} &\approx x_n + \lambda_n (\tilde{x}_n - x_n) , \\
\text{For } i = 1, \ldots, m \quad &
\begin{cases}
p_{i,n+1} &\approx p_{i,n} + \lambda_n (\tilde{p}_{i,n} - p_{i,n}) , \\
q_{i,n+1} &\approx q_{i,n} + \lambda_n (\tilde{q}_{i,n} - q_{i,n}) , \\
z_{i,n+1} &\approx z_{i,n} + \lambda_n (\tilde{z}_{i,n} - z_{i,n}) , \\
y_{i,n+1} &\approx y_{i,n} + \lambda_n (\tilde{y}_{i,n} - y_{i,n}) , \\
v_{i,n+1} &\approx v_{i,n} + \lambda_n (\tilde{v}_{i,n} - v_{i,n}) .
\end{cases}
\end{cases}
\tag{3.130}
\end{align*}
\]

**Theorem 3.48** For Problem \( (3.44) \), suppose that

\[
z \in \text{ran} \left( \partial f + \sum_{i=1}^{m} L_i^* \left( (K_i^* \circ \partial g_i \circ K_i) \sqcup (M_i^* \circ \partial l_i \circ M_i) \right) (L_i \cdot r_i) + \nabla h \right) , \tag{3.131}
\]

and consider the sequences generated by Algorithm 3.47. Then there exists an optimal solution \( \overline{x} \) to \( (3.120) \) and optimal solution \( (\overline{p}, \overline{q}) \) to \( (3.121) \) such that

(i) \( x_n \rightarrow \overline{x} \), \( p_{i,n} \rightarrow \overline{p}_i \) and \( q_{i,n} \rightarrow \overline{q}_i \) for any \( i = 1, \ldots, m \) as \( n \rightarrow +\infty \),

(ii) if \( h \) is uniformly convex at \( \overline{x} \), then \( x_n \rightarrow \overline{x} \) as \( n \rightarrow +\infty \).

Proof. The results is a direct consequence of Theorem 3.38 when taking

\[
A = \partial f, \quad C = \nabla h, \quad \text{and} \quad B_i = \partial g_i, \quad D_i = \partial l_i, \quad i = 1, \ldots, m. \tag{3.132}
\]
3.3 Solving inclusions with parallel sums of linearly composed monotone operators

We also notice that, according to Theorem 20.40 in [11], the operators in (3.132) are maximally monotone, while, by [11, Corollary 16.24], we have \( A^{-1} = \partial f^*, C^{-1} = \partial h^*, B_i^{-1} = \partial g_i^* \) and \( D_i^{-1} = \partial l_i^* \) for \( i = 1, \ldots, m \). Furthermore, by [11, Corollary 18.16], \( C = \nabla h \) is \( \mu^{-1} \)-cocoercive, while, if \( h \) is uniformly convex at \( \tilde{\pi} \in \mathcal{H} \), then \( C = \nabla h \) is uniformly monotone at \( \tilde{\pi} \) (cf. [127, Section 3.4]).

**Remark 3.49** If \( h \in \Gamma(\mathcal{H}) \) is a constant function with \( \nabla h(x) = 0 \) for all \( x \in \mathcal{H} \), then condition (3.129) simplifies to

\[
\max \left\{ \sum_{i=1}^{m} \sigma_i \| L_i \|^2, \max_{j \in \{1, \ldots, m\}} \left\{ \theta_{1,j} \gamma_{1,j} \| K_j \|^2, \theta_{2,j} \gamma_{2,j} \| M_j \|^2 \right\} \right\} < 1.
\]

In this situation, Algorithm 3.47 converges under the relaxed assumption that \( (\lambda_n)_{n \geq 0} \subseteq [\varepsilon, 2 - \varepsilon] \) for \( \varepsilon \in (0, 1) \) (see also Remark 3.39).

**An algorithm of forward-backward-forward type**

On the other hand, when applied to (3.123), the iterative scheme introduced in (3.112) and the corresponding convergence statements read as follows.

**Algorithm 3.50** Let \( x_0 \in \mathcal{H} \), and for every \( i = 1, \ldots, m \), let \( p_{i,0} \in \mathcal{X}_i \), \( q_{i,0} \in \mathcal{Y}_i \), and \( z_{i,0}, y_{i,0}, v_{i,0} \in \mathcal{G}_i \). Set

\[
\beta = \mu + \sqrt{\max \left\{ \sum_{i=1}^{m} \| L_i \|^2, \max_{1 \leq i \leq m} \left\{ \| K_i \|^2, \| M_i \|^2 \right\} \right\}},
\]

let \( \varepsilon \in (0, 1/(\beta + 1)) \), let \( (\gamma_n)_{n \geq 0} \) be a sequence in \( [\varepsilon, (1 - \varepsilon)/\beta] \), and set

\[
(\forall n \geq 0) \quad \begin{aligned}
x_n &\approx \text{Prox}_{\gamma_n f} \left( x_n - \gamma_n \left( C x_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z \right) \right), \\
\tilde{x}_n &\approx \text{Prox}_{\gamma_n g} \left( x_n - \gamma_n \left( C x_n + \sum_{i=1}^{m} L_i^* v_{i,n} - z \right) \right), \\
p_{i,n} &\approx \text{Prox}_{\gamma_{i,n} l_i^*} \left( p_{i,0} + \gamma_{i,n} K_i z_{i,n} \right), \\
q_{i,n} &\approx \text{Prox}_{\gamma_{i,n} l_i^*} \left( q_{i,0} + \gamma_{i,n} M_i y_{i,n} \right), \\
u_{i,n} &\approx \gamma_{i,n} \left( L_i^* p_{i,n} - v_{i,n} - \gamma_{i,n} \left( L_i x_n - r_i \right) \right), \\
u_{i,n} &\approx \gamma_{i,n} \left( L_i^* q_{i,n} - v_{i,n} - \gamma_{i,n} \left( L_i x_n - r_i \right) \right), \\
z_{i,n} &\approx \frac{1 + \gamma_{i,n}}{1 + \gamma_{i,n}} \left( u_{i,n} - \gamma_{i,n}^2 z_{i,n} \right), \\
y_{i,n} &\approx \frac{1 + \gamma_{i,n}}{1 + \gamma_{i,n}} \left( u_{i,n} - \gamma_{i,n} z_{i,n} \right), \\
v_{i,n} &\approx \gamma_{i,n} \left( L_i x_n - r_i - z_{i,n} - \tilde{y}_{i,n} \right), \\
x_{n+1} &\approx x_n + \gamma_n \left( C x_n - C \tilde{x}_n + \sum_{i=1}^{m} L_i^* (v_{i,n} - \tilde{v}_{i,n}) \right), \\
\text{For } i = 1, \ldots, m \\
p_{i,n+1} &\approx p_{i,n} - \gamma_n \left( K_i (z_{i,n} - \tilde{z}_{i,n}) \right), \\
q_{i,n+1} &\approx q_{i,n} - \gamma_n \left( M_i (y_{i,n} - \tilde{y}_{i,n}) \right), \\
z_{i,n+1} &\approx \tilde{z}_{i,n} + \gamma_n \left( K_i^* (p_{i,n} - \tilde{p}_{i,n}) \right), \\
y_{i,n+1} &\approx \tilde{y}_{i,n} + \gamma_n \left( M_i^* (q_{i,n} - \tilde{q}_{i,n}) \right), \\
v_{i,n+1} &\approx \tilde{v}_{i,n} - \gamma_n \left( L_i (x_n - \tilde{x}_n) \right).
\end{aligned}
\]

**Theorem 3.51** For Problem 3.44, suppose that

\[
z \in \text{ran} \left( \partial f + \sum_{i=1}^{m} L_i^* \left( (K_i^* \circ \partial g_i \circ K_i) \square (M_i^* \circ \partial l_i \circ M_i) \right) (L_i \cdot -r_i) + \nabla h \right),
\]

(3.135)
and consider the sequences generated by Algorithm 3.50. Then there exists an optimal solution $\bar{x}$ to (3.120) and optimal solution $(\bar{p}, \bar{q})$ to (3.121), such that

(i) $\sum_{n \geq 0} \| x_n - \bar{x}_n \|^2 < +\infty$, and for $i = 1, \ldots, m,$

$$\sum_{n \geq 0} \| p_{i,n} - \tilde{p}_{i,n} \|^2 < +\infty, \text{ and } \sum_{n \geq 0} \| q_{i,n} - \tilde{q}_{i,n} \|^2 < +\infty,$$

(ii) $x_n \rightharpoonup \bar{x}, \tilde{x}_n \rightharpoonup \bar{x},$ and for $i = 1, \ldots, m,$

$$\left\{ \begin{array}{l}
  p_{i,n} \rightharpoonup \tilde{p}_{i,n}, \\
  \tilde{p}_{i,n} \rightharpoonup \bar{p}_{i,n},
\end{array} \right.$$

and

$$\left\{ \begin{array}{l}
  q_{i,n} \rightharpoonup \tilde{q}_{i,n}, \\
  \tilde{q}_{i,n} \rightharpoonup \bar{q}_{i,n}.
\end{array} \right.$$

Proof. The conclusions follow by using the statements in the proof of Theorem 3.48 and by applying Theorem 3.42. \[\square\]
Numerical experiments

In this chapter, we investigate the numerical performance of the methods proposed in Chapter 2 and Chapter 3 by solving convex minimization problems in different fields of applied mathematics. This also reflects the extensive applicability of our methods which is achieved by the sophisticated splitting philosophy and by the moderate assumptions imposed on the optimization problems.

The numerical tests are made on a system with Intel Core i7-3770 processor having 8GB DDR3 RAM. The Matlab codes and data sets can be found on the attached CD.

4.1 Image processing

For all applications discussed in this section the images have been normalized in order to make their pixels range in the closed interval from 0 to 1.

4.1.1 TV-based image denoising

Our first numerical experiment (cf. [38,39]) aims for the solving of an image denoising problem via total variation regularization (see, for instance, [32,46]). More precisely, we deal with the convex optimization problem

$$\inf_{x \in \mathbb{R}^n} \left\{ \lambda TV(x) + \frac{1}{2} \| x - b \|^2 \right\}, \quad (4.1)$$

where $\lambda \in \mathbb{R}_{++}$ is the regularization parameter, $TV : \mathbb{R}^n \to \mathbb{R}$ is a discrete total variation functional, and $b \in \mathbb{R}^n$ is the observed noisy image. Two such noisy observations are presented in Figure 4.1 where the test image was corrupted by adding Gaussian noise of different intensities.

Here, $x \in \mathbb{R}^n$ represents the vectorized image $X \in \mathbb{R}^{M \times N}$, where $n = MN$ (respectively $n = 3MN$ for color images) and $x_{i,j}$ denotes the normalized value of the pixel located in the $i$-th row and the $j$-th column, for $i = 1, \ldots, M$ and $j = 1, \ldots, N$. Two popular choices for the discrete total variation functional are the isotropic total...
variation $TV_{\text{iso}} : \mathbb{R}^n \to \mathbb{R}$,
\[
TV_{\text{iso}}(x) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} \\
+ \sum_{i=1}^{M-1} |x_{i+1,N} - x_{i,N}| + \sum_{j=1}^{N-1} |x_{M,j+1} - x_{M,j}|,
\]
and the anisotropic total variation $TV_{\text{aniso}} : \mathbb{R}^n \to \mathbb{R}$,
\[
TV_{\text{aniso}}(x) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}| \\
+ \sum_{i=1}^{M-1} |x_{i+1,N} - x_{i,N}| + \sum_{j=1}^{N-1} |x_{M,j+1} - x_{M,j}|,
\]
where in both cases reflexive (Neumann) boundary conditions are assumed. Note that the total variation model was first introduced by Rudin, Osher and Fatemi for image denoising and deconvolution in [115,116].

We introduce $\mathcal{Y} = \mathbb{R}^n \times \mathbb{R}^n$ and define the linear operator $L : \mathbb{R}^n \to \mathcal{Y}$, $x_{i,j} \mapsto (L_1x_{i,j}, L_2x_{i,j})$, where
\[
L_1x_{i,j} = \begin{cases} 
  x_{i+1,j} - x_{i,j}, & \text{if } i < M \\
  0, & \text{if } i = M
\end{cases} \quad \text{and} \quad L_2x_{i,j} = \begin{cases} 
  x_{i,j+1} - x_{i,j}, & \text{if } j < N \\
  0, & \text{if } j = N
\end{cases}.
\]
The operator $L$ represents a discretization of the gradient using reflexive (Neumann) boundary conditions and standard finite differences. One can easily check that $\|L\|^2 \leq 8$ while its adjoint $L^* : \mathcal{Y} \to \mathbb{R}^n$ is given in [46].

Within this example we focus on the anisotropic total variation function which is nothing else than the composition of the $\ell_1$-norm in $\mathcal{Y}$ with the linear operator $L$. Due to the full splitting characteristics of the iterative methods presented in this thesis, we only need to compute the proximal point of the conjugate of the $\ell_1$-norm, the latter being the indicator function of the dual unit ball. Thus, the calculation of the proximal point will result in the computation of a projection which admits an efficient implementation. The more challenging isotropic total variation functional is employed in one of the forthcoming subsections in the context of an image deblurring problem.

Thus, problem (4.1) reads equivalently
\[
\inf_{x \in \mathbb{R}^n} \{ h(x) + g(Lx) \},
\]
where $h : \mathbb{R}^n \to \mathbb{R}$, $h(x) = \frac{1}{2}\|x - b\|^2$, is 1-strongly convex and differentiable with 1-Lipschitzian gradient and $g : \mathcal{Y} \to \mathbb{R}$ is defined as $g(y_1, y_2) = \lambda \|(y_1, y_2)\|_1$. Then its conjugate $g^* : \mathcal{Y} \to \mathbb{R}$ is nothing else than
\[
g^*(p_1, p_2) = (\lambda \| \cdot \|_1)^*(p_1, p_2) = \lambda \left\| \left( \frac{p_1}{\lambda}, \frac{p_2}{\lambda} \right) \right\|_1^* = \delta_S(p_1, p_2) \forall (p_1, p_2) \in \mathcal{Y},
\]
where $S = [-\lambda, \lambda]^n \times [-\lambda, \lambda]^n$. Consequently, for arbitrary $\tau, \sigma \in \mathbb{R}_{++}$, and every
4.1 Image processing

(a) Noisy image, $\sigma = 0.06$  
(b) Noisy image, $\sigma = 0.12$

c) Denoised image, $\lambda = 0.035$  
(d) Denoised image, $\lambda = 0.07$

Figure 4.1: The noisy image in (a) was obtained after adding white Gaussian noise with standard deviation $\sigma = 0.06$ to the original $256 \times 256$ lichtenstein test image\(^1\), (c) shows the denoised image for $\lambda = 0.035$. Likewise, the noisy image when choosing $\sigma = 0.12$ and the denoised one for $\lambda = 0.07$ are shown in (b) and (d), respectively.

\(x, p, q \in \mathbb{R}^n\), it holds

\[
\text{Prox}_{\tau h}(x) = \arg \min_{z \in \mathbb{R}^n} \left\{ \frac{\tau}{2} \| z - b \|^2 + \frac{1}{2} \| z - x \|^2 \right\} = x + \frac{\tau b}{1 + \tau},
\]

\[
\text{Prox}_{\sigma g}(p, q) = \arg \min_{(z_1, z_2) \in S} \frac{1}{2} \| (z_1, z_2) - (p, q) \|^2 = P_{[-\lambda, \lambda]^n \times [-\lambda, \lambda]^n}(p, q).
\]

We solve the regularized image denoising problem with the two Douglas–Rachford type primal-dual methods DR1 (cf. Algorithm 3.29) and DR2 (cf. Algorithm 3.31), the forward-backward-forward type primal dual method FBF (cf. [58, Theorem 3.1]) and its acceleration FBF Acc (cf. Algorithm 3.11), the primal-dual method PD and its accelerated version PD Acc, both given in [48], the alternating minimization algorithm AMA from [120] together with its Nesterov-type acceleration AMA Acc (cf. [78]), as well as the Nesterov (cf. [100]) and FISTA (cf. [15, 16]) algorithm which are operating on the dual problem. A comparison on the obtained results is shown in Table 4.1.

As measure of performance we use the so-called root-mean-square error (RMSE), which is defined as

\[
\text{RMSE}_k = \sqrt{\frac{1}{n} \sum_{i=1}^{n} ((x_k)_i - x^*_i)^2} = \frac{\| x_k - x^* \|}{\sqrt{n}},
\]

where \(x_k\) and \(x^*\) are, respectively, the current iterate at iteration \(k \in \mathbb{N}\) and the unique optimizer.

\(^1\)see http://commons.wikimedia.org/wiki/File:Lichtenstein_img_processing_test.png
\[
\begin{array}{cccccc}
\sigma = 0.12, \lambda = 0.07 & \sigma = 0.06, \lambda = 0.035 \\
\varepsilon = 10^{-4} & \varepsilon = 10^{-6} & \varepsilon = 10^{-4} & \varepsilon = 10^{-6} \\
\hline
\text{DR1} & 1.40s (48) & 3.35s (118) & 1.31s (45) & 2.93s (103) \\
\text{DR2} & 1.24s (75) & 2.82s (173) & 1.12s (66) & 2.57s (147) \\
\text{FBF} [58] & 8.89s (343) & 58.96s (2271) & 4.86s (187) & 41.21s (1586) \\
\text{FBF Acc} & 2.63s (101) & 11.73s (451) & 1.93s (73) & 8.07s (308) \\
\text{PD} [48] & 5.26s (337) & 35.53s (2226) & 2.77s (183) & 25.53s (1532) \\
\text{PD Acc} [48] & 1.42s (96) & 7.26s (447) & 1.20s (70) & 5.44s (319) \\
\text{AMA} [120] & 7.29s (471) & 46.76s (3031) & 3.98s (254) & 34.36s (2184) \\
\text{AMA Acc} [78] & 1.83s (89) & 11.68s (561) & 1.41s (63) & 8.39s (383) \\
\text{Nesterov} [100] & 1.97s (102) & 12.45s (595) & 1.51s (72) & 8.77s (415) \\
\text{FISTA} [15, 16] & 1.71s (100) & 10.92s (645) & 1.14s (70) & 7.41s (429) \\
\end{array}
\]

Table 4.1: Performance evaluation for the images in Figure 4.1. The entries refer to the CPU times in seconds and the number of iterations, respectively, needed in order to attain a root-mean-square error for the iterates below the tolerance \( \varepsilon \).

It shows that the two Douglas–Rachford type methods are performing well against the others, especially when the accuracy increases from \( \varepsilon = 10^{-4} \) to \( \varepsilon = 10^{-6} \). Although DR1 requires less iterations than DR2, the two additional linear operator evaluations, which appear in each iteration, prevent DR1 from being the fastest algorithm in this comparison. One can also see, that Algorithm 3.11 (FBF Acc) is clearly faster than the ordinary method proposed by Combettes and Pesquet (FBF) in [58]. The same applies for the accelerated primal-dual algorithm (PD Acc) proposed by Chambolle and Pock in [48].

4.1.2 TV-based image denoising involving higher-order derivatives

Within this subsection we solve image denoising problems (cf. [37]) where first- and second-order total variation functionals are linked via infimal convolutions. This approach was initially proposed in [47] and further investigated in [117]. The two different convex optimization problems under investigation are (cf. [117])

\[
(\ell_2^2\text{-IC/P}) \quad \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \| x - b \|^2 + \left( (\alpha_1 \cdot \| \cdot \|_1 \circ D_1) \Box (\alpha_2 \cdot \| \cdot \|_1 \circ D_2) \right)(x) \right\}, \quad (4.3)
\]

and

\[
(\ell_2^2\text{-MIC/P}) \quad \inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \| x - b \|^2 + \left( (\alpha_1 \cdot \| \cdot \|_1 \circ L_1) \Box (\alpha_2 \cdot \| \cdot \|_1 \circ L_1) \right)(D_1 x) \right\}.
\]

By making use of [11, Corollary 15.28(i) and Proposition 12.34(i)], one can show that condition (3.124) in Subsection 3.3.4 is fulfilled for both problems above which implies that the infimal convolutions occurring in the objective functions are exact. Therefore, the objective functions above are proper, strongly convex, and lower semicontinuous such that both denoising problems have unique solutions. Here, \( b \in \mathbb{R}^n \) is the observed and vectorized noisy image of size \( M \times N \) (with \( n = MN \) for grayscale and \( n = 3MN \) for colored images). In consideration of the approach
described in [117], for \( y = (y_1^T, \ldots, y_k^T)^T \in \mathbb{R}^{kn} \) and \( \omega = (\omega_1, \ldots, \omega_k) \in \mathbb{R}^k_{++} \), the \( \ell_1 \)-norms on \( \mathbb{R}^{kn} \) are defined as

\[
\| y \|_{1,\omega} = \left\| \left( \omega_1 y_1^2 + \ldots + \omega_k y_k^2 \right)^{\frac{1}{2}} \right\|_1,
\]

where vector multiplications and square roots are understood to be componentwise. Therefore the regularizers correspond to isotropic total variation functionals. For the bounded linear operators occurring in (4.3) and (4.4), we refer to [117, Example 2.2 and Example 3.1]. Therefore, we take the forward difference matrix

\[
D_k := \begin{bmatrix}
-1 & 1 & 0 & 0 & \ldots & 0 \\
0 & -1 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & -1 & 1 & 0 \\
0 & \ldots & 0 & 0 & -1 & 1 \\
0 & \ldots & 0 & 0 & 0 & 0
\end{bmatrix} \in \mathbb{R}^{k \times k}
\]

into account which models the discrete first-order derivative. Note that \(-D_k^T D_k\) is then an approximation of the second-order derivative. By \( A \otimes B \), we denote the Kronecker product of \( A \) and \( B \), and, by letting \( D_x \) and \( D_y \) be the vertical and horizontal difference operators, respectively, we have

\[
D_x = \text{Id}_N \otimes D_M, \quad \mathcal{D}_1 = \begin{bmatrix} D_x & D_y \end{bmatrix}, \quad D_{xx} = \text{Id}_N \otimes (-D_M^T D_M), \quad D_{yy} = (-D_N^T D_N) \otimes \text{Id}_M, \quad \mathcal{D}_2 = \begin{bmatrix} D_{xx} \\ D_{yy} \end{bmatrix}. \tag{4.6}
\]

Following the considerations made in [117], we let \( \omega_1 = (1, 1) \) and \( \omega_2 = (1, 1) \). The operator \( L_1 \) in (4.4) fulfills \( \mathcal{D}_2 = L_1 \mathcal{D}_1 \) and is therefore chosen to be

\[
L_1 = \begin{bmatrix} -D_x^T & 0 \\ 0 & -D_y^T \end{bmatrix}.
\]

For other discrete second-order derivatives also involving mixed partial derivatives (in horizontal-vertical direction or vice versa), we refer to the literature above.

In order to compare our methods from Section 3.3 with algorithms relying on (augmented) Lagrangian and smoothing techniques, note that, using the definition of the infimal convolution, (4.3) and (4.4) can be formulated as constrained problems of the form

\[
(\ell_2^2 \text{-IC/P}) \quad \inf_{x_1, x_2, z_1, z_2} \left\{ \frac{1}{2} \| x_1 + x_2 - b \|_2^2 + \alpha_1 \| z_1 \|_{1,\omega_1} + \alpha_2 \| z_2 \|_{1,\omega_2} \right\} \\
\text{subject to } \begin{bmatrix} \mathcal{D}_1 & 0 \\ 0 & \mathcal{D}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \tag{4.7}
\]

and

\[
(\ell_2^2 \text{-MIC/P}) \quad \inf_{x, y_1, y_2, z} \left\{ \frac{1}{2} \| x - b \|_2^2 + \alpha_1 \| y_1 \|_{1,\omega_1} + \alpha_2 \| z \|_{1,\omega_2} \right\} \\
\text{subject to } \begin{bmatrix} \mathcal{D}_1 & -\text{Id} \\ 0 & L_1 \end{bmatrix} \begin{bmatrix} x \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z \end{bmatrix}, \tag{4.8}
\]

respectively.

By taking into account Problem 3.44, it shows that (4.3) and (4.4) can be considered as special instances of this general problem description. Hence, we let \( \mathcal{Y} = \mathbb{R}^n \times \mathbb{R}^n \),
and introduce the functions $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{2} \|x - b\|^2$, as well as $g : \mathcal{Y} \to \mathbb{R}$, $g = \alpha_1 \| \cdot \|_{1,\omega_1}$, and $l : \mathcal{Y} \to \mathbb{R}$, $l = \alpha_2 \| \cdot \|_{1,\omega_2}$.

The proximal point with respect to $f$ was already given in the subsection before. On the other hand, the proximal point mappings with respect to the conjugates $g^*$ and $l^*$ are projections onto nonempty, closed, and convex sets. Indeed, by making use of convex analysis, the conjugate of $g$, i.e., $g^* : \mathcal{Y} \to \mathbb{R}$, becomes for arbitrary $(p, q) \in \mathcal{Y}$,

$$g^*(p, q) = (\alpha_1 \| \cdot \|_{1,\omega_1})^*(p, q) = \alpha_1(\| \cdot \|_{1,\omega_1})^*\left(\frac{(p, q)}{\alpha_1}\right) = \delta_S(p, q),$$

where (cf. [39])

$$S = \left\{(p, q) \in \mathcal{Y} : \max_{1 \leq i \leq n} \sqrt{p_i^2 + q_i^2} \leq \alpha_1\right\}. \quad (4.9)$$

To this end, by letting $\sigma \in \mathbb{R}_{++}$, and $p, q \in \mathbb{R}^n$ be arbitrary, we have $\text{Prox}_{\sigma g^*}(p, q) = \mathcal{P}_S(p, q)$, where the projection operator $\mathcal{P}_S : \mathcal{Y} \to S$ is defined via

$$(p_i, q_i) \mapsto \alpha_1 \frac{(p_i, q_i)}{\max\left\{\alpha_1, \sqrt{p_i^2 + q_i^2}\right\}}, \quad 1 \leq i \leq n. \quad (4.10)$$

The proximal point with respect to the conjugate $l^*$ can be similarly obtained by using $\alpha_2$ rather than $\alpha_1$.

For our numerical tests we still consider the colored test image lichtenstein (see Figure 4.2) of size $256 \times 256$. By adding white Gaussian noise with standard deviation $0.08$, we obtain the noisy image $b \in \mathbb{R}^n$. The regularization parameters in ($\ell_2^2$-IC/P) and ($\ell_2^2$-MIC/P) are set to $\alpha_1 = 0.06$ and $\alpha_2 = 0.2$. When measuring the quality of the restored images, we use the improvement in signal-to-noise ratio (ISNR), which is given by (cf. [49])

$$\text{ISNR}_k = 10 \log_{10} \left( \frac{\|x - b\|^2}{\|x - x_k\|^2} \right);$$
where $x$, $b$, and $x_k$ are the original, the observed noisy and the reconstructed image at iteration $k \in \mathbb{N}$, respectively.

In Figure 4.3, for solving (4.3) and (4.4), we compare different optimization algorithms with Algorithm 3.47 (FB) and Algorithm 3.50 (FBF) from Section 3.3.

The double smoothing (DS) algorithm as proposed in Section 2.1 is applied to the duals of (4.7) and (4.8) by considering the acceleration strategies in Subsection 2.1.8. One should notice that, since the smoothing parameters are constant, DS solves continuously differentiable approximations of (4.7) and (4.8) and does therefore not necessarily converge to the unique minimizers of (4.3) and (4.4). As a second smoothing algorithm, we make use of the variable smoothing technique (VS) in Section 2.2 which successively reduces the smoothing parameter in each iteration and therefore solves the initial problems as the iteration counter increases. We further consider the primal-dual hybrid gradient method (PDHG) as discussed in [117] which is nothing else than the primal-dual method in [48]. Finally, the alternating direction method of multipliers (ADMM) is applied to (4.7) as discussed in [117]. Here, one needs to apply the Moore-Penrose inverse of a special bounded linear operator which can be implemented efficiently since $D_1^T D_1$ and $D_2^T D_2$ can be diagonalized by the discrete cosine transform. The problem which arises in (4.8), however, is far more difficult to solve with this method (and is therefore not implemented) since the bounded linear operator which needs to be inverted has a more complicated structure. This reveals a typical drawback of ADMM since it does not provide a full splitting like primal-dual or smoothing algorithms.

The FBF method suffers from its additional forward step when taking into account the comparison shown in Figure 4.3. However, many time-intensive steps in this algorithm can be executed in parallel which would reduce the execution time significantly. On the other hand, the FB method performs fast and stable in both examples while considerable differences in the reconstructions for ($\ell_2^2$-IC/P) and ($\ell_2^2$-MIC/P) can not be observed.
Chapter 4 Numerical experiments

(a) Original image  (b) Blurred and noisy image  (c) Reconstructed image

Figure 4.4: Figure (a) shows the clean 256 \times 256 cameraman test image\(^2\), (b) shows the image obtained after multiplying it with a blur operator and adding white Gaussian noise with standard deviation 10\(^{-3}\), and (c) shows the reconstructed image generated by Algorithm 3.29.

4.1.3 TV-based image deblurring

The third numerical experiment in this section concerns the solving of an ill-conditioned linear inverse problem arising in image deblurring. For a given matrix \(A \in \mathbb{R}^{n \times n}\) describing a blur (or averaging) operator and a given vector \(b \in \mathbb{R}^n\) representing the blurred and noisy image, our aim is to estimate the unknown original image \(x \in \mathbb{R}^n\) fulfilling

\[
Ax = b.
\]

To this end, we are solving the regularized, convex, and nondifferentiable problem

\[
\inf_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|_1 + \alpha_2 \|Wx\|_1 + \alpha_1 TV(x) + \delta_{[0,1]^n}(x) \right\}, \tag{4.11}
\]

where the regularization is done by a combination of two functionals with different properties. Here, \(\alpha_1, \alpha_2 \in \mathbb{R}_{++}\) are regularization parameters, \(TV : \mathbb{R}^n \to \mathbb{R}\) is the discrete isotropic total variation functional and \(W : \mathbb{R}^n \to \mathbb{R}^n\) is the discrete Haar wavelet transform with four levels. We would also like to point out that none of the functions occurring in (4.11) is differentiable.

In this example, we make use of the popular cameraman test image. The picture undergoes a Gaussian blur of size 9 \times 9 with standard deviation 4, as done in [39, Section 4.2], yielding a blurring operator \(A\) with \(\|A\|^2 = 1\) and \(A^* = A\). Figure 4.4 shows the original, observed and reconstructed versions of the 256 \times 256 cameraman test image.

Since we are using the isotropic total variation function as regularizer, we have to take the following into account. Recall that \(Y = \mathbb{R}^n \times \mathbb{R}^n\) and consider the bounded linear operator \(L\) defined in Subsection 4.1.1 having norm \(\|L\| \leq \sqrt{8}\). For \(\omega = (1, 1)\), we consider the \(\ell_1\) norms introduced in (4.5) and obtain the equivalent characterization for the isotropic total variation functional of being \(TV_{\omega_0}(x) = \|Lx\|_1\omega\).

Consequently, the optimization problem (4.11) can be equivalently written as

\[
\inf_{x \in \mathbb{R}^n} \left\{ f(x) + g_1(Ax) + g_2(Wx) + g_3(Lx) \right\}, \tag{4.12}
\]

\(^2\)The cameraman test image is part of the image processing toolbox in Matlab.
Figure 4.5: The evolution of the values of the objective function and of the ISNR (improvement in signal-to-noise ratio) for Algorithm 3.29 (DR1), Algorithm 3.31 (DR2) and the forward-backward-forward method (FBF) from [58, Theorem 3.1].

where \( f : \mathbb{R}^n \to \mathbb{R} \), \( f(x) = \delta_{[0,1]^n}(x) \), \( g_1 : \mathbb{R}^n \to \mathbb{R} \), \( g_1(y) = \|y-b\|_1 \), \( g_2 : \mathbb{R}^n \to \mathbb{R} \), \( g_2(y) = \alpha_2\|y\|_1 \), and \( g_3 : \mathcal{Y} \to \mathbb{R} \), \( g_3(y,z) = \alpha_1 \|(y,z)\|_{1,\omega} \). The proximal points of these functions admit explicit representations. Indeed, for every \( p \in \mathbb{R}^n \), we have \( g_1^*(p) = \delta_{[-1,1]^n}(p) + p^T b \) and \( g_2^*(p) = \delta_{[-\alpha_2,\alpha_2]^n}(p) \) (see, for instance [22]), while, for every \( (p,q) \in \mathcal{Y} \), it holds \( g_3^*(p,q) = \delta_S(p,q) \), where \( S \subseteq \mathcal{Y} \) is given in (4.9). To this end, for all \( x, p, q \in \mathbb{R}^n \), it holds

\[
\text{Prox}_{\tau f}(x) = \arg\min_{z \in [0,1]^n} \frac{1}{2} \|z - x\|^2 = \mathcal{P}_{[0,1]^n}(x),
\]

\[
\text{Prox}_{\sigma_1 g_1^*}(p) = \arg\min_{z \in [-1,1]^n} \left\{ \sigma_1 z^T b + \frac{1}{2} \|z - p\|^2 \right\} = \mathcal{P}_{[-1,1]^n}(p - \sigma_1 b),
\]

\[
\text{Prox}_{\sigma_2 g_2^*}(p) = \arg\min_{z \in [-\alpha_2,\alpha_2]^n} \frac{1}{2} \|z - p\|^2 = \mathcal{P}_{[-\alpha_2,\alpha_2]^n}(p),
\]

\[
\text{Prox}_{\sigma_3 g_3^*}(p,q) = \arg\min_{(z_1,z_2) \in S} \frac{1}{2} \|(z_1,z_2) - (p,q)\|^2 = \mathcal{P}_S(p,q),
\]

where the projection operator \( \mathcal{P}_S : \mathcal{Y} \to S \) is defined as in (4.10).

Figure 4.5 shows the performance of Algorithm 3.29 (DR1) and Algorithm 3.31 (DR2) when solving (4.12) for \( \sigma_1 = 0.003 \) and \( \sigma_2 = 0.001 \), by making use of the starting points \( x_0 = b \) and \( (v_{1,0},v_{2,0},v_{3,0}) = (0,0,0) \) and parameters

- **DR1 (Algorithm 3.29):** \( \sigma_1 = 1 \), \( \sigma_2 = 0.05 \), \( \sigma_3 = 0.05 \), \( \tau = 3.99 (\sigma_1 + \sigma_2 + 8\sigma_3)^{-1} \), \( \lambda_n = 1.7 \) for every \( n \geq 0 \),
- **DR2 (Algorithm 3.31):** \( \sigma_1 = 1 \), \( \sigma_2 = 0.05 \), \( \sigma_3 = 0.05 \), \( \tau = 0.99 (\sigma_1 + \sigma_2 + 8\sigma_3)^{-1} \), \( \lambda_n = 1.6 \) for every \( n \geq 0 \).

Furthermore, we compare the two Douglas–Rachford methods with the iterative scheme designed in [58, Theorem 3.1] for

- **FBF (58, Theorem 3.1)):** \( \varepsilon = \frac{1}{50(\sqrt{1+18+1})} \), \( \gamma_n = \frac{1-\varepsilon}{\sqrt{1+18}} \) for every \( n \geq 0 \), within the first 10 seconds when applied to the 256 \( \times \) 256 cameraman test image. It shows that both Douglas–Rachford methods clearly outperform the forward-backward-forward splitting method in terms of function value decrease and improvement in signal-to-noise ratio.
110 Chapter 4 Numerical experiments

(a) Original image  
(b) 80% missing pixels  
(c) Reconstructed image

Figure 4.6: Figure (a) shows the 240 × 256 clean fruits image, (b) shows the same image for which 80% randomly chosen pixels were set to black, and (c) shows the solution generated by Algorithm 3.29 (DR1) after 400 iterations.

4.1.4 TV-based image inpainting

In the last image processing example (cf. [39]), we show how image inpainting problems, which aim for recovering lost information, can be efficiently solved via the primal-dual methods investigated in this work. To this end, we consider the following TV-regularized model

\[
\inf \, TV_{iso}(x),
\]

s.t. \( Kx = b, \)

\[x \in [0, 1]^n\]

where \( TV_{iso} : \mathbb{R}^n \rightarrow \mathbb{R}\) is the isotropic total variation functional and \( K \in \mathbb{R}^{n \times n}\) is a diagonal matrix. Here, for \( i = 1, \ldots, n\), we have \( K_{i,i} = 0\), if the pixel \( i\) in the noisy image \( b \in \mathbb{R}^n\) is lost (in our case set to black) and \( K_{i,i} = 1\), otherwise. The induced linear operator \( K : \mathbb{R}^n \rightarrow \mathbb{R}^n\) fulfills \( \|K\| = 1\), while, in the light of the considerations made in the previous subsections, we have that \( TV_{iso}(x) = \|Lx\|_{1,\omega}\) for \( \omega = (1, 1)\) and all \( x \in \mathbb{R}^n\).

Thus, problem (4.13) can be formulated as

\[
\inf \, \{f(x) + g_1(Lx) + g_2(Kx)\}.
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = \delta_{[0,1]^n}, g_1 : \mathcal{Y} \rightarrow \mathbb{R}, g_1(y_1, y_2) = \|(y_1, y_2)\|_{1,\omega}, \) and \( g_2 : \mathbb{R}^n \rightarrow \mathbb{R}, g_2(y) = \delta_{[b]}(y). \) We solve this problem via Algorithm 3.29 (DR1), while the formulae for the proximal points involved in this iterative scheme have already been given in former subsections. Figure 4.6 shows the original fruits image, the image obtained from it after setting 80% randomly chosen pixels to black and the image reconstructed by Algorithm 3.29.

4.2 Kernel based machine learning

The following numerical experiment (cf. [40]) concerns the solving of the problem of classifying images via support vector machines classification, an approach which

\[^{3}\text{see http://www.hlevkin.com/TestImages/fruits.bmp}\]
4.2 Kernel based machine learning

belongs to the class of kernel based learning methods.

The given data set consisting of 11339 training images and 1850 test images of size 28 × 28 was taken from the website http://www.cs.nyu.edu/~roweis/data.html. The problem we consider is to determine a decision function based on a pool of handwritten digits showing either the number five or the number six, labeled by +1 and −1, respectively (see Figure 4.7). Subsequently, we evaluate the quality of the decision function on the test data set by computing the percentage of misclassified images. In order to reduce the computational effort, we use only half of the available images from the training data set.

Figure 4.7: A sample of images belonging to the classes +1 and −1, respectively.

The classifier functional \( f \) is assumed to be an element of the Reproducing Kernel Hilbert Space (RHKS) \( \mathcal{H}_\kappa \), which in our case is induced by the symmetric and finitely positive definite Gaussian kernel function

\[
\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad \kappa(x, y) = \exp \left( -\frac{\|x - y\|^2}{2\sigma^2_\kappa} \right),
\]

where \( \sigma_\kappa \in \mathbb{R}_{++} \) denotes the kernel parameter. Let \( \langle \cdot, \cdot \rangle_\kappa \) be the inner product on \( \mathcal{H}_\kappa \), \( \| \cdot \|_\kappa \) the corresponding norm and \( K \in \mathbb{R}^{n \times n} \) the Gram matrix with respect to the training data set \( Z = \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \subseteq \mathbb{R}^d \times \{+1, -1\} \), namely the symmetric and positive definite matrix with entries \( K_{ij} = \kappa(X_i, X_j) \) for \( i, j = 1, \ldots, n \). Within this example we make use of the hinge loss \( v : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad v(x, y) = \max\{1 - xy, 0\} \), which penalizes the deviation between the predicted value \( f(x) \) and the true value \( y \in \{+1, -1\} \). The smoothness of the decision function \( f \in \mathcal{H}_\kappa \) is employed by means of the smoothness functional \( \Omega : \mathcal{H}_\kappa \to \mathbb{R}, \quad \Omega(f) = \|f\|_\kappa^2 \), taking high values for nonsmooth functions and low values for smooth ones. The decision function \( f \) we are looking for is the optimal solution of the Tikhonov regularization problem

\[
\inf_{f \in \mathcal{H}_\kappa} \left\{ C \sum_{i=1}^n v(f(X_i), Y_i) + \frac{1}{2}\|f\|_\kappa^2 \right\}, \quad (4.14)
\]

where \( C > 0 \) denotes the regularization parameter controlling the tradeoff between the loss function and the smoothness functional.

The representer theorem (cf. [118]) ensures the existence of a vector of coefficients \( c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n \) such that the minimizer \( f \) of (4.14) can be expressed as a kernel expansion in terms of the training data, i.e., \( f(\cdot) = \sum_{i=1}^n c_i \kappa(\cdot, X_i) \). Thus, the smoothness functional becomes \( \Omega(f) = \|f\|_\kappa^2 = \langle f, f \rangle_\kappa = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \kappa(X_i, X_j) = c^T K c \) and for \( i = 1, \ldots, n \), it holds \( f(X_i) = \sum_{j=1}^n c_j \kappa(X_i, X_j) = (K c)_i \). Hence, in
Thus, for every $c \in \mathbb{R}^n$, it holds (see, also, [33, 41])

$$
g^*(p) = \sup_{z \in \mathbb{R}^n} \left\{ (p, z) - C \sum_{i=1}^{n} v(z_i, Y_i) \right\} = \sum_{i=1}^{n} (Cv(\cdot, Y_i))^*(p_i) = C \sum_{i=1}^{n} v(\cdot, Y_i)^* \left( \frac{p_i}{C} \right)
$$

Thus, for $\sigma \in \mathbb{R}_{++}$ and $c \in \mathbb{R}^n$, we have

$$
\text{Prox}_{\sigma g^*}(c) = \arg\min_{p \in \mathbb{R}^n} \left\{ \sigma C \sum_{i=1}^{n} v(\cdot, Y_i)^* \left( \frac{p_i}{C} \right) + \frac{1}{2} \|p - c\|^2 \right\}
$$

$$
= \arg\min_{p_i \in [-C, 0], i = 1, \ldots, n} \left\{ \sum_{i=1}^{n} \left[ \sigma p_i Y_i + \frac{1}{2} (p_i - c_i)^2 \right] \right\}
$$

$$
= \left( P_{Y_i [-C, 0]} (c_1 - \sigma Y_1), \ldots, P_{Y_n [-C, 0]} (c_n - \sigma Y_n) \right)^T.
$$

With respect to the considered dataset, we denote by

$$
D = \{(X_i, Y_i), \; i = 1, \ldots, 5670\} \subseteq \mathbb{R}^{784} \times \{+1, -1\}
$$

the set of available training data consisting of 2711 images in the class +1 and 2959 images in the class −1. Notice that a sample from each class of images is shown in Figure 4.7. Due to numerical reasons, the images have been normalized (cf. [84]) by dividing each of them by the quantity $\left( \frac{1}{5670} \sum_{i=1}^{5670} ||X_i||^2 \right)^{\frac{1}{2}}$.

In order to specify a good choice for the kernel parameter $\sigma_\kappa \in \mathbb{R}_{++}$ and the tradeoff parameter $C \in \mathbb{R}_{++}$, we tested different combinations of them with the forward-backward (FB) solver given in [122]. The results are shown in Table 4.2, whereby the combination $\sigma_\kappa = 0.25$ and $C = 1$ provides with $0.7027\%$ the lowest
4.3 The generalized Heron problem

We consider the generalized Heron problem as we did in [38], which has been recently investigated in [94,95], and where for its solving subgradient-type methods have been used.

While the classical Heron problem concerns the finding of a point \( \pi \) on a given straight line in the plane such that the sum of distances from \( \pi \) to given points \( u^1, u^2 \) is minimal, the problem that we address here aims to find a point in a closed convex set \( \Omega \subseteq \mathbb{R}^n \) which minimizes the sum of the distances to given closed convex sets \( \Omega_i \subseteq \mathbb{R}^n, i = 1, \ldots, m \).

The distance from a point \( x \in \mathbb{R}^n \) to a nonempty set \( \Omega \subseteq \mathbb{R}^n \) is given by

\[
d(x; \Omega) = (\| \cdot \| \delta_{\Omega})(x) = \inf_{z \in \Omega} \| x - z \|.
\]

Thus the generalized Heron problem reads

\[
\inf_{x \in \Omega} \sum_{i=1}^{m} d(x; \Omega_i),
\]

where the sets \( \Omega \subseteq \mathbb{R}^n \) and \( \Omega_i \subseteq \mathbb{R}^n, i = 1, \ldots, m, \) are nonempty, closed, and convex. We observe that (4.16) perfectly fits into the framework considered in Problem 3.28 when setting

\[
f = \delta_{\Omega}, \text{ and } g_i = \| \cdot \|, l_i = \delta_{\Omega_i} \text{ for all } i = 1, \ldots, m.
\]

However, note that (4.16) cannot be solved via the primal-dual methods in [58] and [122] since they require the presence of at least one strongly convex function.

Table 4.3: Performance evaluation for the SVM problem using \( C = 1 \) and \( \sigma = 0.25 \). The entries refer to the CPU times in seconds and the number of iterations.

<table>
<thead>
<tr>
<th>Method</th>
<th>CPU Time (s) (Iterations)</th>
<th>RMSE ≤ 10^{-3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB [122]</td>
<td>3.07s (113)</td>
<td>19.50s (717)</td>
</tr>
<tr>
<td>FB Acc [26]</td>
<td>95.33s (3522)</td>
<td>348.41s (12923)</td>
</tr>
<tr>
<td>FBF [58]</td>
<td>4.36s (80)</td>
<td>32.92s (606)</td>
</tr>
<tr>
<td>FBF Acc</td>
<td>3.63s (67)</td>
<td>32.90s (606)</td>
</tr>
<tr>
<td>VS</td>
<td>8.03s (300)</td>
<td>108.64s (4076)</td>
</tr>
</tbody>
</table>

misclassification rate. This means that among the 1870 images belonging to the test data set, 13 of them were not correctly classified.

Table 4.3 shows some results when solving the classification problem (4.15) via the variable smoothing algorithm (VS) and via primal-dual methods which are able to perform a forward step on the operator \( \nabla h \). Since the matrix \( K \in \mathbb{R}^{n \times n} \) is positive definite, the function \( h(c) = c^T K c \) is strongly convex as well. Hence there exists a unique solution to (4.15) and we can also apply the accelerated versions of the FB and of the FBF method given in [26] and Algorithm 3.11, respectively. However, we notice that the acceleration of the forward-backward primal-dual method (FB Acc) converges extremely slow in this example.
(cf. Baillon–Haddad Theorem, [8,10]) in each of the infimal convolutions \( \| \cdot \| \Box \delta_{\Omega_i} \), \( i = 1, \ldots, m \), a fact which is obviously not the case. Be aware that

\[
g_i^* : \mathbb{R}^n \to \mathbb{R}, \quad g_i^*(p) = \sup_{x \in \mathbb{R}^n} \{ \langle p, x \rangle - \| x \| \} = \delta_{B(0,1)}(p), \quad i = 1, \ldots, m,
\]

thus the proximal points of \( f \), \( g_i^* \), and \( l_i^* \), \( i = 1, \ldots, m \), can be calculated via projections, in case of the latter via Moreau’s decomposition formula.

In the following we are testing our algorithms on some examples taken from [94,95].

**Example 4.1** (Example 5.5 in [95]) Consider problem (4.16) with the constraint set \( \Omega \) being the closed ball centered at \((5, 0)\) having radius 2, and the sets \( \Omega_i \), \( i = 1, \ldots, 8 \), being pairwise disjoint squares in right position in \( \mathbb{R}^2 \) (i.e., the edges are parallel to the x- and y-axes, respectively), with centers \((-2, 4)\), \((-1, -8)\), \((0, 0)\), \((0, 6)\), \((5, -6)\), \((8, -8)\), \((8, 9)\), and \((9, -5)\), and side length 1, respectively (see Figure 4.8).

![Figure 4.8: Example 4.1. Generalized Heron problem with squares and disc constraint set on the left-hand side, and performance evaluation for the root-mean-square error (RMSE) on the right-hand side.](image.png)

When solving this problem with Algorithm 3.29 (DR1) and Algorithm 3.31 (DR2) and the choices made in (4.17), the following formulae for the proximal points involved in their formulations are necessary for \( x \), \( p \in \mathbb{R}^2 \), and \( \tau, \sigma_i \in \mathbb{R}_{++}, i = 1, \ldots, 8 \):

\[
\text{Prox}_{x_f}(x) = (5, 0) + \arg\min_{y \in B(0,2)} \frac{1}{2} \| y - (x - (5, 0)) \|^2 = (5, 0) + \mathcal{P}_{B(0,2)}(x - (5, 0)),
\]

\[
\text{Prox}_{\Omega, g_i}(p) = \arg\min_{z \in B(0,1)} \frac{1}{2} \| z - p \|^2 = \mathcal{P}_{B(0,1)}(p),
\]

\[
\text{Prox}_{\Omega, l_i}(p) = (p - \sigma_i \text{Prox}_{\sigma_i^{-1}l_i}
\left( \frac{p}{\sigma_i} \right)) = p - \sigma_i \arg\min_{z \in \Omega_i} \frac{1}{2} \| z - \frac{p}{\sigma_i} \|^2 = p - \sigma_i \mathcal{P}_{\Omega_i}
\left( \frac{p}{\sigma_i} \right).
\]

Figure 4.8 gives an insight into the performance of the proposed primal-dual methods when compared with the subgradient algorithm used in [95]. After a few milliseconds, both splitting algorithms reach machine precision with respect to the root-mean-square error where the following parameters are used:
4.4 Portfolio optimization under different risk measures

- DR1: \( \forall i = 1, \ldots, 8, \sigma_i = 0.15, \tau = 2/(\sum_{j=1}^{8} \sigma_j), \lambda_n = 1.5, x_0 = (5, 2), v_{i,0} = 0, \)
- DR2: \( \forall i = 1, \ldots, 8, \sigma_i = 0.1, \tau = 0.24/(\sum_{j=1}^{8} \sigma_j), \lambda_n = 1.8, x_0 = (5, 2), v_{i,0} = 0, \)
- Subgradient (cf. [95, Theorem 4.1]) \( x_0 = (5, 2), \alpha_n = \frac{1}{n}. \)

**Example 4.2** (Example 4.3 in [94]) In this example we solve the generalized Heron problem (4.16) in \( \mathbb{R}^3 \), where the constraint set \( \Omega \) is the closed ball centered at \((0, 2, 0)\) with radius 1, and \( \Omega_i, i = 1, \ldots, 5 \), are cubes in right position with center at \((-4, 0, 0), (-4, 2, -3), (-3, -4, 2), (-5, 4, 4), \) and \((-1, 8, 1)\), and side length 2, respectively.

Figure 4.9 shows the example taken from [94], where we use the following parameters for initialization:

- DR1: \( \forall i = 1, \ldots, 5, \sigma_i = 0.3, \tau = 2/(\sum_{j=1}^{5} \sigma_j), \lambda_n = 1.5, x_0 = (0, 2, 0), v_{i,0} = 0, \)
- DR2: \( \forall i = 1, \ldots, 5, \sigma_i = 0.2, \tau = 0.24/(\sum_{j=1}^{5} \sigma_j), \lambda_n = 1.8, x_0 = (0, 2, 0), v_{i,0} = 0, \)
- Subgradient (cf. [94, Theorem 4.1]) \( x_0 = (0, 2, 0), \alpha_n = \frac{1}{n}. \)

Figure 4.9: Example 4.2. Generalized Heron problem with cubes and ball constraint set on the left-hand side, and performance evaluation for the RMSE on the right-hand side.

Once again, after a few milliseconds, the Douglas–Rachford type primal-dual methods DR1 and DR2 reach machine precision, whereas the method proposed in [94] has not terminated after passing the 10 seconds barrier.

4.4 Portfolio optimization under different risk measures

In financial mathematics, quantifying the risk of future random outcomes is a principal concern for decision makers who, naturally, have their own attitude towards risk. In the classical portfolio theory by Markowitz (cf. [90]), the variance was used to measure the uncertainty of prospective outcomes. However, as was already discussed by Markowitz in [91], asymmetry is a desirable property for risk measures since investors have differing stances on rising or falling courses.

In the following, by making use of primal-dual methods, we solve portfolio optimization problems under different convex risk measures as described in our paper [36].
Chapter 4 Numerical experiments

For preliminaries on notations and properties of probability spaces, the Optimized Certainty Equivalent, or utility functions, we refer to Section 1.2.

We consider a portfolio with a number of \( N \geq 1 \) different positions with returns \( R_i \in L^2, i = 1, \ldots, N \), a nonzero vector of expected returns \( \mu = (\mathbb{E}[R_1], \ldots, \mathbb{E}[R_N])^T \), and \( \mu^* \leq \max_{i = 1, \ldots, N} \mathbb{E}[R_i] \) a given lower bound for the expected return of the portfolio. In the following, by making use of different utility functions, we are solving the optimization problem

\[
\inf_{x^T \mu \geq \mu^*, \ x^T \mathbb{1}_N = 1, \ x = (x_1, \ldots, x_N)^T \in \mathbb{R}_N^N} \rho_u \left( \sum_{i=1}^N x_i R_i \right), \tag{4.18}
\]

which assumes the minimization of the risk of the portfolio subject to constraints on the expected return of the portfolio and on the budget. Recall that \( \mathbb{1}_N \) denotes the vector in \( \mathbb{R}^N \) having all entries equal to 1. The constraint \( x = (x_1, \ldots, x_N)^T \in \mathbb{R}_N^N \) means that we are not allowing short sales in these particular examples. However, they could be easily considered as well if desired. By using (1.20), we obtain the following reformulation of problem (4.18)

\[
\inf_{x^T \mu \geq \mu^*, \ x^T \mathbb{1}_N = 1, \ x = (x_1, \ldots, x_N)^T \in \mathbb{R}_N^N, \ \lambda \in \mathbb{R}} \left\{ \lambda + \mathbb{E} \left[ u \left( \sum_{i=1}^N x_i R_i + \lambda \right) \right] \right\}, \tag{4.19}
\]

which will prove to be more suitable for being solved by means of primal-dual proximal splitting algorithms. In this sense, the following result, which relates the optimal solutions of the two optimization problems is of certain importance.

**Proposition 4.3** The following statements are true.

(a) If \((\bar{x}, \lambda)\) is an optimal solution to (4.19), for \( \bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)^T \), then \( \bar{x} \) is an optimal solution to (4.18).

(b) If \( \bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)^T \) is an optimal solution to (4.18) and

\[
\bar{\lambda} \in \arg \min_{\lambda \in \mathbb{R}} \left\{ \lambda + \mathbb{E} \left[ u \left( \sum_{i=1}^N \bar{x}_i R_i + \lambda \right) \right] \right\},
\]

then \((\bar{x}, \bar{\lambda})\) is an optimal solution to (4.19).

**Proof.** Denote by \( \mathcal{X} = \left\{ x \in \mathbb{R}_N^N : x^T \mu \geq \mu^*, \ x^T \mathbb{1}_N = 1 \right\} \).

(a) Since \((\bar{x}, \bar{\lambda}) \in \mathcal{X} \times \mathbb{R} \) is an optimal solution to (4.19), we have for every \((x, \lambda) \in \mathcal{X} \times \mathbb{R} \)

\[
\lambda + \mathbb{E} \left[ u \left( \sum_{i=1}^N x_i R_i + \lambda \right) \right] \geq \bar{\lambda} + \mathbb{E} \left[ u \left( \sum_{i=1}^N \bar{x}_i R_i + \lambda \right) \right] \geq \rho_u \left( \sum_{i=1}^N \bar{x}_i R_i \right).
\]

Passing to the infimum over \( \lambda \in \mathbb{R} \) yields

\[
\rho_u \left( \sum_{i=1}^N x_i R_i \right) \geq \rho_u \left( \sum_{i=1}^N \bar{x}_i R_i \right) \ \forall x \in \mathcal{X},
\]

hence, \( \bar{x} \in \mathcal{X} \) is an optimal solution to (4.18).
4.4 Portfolio optimization under different risk measures

(b) The conclusion follows by noticing that for every \( (x, \lambda) \in X \times \mathbb{R} \), we have

\[
\lambda + \mathbb{E} \left[ u \left( \sum_{i=1}^{N} x_i R_i + \lambda \right) \right] \geq \rho_u \left( \sum_{i=1}^{N} R_i x_i \right) \geq \rho_u \left( \sum_{i=1}^{N} x_i R_i \right) = \lambda + \mathbb{E} \left[ u \left( \sum_{i=1}^{N} x_i R_i + \lambda \right) \right],
\]

which completes the proof.

**Remark 4.4** A sufficient condition guaranteeing that

\[
\arg \min_{\lambda \in \mathbb{R}} \{ \lambda + \mathbb{E} \left[ u (X + \lambda) \right] \} \neq \emptyset \quad \forall X \in L^2
\]

was given in [31, Theorem 4] and reads

\[
\{ d \in \mathbb{R} : u_\infty (d) = -d \} = \{ 0 \}, \tag{4.20}
\]

where \( u_\infty : \mathbb{R} \to \mathbb{R} \), \( u_\infty (d) = \sup \{ u(x+d) - u(x) : x \in \text{dom} \ u \} \), denotes the recession function of the function \( u \). Moreover, in the light of the same result, it follows that under (4.20),

\[
\rho_u (X) = \sup_{\Xi \in L^2} \{ \langle X, \Xi \rangle - \mathbb{E} [ u^*(\Xi) ] \} \quad \forall X \in L^2,
\]

thus \( \rho_u \) is lower semicontinuous. Since \( X = \{ x \in \mathbb{R}^N : x^T \mu \geq \mu^* , x^T 1^N = 1 \} \) is compact, this further implies that (4.18) has an optimal solution and, consequently, that (4.19) has an optimal solution, too. All particular convex utility functions we deal with in this section fulfill condition (4.20).

According to Proposition 4.3, determining an optimal solution to problem (4.19) will lead to an optimal solution to the portfolio optimization problem (4.18). However, as we will show in the following, problem (4.19) is a particular case of Problem 3.4, thus it can be solved by primal-dual proximal splitting methods. In order to show this, let us first consider the bounded linear operator

\[
K : \mathbb{R}^N \times \mathbb{R} \to L^2, \quad (x_1, \ldots, x_n, \lambda) \mapsto \sum_{i=1}^{N} x_i R_i + \lambda.
\]

In order to determine its adjoint operator \( K^* : L^2 \to \mathbb{R}^N \times \mathbb{R} \), we use that

\[
\langle K(x, \lambda), Z \rangle = \int_{\Omega} \left( \sum_{i=1}^{N} x_i R_i(\omega) + \lambda \right) Z(\omega) \, d\mathbb{P}(\omega) = \sum_{i=1}^{N} x_i \langle R_i , Z \rangle + \lambda \langle 1, Z \rangle
\]

\[= \langle (x, \lambda), K^* Z \rangle,
\]

for all \( (x, \lambda) \in \mathbb{R}^N \times \mathbb{R} \) and all \( Z \in L^2 \). Thus, we get

\[
K^* Z = (\langle R_1, Z \rangle, \ldots, \langle R_N, Z \rangle, \mathbb{E} [ Z ])^T \quad \forall Z \in L^2.
\]

Further, we introduce the closed convex sets

\[
S = \left\{ x \in \mathbb{R}^N : x^T \mu \geq \mu^* \right\}, \quad T = \left\{ x \in \mathbb{R}^N : x^T 1^N = 1 \right\}.
\]
and obtain the unconstrained problem
\[
\inf_{(x,\lambda) \in \mathbb{R}^N \times \mathbb{R}} \left\{ \delta_{\mathbb{R}^+} (x) + \lambda + \delta_{S \times \mathbb{R}} (x, \lambda) + \delta_{T \times \mathbb{R}} (x, \lambda) + (\mathbb{E} [u \circ K]) (x, \lambda) \right\}.
\] (4.21)

When calculating the proximal points of these functions, one has to project onto \( \mathbb{R}^N \), \( S \), and \( T \), where explicit formulae exist (cf. [11, Examples 3.21 and 28.16]). The proximal point with respect to the function \( E[u] \) can be obtained via the following proposition (cf. [36]), which makes use of the interchangeability of integration and minimization derived in [114, Theorem 14.60].

**Proposition 4.5** For arbitrary random variables \( X \in L^2 \) and \( \gamma \in \mathbb{R}^{++} \), it holds
\[
\text{Prox}_{\gamma E[u]} (X)(\omega) = \text{Prox}_{\gamma u} (X(\omega)) \text{ \( \forall \omega \in \Omega \) a.s.,}
\] (4.22)

**Proof.** We have
\[
\text{Prox}_{\gamma E[u]} (X) = \arg \min_{Y \in L^2} \left\{ \gamma \mathbb{E} [u(Y)] + \frac{1}{2} \|Y - X\|^2 \right\}
\]
\[
= \arg \min_{Y \in L^2} \left\{ \gamma \int_{\Omega} u(Y(\omega)) d\mathbb{P}(\omega) + \frac{1}{2} \int_{\Omega} (Y(\omega) - X(\omega))^2 d\mathbb{P}(\omega) \right\}
\]
\[
= \arg \min_{Y \in L^2} \int_{\Omega} \left( \gamma u(Y(\omega)) + \frac{1}{2} (Y(\omega) - X(\omega))^2 \right) d\mathbb{P}(\omega).
\]

Hence, using the interchangeability of integration and minimization (see [114, Theorem 14.60]), we have
\[
\text{Prox}_{\gamma E[u]} (X)(\omega) = \arg \min_{y \in \mathbb{R}} \left\{ \gamma u(y) + \frac{1}{2} (y - X(\omega))^2 \right\} = \text{Prox}_{\gamma u} (X(\omega)) \text{ \( \forall \omega \in \Omega \) a.s.,}
\]
which completes the proof. \( \blacksquare \)

In what follows, we provide explicit formulae for the proximal points of some popular convex utility functions considered in the literature, which will be of importance for the numerical experiments presented later and which involve the convex risk measures relying on them.

**Example 4.6** (Piecewise linear utility) For \( \gamma_2 < -1 < \gamma_1 \leq 0 \), we consider the piecewise linear utility function
\[
u_1 : \mathbb{R} \to \mathbb{R}, \quad u_1(t) = \begin{cases} 
\gamma_2 t, & \text{if } t \leq 0 \\
\gamma_1 t, & \text{if } t > 0 
\end{cases} = \gamma_1 [t]_+ - \gamma_2 [t]_-.
\]
Assumption 1.16 is fulfilled since \( u_1(0) = 0 \) and \( -1 \in \partial u_1(0) = [\gamma_2, \gamma_1] \) and, since for all \( d \in \mathbb{R} \) (see [31])
\[
(u_1)_\infty (d) = \begin{cases} 
\gamma_2 d, & \text{if } d < 0 \\
0, & \text{if } d = 0 \\
\gamma_1 d, & \text{if } d > 0 
\end{cases}
\]
condition (4.20) is fulfilled, as well. Hence, \( u_1 \) gives rise to the lower semicontinuous coherent risk measure
\[
\rho_{u_1}(X) = \inf_{\lambda \in \mathbb{R}} \left\{ \lambda + \gamma_1 \mathbb{E} [X + \lambda]_+ - \gamma_2 \mathbb{E} [X + \lambda]_- \right\} \forall X \in L^2.
\] (4.23)
For every \( \gamma \in \mathbb{R}_{++} \) and \( t \in \mathbb{R} \), it holds
\[
\text{Prox}_{\gamma u_1}(t) = \arg \min_{s \in \mathbb{R}} \left\{ \gamma \left( \gamma_1 [s]_+ - \gamma_2 [s]_- \right) + \frac{1}{2} (s - t)^2 \right\} = \begin{cases} 
  t - \gamma \gamma_2, & \text{if } t < \gamma \gamma_2 \\
  0, & \text{if } t \in [\gamma \gamma_2, \gamma \gamma_1] \\
  t - \gamma \gamma_1, & \text{if } t > \gamma \gamma_1
\end{cases}
\]

When setting \( \gamma_1 = 0 \) and \( \gamma_2 = -\frac{1}{\alpha} \) for some \( \alpha \in (0,1) \), the convex risk measure (4.23) becomes the classical so-called Conditional Value-at-Risk at level \( \alpha \) (see, for example, [110,111])
\[
\text{CVaR}_\alpha : L^2 \rightarrow \mathbb{R}, \quad \text{CVaR}_\alpha(X) = \inf_{\lambda \in \mathbb{R}} \left\{ \lambda + \frac{1}{1 - \alpha} \mathbb{E} [X + \lambda]_- \right\}. \quad (4.24)
\]
The infimum in the expression of the Conditional Value-at-Risk is attained for every \( X \in L^2 \) at the so-called Value-at-Risk at level \( \alpha \), i.e.,
\[
\text{VaR}_\alpha(X) = \arg \min_{\lambda \in \mathbb{R}} \left\{ \lambda + \frac{1}{1 - \alpha} \mathbb{E} [X + \lambda]_- \right\}.
\]

**Example 4.7** (Exponential utility function) Consider the exponential utility function \( u_2 : \mathbb{R} \rightarrow \mathbb{R}, \ u_2(t) = \exp(-t) - 1 \). It fulfills Assumption 1.16 and, since \( (u_2)_\infty = \delta_{[0, +\infty)} \), condition (4.20) is fulfilled, as well. It gives rise via (1.20) to the so-called entropic risk measure
\[
\rho_{u_2}(X) = \inf_{\lambda \in \mathbb{R}} \left\{ \lambda + \mathbb{E} [\exp(-X - \lambda) - 1] \right\} \quad \forall X \in L^2, \quad (4.25)
\]
which is a lower semicontinuous convex risk measure. For arbitrary \( \gamma \in \mathbb{R}_{++} \) and \( t \in \mathbb{R} \), it holds
\[
\text{Prox}_{\gamma u_2}(t) = \arg \min_{s \in \mathbb{R}} \left\{ \gamma (\exp(-s) - 1) + \frac{1}{2} (s - t)^2 \right\} = W(\gamma \exp(-t)) + t,
\]
where \( W \) denotes the Lambert \( W \) function (cf. [61]). Although no closed form expression for this function can be given, the Symbolic Math Toolbox in Matlab provides the routine \texttt{lambertw} to compute \( \text{Prox}_{\gamma u_2}(t) \). Alternatively, these proximal points can efficiently be calculated by applying Newton’s method under the use of previous iterates as starting points.

**Example 4.8** (Indicator utility function) By choosing the utility function \( u_3 : \mathbb{R} \rightarrow \mathbb{R}, \ u_3(t) = \delta_{\mathbb{R}_+}(t) \), one has \( (u_3)_\infty = \delta_{\mathbb{R}_+} \), thus, both Assumption 1.16 and condition (4.20) are fulfilled. It gives rise to the so-called worst-case risk measure
\[
\rho_{u_3}(X) = \inf_{\lambda \in \mathbb{R}} \scriptstyle{\underbrace{\lambda}_{X+\lambda\geq0}} = -\text{essinf} X = \text{esssup}(-X) \quad \forall X \in L^2, \quad (4.26)
\]
which is a lower semicontinuous convex risk measure. For arbitrary \( \gamma \in \mathbb{R}_{++} \) and \( t \in \mathbb{R} \), it holds
\[
\text{Prox}_{\gamma u_3}(t) = \arg \min_{s \in \mathbb{R}} \left\{ \gamma \delta_{\mathbb{R}_+}(s) + \frac{1}{2} (s - t)^2 \right\} = \mathcal{P}_{\mathbb{R}_+}(t).
\]
Example 4.9 (Quadratic utility function) For a fixed $\beta \in \mathbb{R}_{++}$, we consider the quadratic utility function

$$u_4 : \mathbb{R} \to \mathbb{R}, \quad u_4(t) = \begin{cases} \frac{\beta}{2} t^2 - t, & \text{if } t \leq \frac{1}{\beta} \\ -\frac{1}{2\beta}, & \text{if } t > \frac{1}{\beta}. \end{cases}$$

Obviously, $(u_4)_\infty = \delta_{[0, \infty)}$, thus, both Assumption 1.16 and condition (4.20) are also fulfilled for this utility function. For arbitrary $\gamma \in \mathbb{R}_{++}$ and $t \in \mathbb{R}$, it holds

$$\operatorname{Prox}_{\gamma u_4}(t) = \arg\min_{s \in \mathbb{R}} \left\{ \gamma u_4(s) + \frac{1}{2} (s - t)^2 \right\} = \begin{cases} \frac{t + \sqrt{t^2 + 2\gamma \beta}}{2\beta}, & \text{if } t \leq \frac{1}{\beta} \\ t, & \text{if } t > \frac{1}{\beta}. \end{cases}$$

Example 4.10 (Logarithmic utility function) For $\theta \in \mathbb{R}_{++}$, we consider the logarithmic utility function

$$u_5 : \mathbb{R} \to \mathbb{R}, \quad u_5(t) = \begin{cases} -\theta \ln \left( 1 + \frac{t}{\theta} \right), & \text{if } t > -\theta \\ +\infty, & \text{if } t \leq -\theta. \end{cases}$$

For this special utility function, one can also show that $(u_5)_\infty = \delta_{[0, \infty)}$, hence that (4.20) is fulfilled. The properties in Assumption 1.16 hold as well and therefore, via (1.20), we obtain the convex risk measure

$$\rho_{u_5}(X) = \inf_{X + \lambda > -\theta} \{ \lambda - \theta \mathbb{E} \left[ \ln \left( 1 + \frac{X + \lambda}{\theta} \right) \right] \} \quad \forall X \in L^2.$$ 

The proximal points of the logarithmic utility function take an explicit expression. For arbitrary $\gamma \in \mathbb{R}_{++}$ and $t \in \mathbb{R}$, it holds

$$\operatorname{Prox}_{\gamma u_5}(t) = \arg\min_{s \in \mathbb{R}} \left\{ -\gamma \theta \ln \left( 1 + \frac{s}{\theta} \right) + \frac{1}{2} (s - t)^2 \right\} = \frac{t - \theta}{2} + \sqrt{\frac{(\theta - t)^2}{4} + \theta(\gamma + t)}.$$ 

For the experiments described as follows we took weekly opening courses over the last 13 years from assets belonging to the indices DAX and NASDAQ, in order to obtain the returns $R_i \in \mathbb{R}^{|\Omega|}$, $i = 1, \ldots, N$, for $|\Omega| = 689$ and $N = 106$. The data was provided by the Yahoo finance database. Assets which do not support the required historical information like Volkswagen AG (DAX) or Netflix, Inc. (NASDAQ) were not taken into account.

<table>
<thead>
<tr>
<th>$\mu^*$</th>
<th>linear ($\alpha = 0.95$)</th>
<th>exponential</th>
<th>indicator</th>
<th>quadr. ($\beta = 1$)</th>
<th>log. ($\theta = 5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.14s (500)</td>
<td>0.18s (402)</td>
<td>- (&gt; 15000)</td>
<td>0.05s (170)</td>
<td>0.53s (1891)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.15s (520)</td>
<td>0.15s (336)</td>
<td>- (&gt; 15000)</td>
<td>0.06s (196)</td>
<td>0.38s (1335)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.33s (1202)</td>
<td>0.31s (682)</td>
<td>- (&gt; 15000)</td>
<td>0.06s (186)</td>
<td>0.72s (2570)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.32s (1164)</td>
<td>0.40s (885)</td>
<td>- (&gt; 15000)</td>
<td>0.08s (272)</td>
<td>1.07s (3820)</td>
</tr>
<tr>
<td>1.1</td>
<td>0.41s (1526)</td>
<td>6.80s (15222)</td>
<td>- (&gt; 15000)</td>
<td>0.14s (486)</td>
<td>1.18s (4198)</td>
</tr>
<tr>
<td>1.3</td>
<td>0.42s (1570)</td>
<td>5.45s (12155)</td>
<td>- (&gt; 15000)</td>
<td>0.41s (1476)</td>
<td>6.61s (23547)</td>
</tr>
</tbody>
</table>

Table 4.4: CPU times in seconds and the number of iterations when solving the portfolio optimization problem (4.18) under different utility functions.
For solving the portfolio optimization problem (4.18), we took different convex risk measures into consideration which were induced by linear, exponential, indicator, quadratic, and logarithmic utility functions. We applied Algorithm 3.29 (DR1) to the unconstrained problem in (4.21), while using formulae for the proximal points of each utility function given in the examples above. The values of the expected returns associated with $R_i, i = 1, \ldots, N$, ranged from $-0.2690$ (Commerzbank AG, DAX) to $1.4156$ (priceline.com Incorporated, NASDAQ).

Computational results on this problem are reported in Table 4.4 for different values of $\mu^*$. We terminate the algorithm when subsequent iterates start to stay within an accuracy level of 1% with respect to the set of constraints and to the optimal objective value. It shows that the worst-case risk measure, which is obtained by using the indicator utility, performs poorly on the given dataset, while it seems that the algorithm is sensitive to the lower bound on the expected return $\mu^*$.

### 4.5 Clustering

In cluster analysis one aims for grouping a set of points such that points within the same group (usually measured via distance functions) are more similar to each other than to points in other groups. Clustering can be formulated as a convex optimization problem (see, for instance, [50, 83, 87]). In this example (cf. [26]), we are solving the problem

$$
\inf_{x_i \in \mathbb{R}^n, i=1, \ldots, m} \left\{ \frac{1}{2} \sum_{i=1}^{m} \|x_i - u_i\|^2 + \gamma \sum_{i<j} \omega_{ij} \|x_i - x_j\|_p \right\},
$$

where $\gamma \in \mathbb{R}_+$ is a tuning parameter, $p \in \{1, 2\}$, and $\omega_{ij} \in \mathbb{R}_+$ represent weights on the terms $\|x_i - x_j\|_p$, for $i, j \in \{1, \ldots, m\}, i < j$. For each given point $u_i \in \mathbb{R}^n, i = 1, \ldots, m$, the variable $x_i \in \mathbb{R}^n$ corresponds to the associated cluster center. In [83], the authors consider $\ell_1, \ell_2,$ and $\ell_\infty$ norms on the penalty terms $x_i - x_j$ while in [87] arbitrary $\ell_p$ norms were taken into account. Since the objective function is strongly convex, there exists a unique solution to (4.27).

The tuning parameter $\gamma \in \mathbb{R}_+$ plays a central role for the results on the clustering problem. Taking $\gamma = 0$, each cluster center $x_i$ will coincide with the associated point $u_i$. As $\gamma$ increases, the cluster centers will start to coalesce where two points $u_i, u_j$ are said to belong to the same cluster when $x_i = x_j$. One obtains a single cluster containing all points when $\gamma$ becomes sufficiently large.

Moreover, the choice on the weights is important as well since cluster centers may coalesce promptly as $\gamma$ passes certain critical values. For our weights, we use a $K$-
Table 4.5: Performance evaluation for the clustering problem. The entries refer to the CPU times in seconds and the number of iterations, respectively, needed in order to attain a root-mean-square error for the iterates below the tolerance $\varepsilon$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$p = 2$, $\gamma = 5.2$</th>
<th>$p = 1$, $\gamma = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varepsilon = 10^{-4}$</td>
<td>$\varepsilon = 10^{-8}$</td>
</tr>
<tr>
<td>DR1</td>
<td>0.78s (216)</td>
<td>1.68s (460)</td>
</tr>
<tr>
<td>DR2</td>
<td>0.61s (323)</td>
<td>1.20s (644)</td>
</tr>
<tr>
<td>FBF [58]</td>
<td>7.67s (2123)</td>
<td>17.58s (4879)</td>
</tr>
<tr>
<td>FBF Acc</td>
<td>5.05s (1384)</td>
<td>10.27s (2801)</td>
</tr>
<tr>
<td>FB [122]</td>
<td>2.48s (1353)</td>
<td>5.72s (3090)</td>
</tr>
<tr>
<td>FB Acc [26]</td>
<td>2.04s (1102)</td>
<td>4.11s (2205)</td>
</tr>
<tr>
<td>PD [48]</td>
<td>1.48s (780)</td>
<td>3.26s (1708)</td>
</tr>
<tr>
<td>PD Acc [48]</td>
<td>1.28s (671)</td>
<td>3.14s (1649)</td>
</tr>
<tr>
<td>AMA [120]</td>
<td>13.53s (7209)</td>
<td>31.09s (16630)</td>
</tr>
<tr>
<td>AMA Acc [78]</td>
<td>3.10s (1639)</td>
<td>15.91s (8163)</td>
</tr>
<tr>
<td>Nesterov [100]</td>
<td>7.85s (3811)</td>
<td>42.69s (21805)</td>
</tr>
<tr>
<td>FISTA [15, 16]</td>
<td>7.55s (4055)</td>
<td>51.01s (27356)</td>
</tr>
</tbody>
</table>

We take the values $K = 10$ and $\phi = 0.5$ which are the best ones reported in [50] on a similar dataset.

Let $k$ be the number of nonzero weights $\omega_{ij}$. Then, introducing a linear operator $A : \mathbb{R}^{mn} \rightarrow \mathbb{R}^{kn}$, problem (4.27) can be equivalently written as

$$\inf_{x \in \mathbb{R}^{mn}} \{ f(x) + g(Ax) \},$$

the function $f$ being 1-strongly convex and differentiable with 1-Lipschitz continuous gradient. Also, by taking $p \in \{1, 2\}$, the proximal points with respect to $g^*$ admit explicit representations.

For our numerical tests, we consider the standard dataset consisting of two interlocking half moons in $\mathbb{R}^2$, each of them being composed of 100 points (see Figure 4.10). Our stopping criterion asks the root-mean-square error (RMSE) to be less than or equal to a given bound $\varepsilon$ which is either $\varepsilon = 10^{-4}$ or $\varepsilon = 10^{-8}$. As tuning parameters, we use $\gamma = 4$ for $p = 1$ and $\gamma = 5.2$ for $p = 2$ since both choices lead to a correct separation of the input data into the two half moons.

By taking into consideration the results given in Table 4.5, it shows that the two Douglas–Rachford type primal-dual methods are superior to all other algorithms within this comparison. One can also see that the accelerations of the forward-backward-forward (FBF) and of the forward-backward (FB) type primal-dual methods have a positive effect on both CPU times and required iterations compared with the regular methods. This characteristic is also achieved by the popular primal-dual (PD) method due to Chambolle and Pock and its acceleration (PD Acc), both methods...
given in [48]. The alternating minimization algorithm (AMA, cf. [120]) converges slow in this example while its Nesterov-type acceleration (AMA Acc, cf. [78]), however, performs better. The two accelerated proximal gradient methods FISTA (cf. [15, 16]) and the one we called Nesterov (cf. [100]), which are both solving the dual problem, perform surprisingly bad in this case.
1. We consider the convex optimization problem

\[
(P) \quad \inf_{x \in \mathcal{H}} \{ f(x) + g(Kx) \},
\]

where \( \mathcal{H} \) is a real Hilbert space, \( f \in \Gamma(\mathcal{H}) \), \( g \in \Gamma(\mathbb{R}^m) \), and \( K : \mathcal{H} \to \mathbb{R}^m \) is a linear operator fulfilling \( K(\text{dom } f) \cap \text{dom } g \neq \emptyset \). Furthermore, we assume that \( \text{dom } f \) and \( \text{dom } g \) are bounded and assign the (Fenchel) dual problem

\[
(D) \quad \sup_{p \in \mathbb{R}^m} \{-f^*(K^*p) - g^*(-p)\}
\]

to \( (P) \), where \( f^* : \mathcal{H} \to \overline{\mathbb{R}} \) and \( g^* : \mathbb{R}^m \to \overline{\mathbb{R}} \) denote the conjugate functions of \( f \) and \( g \), respectively. We then develop an algorithm for solving \( (P) \) by making use of an approach which regularizes the dual objective function two times into a differentiable strongly convex one with Lipschitz continuous gradient. A fast gradient method by Nesterov (cf. [98]) then solves the doubly regularized dual problem and allows the reconstruction of an approximately optimal primal solution.

2. We show that the double smoothing approach establishes a rate of convergence of \( \mathcal{O}\left(\frac{1}{\varepsilon} \ln \left(\frac{1}{\varepsilon}\right)\right) \) with respect to the primal objective function values where \( \varepsilon > 0 \) is the desired accuracy. By strengthening the assumptions on \( f \) and/or \( g \) in view of strong convexity and/or Fréchet differentiability, we are able to improve this rate of convergence to \( \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}} \ln \left(\frac{1}{\varepsilon}\right)\right) \) or even to \( \mathcal{O}\left(\ln \left(\frac{1}{\varepsilon}\right)\right) \).

3. We consider the convex optimization problem

\[
\inf_{x \in \mathcal{H}} \{ f(x) + g(Kx) \},
\]

where \( \mathcal{H} \) and \( \mathcal{G} \) are real Hilbert spaces, \( K : \mathcal{H} \to \mathcal{G} \) is a bounded linear operator, and \( f \in \Gamma(\mathcal{H}) \) as well as \( g \in \Gamma(\mathcal{G}) \) are functions such that \( f \) is \( L_f \)-Lipschitz continuous and \( g \) is \( L_g \)-Lipschitz continuous for some real constants \( L_f, L_g \in \mathbb{R}_{++} \). We develop a smoothing strategy for solving this optimization problem which firstly regularizes the functions \( f \) and \( g \) by approximating them via their Moreau envelopes. An accelerated gradient method by Nesterov (cf. [97]) is then applied to the obtained problem. Our approach enables a successive reduction of the smoothing parameters involved in this regularization from iteration to iteration. This ensures, in contrast to the double smoothing approach, where constant smoothing parameters are used, that the primal objective can be approximated more accurately by continuously differentiable functions as the iteration counter increases. Dependent on the choice of constant or variable smoothing parameters, we derive rates of convergence for the nonregularized objective function of \( \mathcal{O}\left(\frac{1}{k}\right) \) or \( \mathcal{O}\left(\frac{\ln k}{k}\right) \), respectively. Here, \( k \) denotes the number of iterations.
4. In the context of the variable smoothing algorithm, we also consider convex optimization problems in the shape of

$$\inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(K_i x) \right\},$$

where $\mathcal{H}$ and $\mathcal{G}_i$, $i = 1, \ldots, m$, are real Hilbert spaces, $f : \mathcal{H} \to \mathbb{R}$ and $g_i : \mathcal{G}_i \to \mathbb{R}$ are convex and Lipschitz continuous functions, and $K_i : \mathcal{H} \to \mathcal{G}_i$, is a bounded linear operator for each $i = 1, \ldots, m$. Additionally, we derive algorithms for the special case when the function $f$ is already known to be differentiable with Lipschitz continuous gradient. In this particular situation, one can remove the assumption of Lipschitz continuity imposed on $f$.

5. In terms of the forward-backward-forward method by Combettes and Pesquet in [58], we consider the convex optimization problem

$$(P) \quad \inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \square l_i)(L_i x - r_i) + h(x) - \langle x, z \rangle \right\},$$

and its dual

$$(D) \quad \sup_{(v_1, \ldots, v_m) \in \mathcal{G}_1 \times \cdots \times \mathcal{G}_m} \left\{ - (f^* \square h^*) \left( z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} (g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle) \right\},$$

where $\mathcal{H}$ and $\mathcal{G}_i$, $i = 1, \ldots, m$, are real Hilbert spaces. Here, we let $f, h \in \Gamma(\mathcal{H})$ such that $h$ is differentiable with $\mu$-Lipschitz continuous gradient, and, for each $i = 1, \ldots, m$, we let $g_i, l_i \in \Gamma(\mathcal{G}_i)$ such that $l_i$ is $\nu_i^{-1}$-strongly convex, while $L_i : \mathcal{H} \to \mathcal{G}_i$ is a bounded linear operator. By introducing the notion of the primal-dual gap restricted to some bounded set, we show that, under certain conditions, this gap can be bounded above by some term which decreases with an order of $O(\frac{1}{n})$, where $n$ is the iteration counter.

6. By still considering the primal-dual method due to Combettes and Pesquet in [58] in terms of solving a primal-dual system of monotone inclusions, we propose accelerations for this method under the additional assumption of strong monotonicity. The problem involves the primal inclusion

$$\text{find } \pi \in \mathcal{H} \text{ such that } z \in A\pi + \sum_{i=1}^{m} L_i^* ((B_i \square D_i)(L_i \pi - r_i)) + C\pi,$$

and the dual inclusion

$$\text{find } v_i \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \left\{ \begin{aligned} z - \sum_{i=1}^{m} L_i^* v_i & \in Ax + Cx, \\ v_i & \in (B_i \square D_i)(L_i x - r_i), \quad i = 1, \ldots, m, \end{aligned} \right\},$$

where for each $i = 1, \ldots, m$, the mapping $L_i : \mathcal{H} \to \mathcal{G}_i$ is bounded linear and the operators $A$ and $B_i$ are set-valued maximally monotone while $C$ and $D_i^{-1}$ are assumed to be single-valued monotone Lipschitzian. To this end, whenever $A + C$ is $\rho$-strongly monotone for some $\rho \in \mathbb{R}_{++}$, we can show that the sequence of iterates $(x_n)_{n \geq 0}$ in our new algorithm converges to an optimal primal solution such that the norm distance decreases with an order of $O(\frac{1}{n})$. In the situation when $A + C$ and $B_i^{-1} + D_i^{-1}$ are strongly monotone for each $i = 1, \ldots, m$, we obtain linear convergence for a sequence $(x_n, v_{1,n}, \ldots, v_{m,n})_{n \geq 0}$ to the unique primal-dual solution of our problem.
7. We develop two different Douglas–Rachford type methods for solving a general primal-dual system of monotone inclusions under minimal assumptions on the given operators. Therefore, by considering the real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}_i$, $i = 1, \ldots, m$, we solve the primal inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } z \in A x + \sum_{i=1}^{m} L_i^* (B_{\bigsquare} D_i) (L_i x - r_i),$$

together with the dual inclusion

$$\text{find } \nu_1 \in \mathcal{G}_1, \ldots, \nu_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \left\{ z - \sum_{i=1}^{m} L_i^* \nu_i \in A x, \left\{ \nu_i \in (B_{\bigsquare} D_i) (L_i x - r_i), \ i = 1, \ldots, m, \right. \right. \right.$$

where for each $i = 1, \ldots, m$, the set-valued operators $A$, $B_i$, and $D_i$ are only assumed to be maximally monotone. To prove their asymptotic convergence, we use a common product Hilbert space strategy by reformulating the corresponding inclusion problem reasonably such that an error tolerant Douglas–Rachford algorithm can be applied to it.

8. In terms of the two Douglas–Rachford type primal-dual methods for solving monotone inclusions, we also consider the important scenario when these are applied to convex minimization problems. Therefore, we aim for solving

$$\text{(P)} \quad \inf_{x \in \mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} (g_i \bigsquare l_i) (L_i x - r_i) - \langle x, z \rangle \right\},$$

as well as its associated dual problem

$$\text{(D)} \quad \sup_{(v_1, \ldots, v_m) \in \mathcal{G}_1 \times \cdots \times \mathcal{G}_m} \left\{ -f^* \left( z - \sum_{i=1}^{m} L_i^* v_i \right) - \sum_{i=1}^{m} (g_i^*(v_i) + l_i^*(v_i) + \langle v_i, r_i \rangle) \right\},$$

where, for each $i = 1, \ldots, m$, the functions $f$, $g_i$, and $l_i$ are only assumed to be proper, convex, and lower semicontinuous. Instead of being obliged to determine resolvents of maximally monotone operators, the methods now ask for the proximal point mappings of these functions which are sometimes known to take explicit expressions, as is the case in the majority of our experiments in Chapter 4. Conditions which ensure the equivalence between solving monotone inclusions and convex optimization problems of this type are discussed as well.

9. We develop two primal-dual methods for solving monotone inclusion problems having parallel sums of linearly composed monotone operators in their formulation, i.e., parallel sums of the type

$$\left( K_i^* \circ B_i \circ K_i \right) \bigsquare \left( M_i^* \circ D_i \circ M_i \right), \ i = 1, \ldots, m,$$

the operators $B_i$ and $D_i$ being maximally monotone, while $K_i$ and $M_i$ are assumed to be bounded and linear for each $i = 1, \ldots, m$. These methods exploit the approach of reformulating the inclusion problems in a product Hilbert space and rely, on the one hand, on an error tolerant forward-backward method and, on the other hand, on an error tolerant forward-backward-forward method which provide asymptotic convergence.
10. Their application to convex optimization problems in the shape of

$$\inf_{x \in H} \left\{ f(x) + \sum_{i=1}^{m} \left( (g_i \circ K_i) \square (l_i \circ M_i) \right) (L_i x - r_i) + h(x) - \langle x, z \rangle \right\},$$

is given a special attention. This is reasonable since parallel sums of linearly composed monotone operators are inspired by a real-world application in imaging (cf. [47, 117]), where first- and second-order total variation functionals are linked via infimal convolutions in order to reduce staircasing effects in the reconstructed images. We additionally provide conditions under which at least weak convergence of these methods to a primal-dual solution is guaranteed.

11. We investigate numerical experiments which arise in the fields of image processing, machine learning, location theory, portfolio optimization and clustering. As measure of performance, we use the root-mean-square error (RMSE) to a desired solution, or, in the case of solving image processing problems, we might use the improvement in signal-to-noise ratio (ISNR).
## Symbols and notation

### Spaces and basics

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{H}$</td>
<td>The Hilbert space $\mathcal{H}$</td>
<td>5</td>
</tr>
<tr>
<td>$\mathcal{H}_V$</td>
<td>The Hilbert space $\mathcal{H}$ with scalar product $\langle x, y \rangle_{\mathcal{H}_V} = \langle x, V y \rangle$</td>
<td>69</td>
</tr>
<tr>
<td>$\mathcal{H} \oplus G$</td>
<td>The Hilbert direct sum of two Hilbert spaces $\mathcal{H}$ and $G$</td>
<td>5</td>
</tr>
<tr>
<td>$\mathcal{H} \times G$</td>
<td>The Cartesian product of two Hilbert spaces $\mathcal{H}$ and $G$</td>
<td>5</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>The set of real numbers</td>
<td>5</td>
</tr>
<tr>
<td>$\mathbb{R}_+$</td>
<td>The set of positive real numbers</td>
<td>5</td>
</tr>
<tr>
<td>$\mathbb{R}_{++}$</td>
<td>The set of strictly positive real numbers, $\mathbb{R}<em>{++} = \mathbb{R}</em>+ \setminus {0}$</td>
<td>5</td>
</tr>
<tr>
<td>$\mathbb{N}$</td>
<td>The set of natural numbers, $\mathbb{N} = {1, 2, \ldots}$</td>
<td>5</td>
</tr>
<tr>
<td>$(\Omega, \mathfrak{F}, \mathbb{P})$</td>
<td>The probability space</td>
<td>11</td>
</tr>
<tr>
<td>$L^2(\Omega, \mathfrak{F}, \mathbb{P})$</td>
<td>The set of functions $X : \Omega \rightarrow \mathbb{R}$ such that $X$ is square integrable with respect to $\mathbb{P}$</td>
<td>11</td>
</tr>
<tr>
<td>$\langle \cdot, \cdot \rangle$</td>
<td>The scalar product</td>
<td>5</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\cdot</td>
</tr>
<tr>
<td>$\rightarrow$</td>
<td>Strong convergence</td>
<td>5</td>
</tr>
<tr>
<td>$\rightharpoonup$</td>
<td>Weak convergence</td>
<td>5</td>
</tr>
<tr>
<td>$\forall$</td>
<td>For all</td>
<td>5</td>
</tr>
<tr>
<td>$\in$</td>
<td>In</td>
<td>5</td>
</tr>
<tr>
<td>$\exists$</td>
<td>There exists (at least one)</td>
<td>5</td>
</tr>
<tr>
<td>$\pm \infty$</td>
<td>Plus and minus infinite, respectively</td>
<td>5</td>
</tr>
<tr>
<td>$(x_n)_{n \geq 0}$</td>
<td>A sequence of vectors starting with $x_0$</td>
<td>5</td>
</tr>
<tr>
<td>$x_n \approx y_n$</td>
<td>The sequences $(x_n)<em>{n \geq 0}$ and $(y_n)</em>{n \geq 0}$ fulfill $\sum_{n \geq 0}</td>
<td></td>
</tr>
<tr>
<td>$1^n$</td>
<td>The element in $\mathbb{R}^n$ with all entries equal to 1</td>
<td>5</td>
</tr>
</tbody>
</table>

### Functions

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Gamma(\mathcal{H})$</td>
<td>The set of proper, convex, and lower semicontinuous functions</td>
<td>7</td>
</tr>
<tr>
<td>$f : \mathcal{H} \rightarrow \mathbb{R}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{dom } f$</td>
<td>The domain of a function $f$</td>
<td>6</td>
</tr>
<tr>
<td>$\text{epi } f$</td>
<td>The epigraph of a function $f$</td>
<td>6</td>
</tr>
<tr>
<td>$\text{arg min } f$</td>
<td>The set of minimizers of a function $f$</td>
<td>5</td>
</tr>
<tr>
<td>Symbols and notation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>$f \Box g$</td>
<td>The infimal convolution of two functions $f$ and $g$. p. 8</td>
<td></td>
</tr>
<tr>
<td>$\gamma f$</td>
<td>The Moreau envelope of index $\gamma$ of a function $f$. p. 8</td>
<td></td>
</tr>
<tr>
<td>$\text{Prox}_f$</td>
<td>The proximity operator of a function $f$. p. 8</td>
<td></td>
</tr>
<tr>
<td>$f^*$</td>
<td>The conjugate of a function $f$. p. 7</td>
<td></td>
</tr>
<tr>
<td>$f^{**}$</td>
<td>The biconjugate of a function $f$. p. 7</td>
<td></td>
</tr>
<tr>
<td>$\partial f$</td>
<td>The subdifferential of a function $f$. p. 7</td>
<td></td>
</tr>
<tr>
<td>$\nabla f$</td>
<td>The gradient of a function $f$. p. 7</td>
<td></td>
</tr>
<tr>
<td>$d(x; C)$</td>
<td>The distance function of a point $x$ to a set $C$. p. 113</td>
<td></td>
</tr>
<tr>
<td>$\delta_C$</td>
<td>The indicator function of a set $C$. p. 9</td>
<td></td>
</tr>
<tr>
<td>$\sigma_C$</td>
<td>The support function of a set $C$. p. 9</td>
<td></td>
</tr>
<tr>
<td>$f_\infty$</td>
<td>The recession function of a function $f$. p. 117</td>
<td></td>
</tr>
<tr>
<td>$f \circ A$</td>
<td>The composition of a function $f$ with a linear operator $A$. p. 8</td>
<td></td>
</tr>
<tr>
<td>$TV(x)$</td>
<td>The total variation functional of an image $x \in \mathbb{R}^n$. p. 102</td>
<td></td>
</tr>
<tr>
<td>ISNR</td>
<td>The improvement in signal-to-noise ratio. p. 106</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>The root-mean-square error. p. 103</td>
<td></td>
</tr>
<tr>
<td>$\mathbb{E}[X]$</td>
<td>The expectation value of a random variable $X$. p. 11</td>
<td></td>
</tr>
<tr>
<td>$\text{VaR}_\alpha(X)$</td>
<td>The Value-at-Risk of a random variable $X$ for $\alpha \in (0, 1)$. p. 119</td>
<td></td>
</tr>
<tr>
<td>$\text{CVaR}_\alpha(X)$</td>
<td>The Conditional Value-at-Risk of a random variable $X$ for $\alpha \in (0, 1)$. p. 119</td>
<td></td>
</tr>
<tr>
<td>$\text{essinf}(X)$</td>
<td>The essential infimum of a random variable $X$. p. 11</td>
<td></td>
</tr>
<tr>
<td>$\text{esssup}(X)$</td>
<td>The essential supremum of a random variable $X$. p. 11</td>
<td></td>
</tr>
<tr>
<td>$\inf, \min$</td>
<td>The infimum and minimum, respectively. p. 5</td>
<td></td>
</tr>
<tr>
<td>$\lim\inf$</td>
<td>The limit inferior. p. 5</td>
<td></td>
</tr>
<tr>
<td>$\sup, \max$</td>
<td>The supremum and maximum, respectively. p. 5</td>
<td></td>
</tr>
<tr>
<td>$\lim\sup$</td>
<td>The limit superior. p. 5</td>
<td></td>
</tr>
<tr>
<td>$(P)$</td>
<td>The primal optimization problem. p. 5</td>
<td></td>
</tr>
<tr>
<td>$v(P)$</td>
<td>The optimal objective value of $(P)$. p. 5</td>
<td></td>
</tr>
<tr>
<td>$(D)$</td>
<td>The dual optimization problem. p. 5</td>
<td></td>
</tr>
<tr>
<td>$v(D)$</td>
<td>The optimal objective value of $(D)$. p. 5</td>
<td></td>
</tr>
</tbody>
</table>

### Sets

- $2^\mathcal{H}$: The power set of $\mathcal{H}$. p. 10
- $C \times D$: The Cartesian product of two sets $C$ and $D$. p. 5
- $C + D$: The Minkowski sum of two sets $C$ and $D$. p. 5
- $\lambda C$: The scaled set $C$ by a real constant $\lambda$. p. 5
- $z + C$: The translation of a set $C$ by a vector $z$. p. 5
- aff $C$: The affine hull of a set $C$. p. 5
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>cone $C$</td>
<td>The conical hull of a set $C$</td>
<td>5</td>
</tr>
<tr>
<td>int $C$</td>
<td>The interior of a set $C$</td>
<td>6</td>
</tr>
<tr>
<td>ri $C$</td>
<td>The relative interior of a set $C$</td>
<td>6</td>
</tr>
<tr>
<td>sqri $C$</td>
<td>The strong quasi-relative interior of a set $C$</td>
<td>6</td>
</tr>
<tr>
<td>cl $C$</td>
<td>The closure of a set $C$</td>
<td>6</td>
</tr>
<tr>
<td>$P_C$</td>
<td>The projection onto a nonempty closed convex set $C$</td>
<td>9</td>
</tr>
<tr>
<td>$C^\perp$</td>
<td>The orthogonal complement of a set $C$</td>
<td>5</td>
</tr>
<tr>
<td>$N_C$</td>
<td>The normal cone of a set $C$</td>
<td>5</td>
</tr>
<tr>
<td>$B(x,r)$</td>
<td>The closed ball with center $x$ and radius $r \in \mathbb{R}_+$</td>
<td>5</td>
</tr>
<tr>
<td>$[x,y]$</td>
<td>The closed interval between $x$ and $y$</td>
<td>5</td>
</tr>
<tr>
<td>$(x,y)$</td>
<td>The open interval between $x$ and $y$</td>
<td>5</td>
</tr>
</tbody>
</table>

### Set-valued operators

$A : \mathcal{H} \rightarrow 2^\mathcal{G}$ The set-valued operator $A$ mapping from $\mathcal{H}$ to $\mathcal{G}$................. p. 10

dom $A$ The domain of a set-valued operator $A$ .................................. p. 10

ran $A$ The range of a set-valued operator $A$ .................................... p. 10

gra $A$ The graph of a set-valued operator $A$ .................................... p. 10

zer $A$ The set of zeros of a set-valued operator $A$ .......................... p. 10

$A^{-1}$ The inverse of a set-valued operator $A$ ................................. p. 10

$A + B$ The sum of a set-valued operator $A$ with a set-valued operator $B$ p. 11

$A \square B$ The parallel sum of a set-valued operator $A$ with a set-valued operator $B$ p. 11

$\lambda A$ The scaling of a set-valued operator $A$ by $\lambda \in \mathbb{R}$ ................................. p. 10

$J_A$ The resolvent of a set-valued operator $A$ ................................. p. 10

$R_A$ The reflected resolvent of a set-valued operator $A$ ................................. p. 10

### Single-valued operators

$T : \mathcal{H} \rightarrow \mathcal{G}$ The single-valued operator $T$ mapping from $\mathcal{H}$ to $\mathcal{G}$ ............... p. 9

Fix $T$ The set of fixed points of a single-valued operator $T$ .................. p. 9

ker $L$ The kernel of a linear operator $L$ ......................................... p. 6

$\|L\|$ The norm of a linear operator $L$ ........................................... p. 6

$L^*$ The adjoint of a linear operator $L$ ........................................ p. 6

$A^T$ The transpose of a matrix $A$ ................................................. p. 6

$A \otimes B$ The Kronecker product of two matrices $A$ and $B$ ................ p. 105

Id The identity operator ........................................................................ p. 10
Bibliography


Index

K-nearest neighbors, 121
\(\beta\)-inverse strongly monotone operator, 10
\(\epsilon\)-accuracy, 19
adjoint, 6
affine hull, 6
almost surely, 11
arg min, 5
averaging operator, 108
biconjugate, 7
blur operator, 108
bounded linear operator, 6
capital requirement, 12
Cauchy–Schwarz inequality, 5
classical Heron problem, 113
classification, 110
closed
  ball, 5
  set, 6
  unit ball, 5
cluster center, 121
clustering, 121
cocoercive operator, 10
concave function, 6
  strictly concave function, 6
concave utility function, 12
Conditional Value-at-Risk, 119
cone, 6
conjugate function, 7
continuous linear operator, 6
convex function, 6
  strictly convex function, 6
  strongly convex function, 6
convex set, 6
convex utility function, 12
decision function, 111
difference operator
  horizontal, 105
  vertical, 105
discrete first-order derivative, 105
discrete second-order derivative, 105
distance function, 113
domain
  of a function, 6
  of an operator, 10
dual
  optimal objective value, 5
  problem, 5
dual unit ball, 102
entropic risk measure, 119
epigraph, 6
essential infimum, 11
essential supremum, 11
Euclidean space, 5
expectation value, 11
expected returns, 116
fast gradient method, 19, 21
Fenchel–Moreau Theorem, 7
Fermat’s rule, 8
firmly nonexpansive operator, 9, 10
fixed point, 9
fixed point set, 9
forward difference matrix, 105
future net worth, 12
Gaussian noise, 106
generalized Heron problem, 113
Gram matrix, 111
graph of a set-valued operator, 10
Haar wavelet transform, 108
Hilbert
  direct sum, 5
  space, 5
hinge loss, 111
identity operator, 10
image deblurring problem, 108
image denoising problem, 101, 104
implementation errors, 84
increasing function, 6
indicator function, 9
infimal convolution, 8
exact infimal convolution, 8
infimum, 5
inner product, 5, 12, 111
interior, 6
inverse of a set-valued operator, 10
ISNR, 106
kernel function, 111
kernel of a linear operator, 6
Kronecker product, 105
learning method, 111
linear operator, 6
Lipschitz continuous, 9
Lipschitz continuous operator, 9
lower semicontinuity
at a point, 7
of a function, 7
machine precision, 114
matrix, 6
maximum, 5
minimum, 5
Minkowski sum, 5
monotone operator, 10
maximally monotone, 10
strongly monotone, 10
uniformly monotone, 10
Moreau envelope, 8, 16, 34
Moreau’s decomposition formula, 9, 10
natural numbers, 5
Neumann boundary condition, 102
nonexpansive operator, 10
norm, 5, 12, 111
norm of a linear operator, 6
normal cone, 6
normalization condition, 12
optimal solution, 5, 19
Optimized Certainty Equivalent, 12
orthogonal complement, 6
parallel sum, 11
portfolio optimization problem, 115
primal
optimal objective value, 5
problem, 5
primal-dual gap function, 51
primal-dual solution, 48, 84
PRISMA algorithm, 33
probability measure, 11
probability space, 11
projection operator, 9
proper function, 6
proximal point, 9
proximal point mapping, 9
random variable, 11
range of a set-valued operator, 10
rate of convergence, 33, 36
real numbers, 5
extended real line, 5
positive real numbers, 5
strictly positive real numbers, 5
recession function, 117
reflected resolvent, 10
reflexive boundary condition, 102
regularization parameter, 101, 106, 108, 111
relative interior, 6
representer theorem, 111
Reproducing Kernel Hilbert Space, 111
resolvent, 10
right position, 114
risk function, 12
cash-invariant risk function, 12
coherent risk measure, 12
convex risk function, 12
convex risk measure, 12
monotone risk function, 12
positively homogeneous risk function, 12
RMSE, 103
scaled set, 6
self-adjoint operator, 68, 75
set of zeros, 10
set-valued operator, 10
single-valued operator, 9
skew symmetric operator, 67, 75
smoothed convex function, 16
smoothing constants, 18
smoothing parameters
  constant, 23, 29, 31, 39, 44
  variable, 42
smoothness functional, 111
Stolz–Cesàro Theorem, 55, 60
strictly increasing function, 6
strong convergence, 5
strong duality, 5
strong quasi-relative interior, 6
strongly positive operator, 68, 76
subdifferentiable, 8
subdifferential, 8
subgradient, 7
sublinearity, 12
subsequence, 27
sum of two set-valued operators, 11
support vector machines, 110
supremum, 5
test data set, 111, 113
Tikhonov regularization problem, 111
total variation
  anisotropic, 102
  involving 2nd-order derivative, 104
  isotropic, 102
training data set, 111, 112
transpose, 6
tuning parameter, 121
upper bound, 21, 29
upper semicontinuous function, 16
utility function
  exponential, 119
  indicator, 119
  linear, 118
  logarithmic, 120
  quadratic, 120
Value-at-Risk, 119
weak convergence, 5
weak duality, 5, 19
weakly compact set, 27
worst-case risk measure, 119
Young–Fenchel inequality, 7
Curriculum Vitae

Personal Information

Last name    Hendrich
First name   Christopher
Year of birth 1988
Citizenship German

Studies & Education

04/2012–07/2014 Technische Universität Chemnitz, Germany
  Degree program: PhD Program in Mathematics
  Thesis: Proximal Splitting Methods in Nonsmooth Convex Optimization (supervisor: Prof. Dr. Radu Ioan Boţ)
  Degree: Doctor rerum naturalium (Dr.rer.nat.)
  Rating: summa cum laude (scl)

10/2007–03/2012 Technische Universität Chemnitz, Germany
  Degree program: Business Mathematics
  Thesis: A Double Smoothing Technique for Solving Nondifferentiable Convex Optimization Problems
  Degree: Dipl.-Math.oec.
  Rating: passed with distinction
  Average grade: 1.1

  Abitur (university-entrance diploma)

Research Fellowships

04/2012–07/2014 Granted with a PhD fellowship from the Free State of Saxony, Germany

Published or Accepted Articles in Journals


Submitted Preprints


Conference Talks


Chemnitz, July 17, 2014