Dissertation on

The Augmented Lagrangian Method and Associated Evolution Equations

submitted to the Faculty of Mathematics, Computer Science and Physics of the University of Innsbruck in partial fullfilment of the requirements for the degree of *doctor rerum naturalium* by

DI Klaus Frick

Innsbruck, September 2008

Supervisor: Univ.-Prof. Dr. Otmar Scherzer
2^{nd} Referee: Univ.-Prof. Dr. Karl Kunisch

Danksagung

Diese Arbeit wurde im Rahmen des Nationalen Forschungsnetzwerks NFN S92 *Industrial Geometry* (Unterprojekte NFN S9203 und NFN S9207) des österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (FWF) verfaßt. Ich bedanke mich bei meinem Betreuer Otmar Scherzer für die Möglichkeit in diesem Projekt mitzuarbeiten, für die hervorragende Betreuung und vor allem für den großen Freiraum der mir für meine wissenschaftlichen Arbeit eingeräumt wurde.

Weiterer Dank gebührt den Kollegen und Kolleginnen in der Infmath Imaging Gruppe am Institut für Mathematik der Universität Innsbruck. Ohne die unzähligen Diskussionen während der Arbeit, während des gemeinsamen Kochens und Essens sowie beizeiten in gemütlicher Runde am Abend wäre an einen Abschluß dieser Arbeit nicht zu denken gewesen.

Besonderer Dank gilt meinen Eltern, die mich in meinem Vorhaben diese Dissertation zu schreiben zu jedem Zeitpunkt unterstützt haben und mir zur Seite gestanden sind. Ihnen sei daher diese Arbeit gewidmet.

Zusammenfassung

Viele Probleme der mathematischen Bildverarbeitung können mit Hilfe von der Multiskalenanalyse modelliert werden, allen voran Bildentrauschung und Bildkompression. Neben klassischen Multiskalenmethoden, wie etwa parabolischen Partiellen Differentialgleichungen oder Wavelets, haben in den letzten Jahren vor allem *inverse Skalenräume* zunehmend Verwendung gefunden. Die vorliegende Dissertation beinhaltet eine Abhandlung über inverse Skalenräume wobei zwei Kernthemen formuliert werden können:

Ein *diskreter inverser Skalenraum* bezeichnet eine Folge von (Bild)Rekonstruktionen, wobei Folgenglieder mit hohem Index mehr Details aufweisen (auf einen feineren Skala liegen) als solche mit niederem Index. Dies steht im Gegensatz zu herkömmlichen Skalenräumen, womit der Terminus *invers* gerechtfertigt wird. In dieser Arbeit wird gezeigt, dass Iterationsverfahren aus der Optimierungstheorie, so genannte *Augmented Lagrangian* Algorithmen, diskrete inverse Skalenräume definieren.

Weiters wird ein kontinuierliches Modell (*stetige inverse Skalenräume*) entwickelt, welches durch abstrakte Differentialgleichungen definiert ist. Es wird gezeigt, dass Lösungen dieser Gleichungen existieren und durch den diskreten Iterationsprozess approximiert werden können.

Abstract

Numerous problems arising in mathematical imaging can be modeled by means of multiscale analysis, above all image denoising and image compression. Aside to standard multiscale methods, such as parabolic partial differential equations or wavelets, *inverse scalce spaces* have gained much popularity in the recent years. This dissertation contains a treatise on inverse scale spaces with emphasis placed on two core aspects:

A *discrete inverse scale space* denotes a sequence of (image)reconstructions, where elements with a high index are considered to feature more details (to lie on a finer scale) than those with a low index. This stays in contrast to conventional scale spaces and hence justifies the term *inverse*. In the present work it will be shown that iteration processes used in optimization theory, so called *augmented Lagrangian* algorithms, define inverse scale spaces.

Moreover, a continuous model (*continuous inverse scale space*) will be introduced which is given by abstract differential equations. It is shown that solutions of these equations exist and can be approximated by the discrete iteration process.

Contents

1 Introduction

This thesis deals with an *iterative* method for solving *linear, ill-posed* operator equations on Banach spaces. The iteration process is defined in a variational framework and is of *Augmented Lagrangian* type. Moreover, the link to a class of *evolution equations* is established, generalizing the abstract Cauchy problem and its relation to the proximal point method.

In this thesis topics of various areas in applied mathematics are brought together, including mathematical image processing, regularization of ill-posed problems and optimization theory. In order to delimit the scope of this work, we shall first give an overview over these fields.

Variational Filtering *Image restoration* is concerned with retrieving a visually appealing image from a degraded (*noisy*) capture. These degradations may originate from blurs (e.g. due to abberations of optical devices), lack of image information (e.g. due to occlusion of a scene by an object) or other perturbations.

Probably the simplest model assumption for the image restoration problem states that a noisy observation f is (pointwise) composed by the true image u and a noisy signal n, that is,

$$f = u + n. \tag{1.1}$$

In this situation the image restoration problem is often referred to *image denoising*.

Figure 1.1: Left: true image u. Right: noisy image $f = u + n$ with 10% Gaussian noise n.

A widely used approach to tackle the image denoising problem is *variational filtering*. Here, images are assumed to be real valued functions defined on an open and bounded set $\Omega \subset \mathbb{R}^2$ which are assigned to an ambient function space, for instance $\mathrm{L}^2(\Omega)$. Furthermore, it is assumed that true images differ from noisy ones by additional smoothness properties, that is, true images belong to a proper subspace $U \subsetneq \mathrm{L}^2(\Omega)$.

Assume that $J : U \to \mathbb{R}$ is a given functional and $\alpha > 0$. For the noisy image $f \in \mathrm{L}^2(\Omega)$, the variational filtering technique consists in computing

$$u_\alpha = \operatorname*{argmin}_{u \in U} \frac{1}{2} \int_\Omega |u - f|^2 \, \mathrm{d}x + \alpha J(u). \tag{1.2}$$

Then, the element u_α (if it exists) is assumed to approximate the true image u. The functional J in (1.2) is chosen, such that strong oscillations in u are penalized and α governs the trade-off between smoothness of u (large α) and data fidelity (small α). The functional J is often of the form

$$J(u) = \int_\Omega g(|\nabla u|_2)\,\mathrm{d}x \tag{1.3}$$

where $g : [0,\infty) \to \mathbb{R}$ is a given scaling function. There exists a large amount of different choices for g in (1.3) some of which are collected in Table 1.1. In [108], Radmoser et al. formulated general conditions on g such that (1.2) is well defined, in the sense that $u_\alpha \in \mathrm{H}^1(\Omega)$ exists and is unique.

The function $g(s)$, $s \geq 0$ ($\gamma > 0$)	The space U	Notes
s^p $(1 < p < \infty)$	$\mathrm{W}^{1,p}(\Omega)$	
s	$\mathrm{BV}(\Omega)$	Rudin et al. [116].
$s + \frac{\gamma}{2}s^2$	$\mathrm{H}^1(\Omega)$	Ito & Kunisch [84].
$\log\left(1+s^2\right)$		Perona & Malik [105]
$\begin{cases} \frac{1}{2\gamma}s^2 & \text{if } s \leq \gamma \\ \frac{\gamma}{2}s^2 + \frac{1}{2}\left(\frac{1}{\gamma} - \gamma\right) & \text{if } s \geq \frac{1}{\gamma} \\ s - \frac{\gamma}{2} & \text{else.} \end{cases}$	$\mathrm{H}^1(\Omega)$	Geman & Yang [63] ($\gamma \leq 1$)

Table 1.1: Scaling functions $g : [0,\infty) \to [0,\infty)$ and corresponding subspaces $U \subset \mathrm{L}^2(\Omega)$ with $u_\alpha \in U$.

Figure 1.2 shows a denoising result for the Rudin – Osher – Fatemi (ROF) Model ($g(s) = s$) for the noisy image f in Figure 1.1.

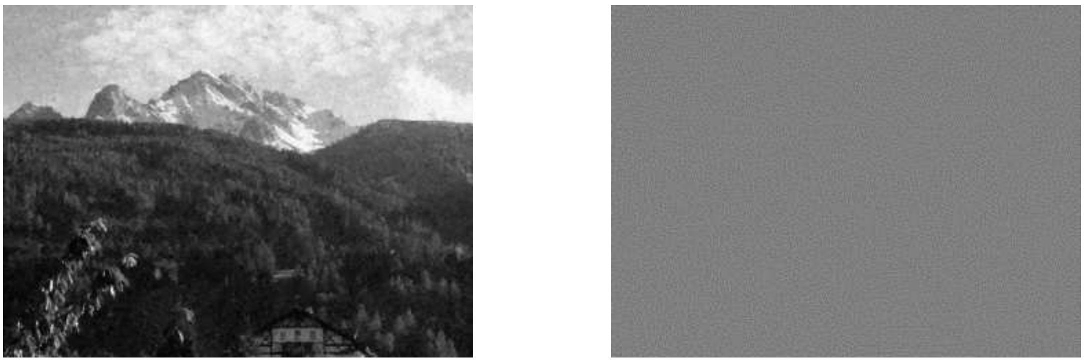

Figure 1.2: Left: Minimizer u_α of (1.2) for $g(s) = s$ and $\alpha = 0.01$. The noisy image f is as in Figure 1.1 with $0 \leq f \leq 1$. Right: Noise part $f - u_\alpha$.

We note, that all models listed in Table 1.1 merely depend on $|\nabla u|_2$, i.e. they are independent of the direction of the gradient. Such models are called *isotropic*. Typical *anisotropic* models are of the form

$$J(u) = \int_\Omega \boldsymbol{\nabla} u^T A \nabla u \,\mathrm{d}x$$

where $A = A(x) \in \mathbb{R}^{2\times 2}$ is a (possibly spacially variing) positive definite matrix. We refer to Weickert [123] for a treatise on anisotropic models of this type. Anisotropic models generalizing the (ROF) model have been studied by Esedoḡlu & Osher in [54]. For instance the case when

$$J(u) = \int_\Omega |\nabla u|_p \, \mathrm{d}x, \quad 1 \le p \le \infty$$

is covered by their analyis.

More recently, many authors focused rather on modifications of the data-fidelity term (i.e. the squared L^2-distance in (1.2)) than on coming up with new functions g. Aside to obvious generalizations (as e.g. the L^1-BV model of Chan & Esedoḡlu in [39]), the inspiring work of Meyer in [95] should be mentioned.

There, the noise (and texture) part of an image is modelled as an *oscillating function* with zero mean. In order to measure such functions, Meyer introduced the *g-norm* that is less sensitive to strong oscillations than for instance the squared L^2- distance and thus better suited for modelling noise. We refer to Chapter 4 of this thesis for a detailed analysis (and slight generalization) of Meyer's g-norm.

Scale Space Methods Revisiting the minimization problem (1.2) where we assume that J is given by (1.3), we shall assume that for a given $\alpha > 0$ a minimizer u_α exists. Formal computation of the optimality condition for (1.2) gives

$$0 = u_\alpha - f - \alpha \mathrm{div}\left(g'(|\nabla u_\alpha|) \frac{\nabla u_\alpha}{|\nabla u_\alpha|} \right) =: u_\alpha - f + \alpha \partial J(u_\alpha)$$

where we assumed that the normal trace of ∇u_α vanishes on the boundary $\partial\Omega$. Consequently this results in

$$\frac{u_\alpha - f}{\alpha} = -\partial J(u_\alpha). \tag{1.4}$$

Thus the minimization problem (1.2) can be considered an implicit time step for the evolution equation (for the solution at time $t = \alpha$)

$$\begin{aligned} \frac{du}{dt} &= -\partial J(u), && (1.5\text{a}) \\ \nabla u \cdot \nu &= 0 \text{ on } \partial\Omega, && (1.5\text{b}) \\ u(0) &= f. && (1.5\text{c}) \end{aligned}$$

Here ν denotes the outer unit normal vector at $\partial\Omega$. The operator ∂J can be interpreted as the gradient of the functional J as defined in (1.3). Thus (1.5) constitutes a steepest descent (or gradient flow) equation for J, that is, the solutions $\{u(t)\}_{t\ge 0}$ tend to minimize J as $t \to \infty$. Stopping the evolution at a suitable (and finite) time $t_0 > 0$ is believed to yield an approximation of the true image u.

Performing n implicit time steps for (1.5) results in *iterative minimization* of (1.2). That is, for given $t > 0$ we compute for $k = 1, \ldots n$

$$\begin{aligned} u_{0,n}(t) &= f, && (1.6\text{a}) \\ u_{k,n}(t) &= \operatorname*{argmin}_{u \in U} \frac{1}{2} \int_\Omega |u - u_{k-1,n}|^2 \, \mathrm{d}x + \frac{t}{n} \int_\Omega g(|\nabla u|) \, \mathrm{d}x. && (1.6\text{b}) \end{aligned}$$

This iteration method is also well known in optimization theory; in this context it is referred to as the *proximal point method*. The function $u_{n,n}(t)$ is considered as an approximation of the solution of (1.5) at time $t > 0$. Indeed, for suitably well behaved functions g it follows on the one hand that Equation (1.5) admits a unique solution $u(t)$ (see e.g. Brézis [26]) and on the other hand that the implicit scheme converges, that is

$$\lim_{n\to\infty} u_{n,n}(t) = u(t).$$

This is the main assertion in the celebrated work [44] of Crandall & Liggett.

In Table 1.2 some common scaling functions g and the corresponding differential operators ∂J are listed. We note that in the Perona – Malik case, the assumptions in [44] are not satisfied. For an analysis of the Perona – Malik model and the associated evolution equation we refer to Catté et al. [36] as well as Scherzer & Weickert [117].

The function $g(s)$, $s \in \mathbb{R}$	The operator $-\partial J(u)$	Notes
s^p $(1 < p < \infty)$	$\mathrm{div}\left(\lvert\nabla u\rvert^{p-2}\,\nabla u\right)$	p-Laplacian equation.
s	$\mathrm{div}\left(\frac{\nabla u}{\lvert\nabla u\rvert}\right)$	Total variation flow [10, 11].
$\log\left(1+s^2\right)$	$\mathrm{div}\left(\frac{\nabla u}{1+\lvert\nabla u\rvert^2}\right)$	Perona – Malik equation [105].

Table 1.2: Scaling functions $g : \mathbb{R} \to \mathbb{R}$ and corresponding operators in ∂J.

With Equation (1.5) an operator $\mathcal{R}_t : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ is associated, that maps a initial (noisy) image to the solution of (1.5) at time $t > 0$. The family $\{\mathcal{R}_t\}_{t\geq 0}$ constitutes a *continuous semigroup* of nonlinear operators. In particular it follows for each $t, s \geq 0$

$$\lim_{t\to 0^+} \mathcal{R}_t(f) = f \quad \text{and} \quad \mathcal{R}_s(\mathcal{R}_t(f)) = \mathcal{R}_{s+t}(f).$$

According to the definitions in the (latest version of) the book of Morel et al. [71, Chap. 21], an image scale space satisfying this property is referred to as *recursive*. In other words, this means that at the beginning of the evolution, all information is contained in the scale space, however, at each positive time $t > 0$ the solution $\mathcal{R}_t(f)$ can be computed from $\mathcal{R}_{t-\delta t}(f)$ for arbitrary small $\delta t > 0$ and thus information is lost during the evolution.

In [103], Osher et al. introduced an iterative method for image denoising which invokes the noisy data f in each iteration step. After initializing $w_0 = 0$ the Algorithm computes for $n = 0, 1, 2, \ldots$

$$u_{n+1} = \operatorname*{argmin}_{u\in U} \frac{1}{2}\int_\Omega |f + w_n - u|^2 \,\mathrm{d}x + \alpha J(u), \tag{1.7a}$$

$$w_{n+1} = w_n + f - u_{n+1}. \tag{1.7b}$$

That is, in the n-th step the new iterate u_{n+1} is computed by solving (1.2) with the signal $f + w_n$, the original image f *enhanced* by the accumulated error w_n.

Figure 1.3 displays the first two steps in the iteration. As it becomes (visually) obvious, the minimiziers u_k in (1.7) contain more details the longer the iteration lasts. Indeed, as it

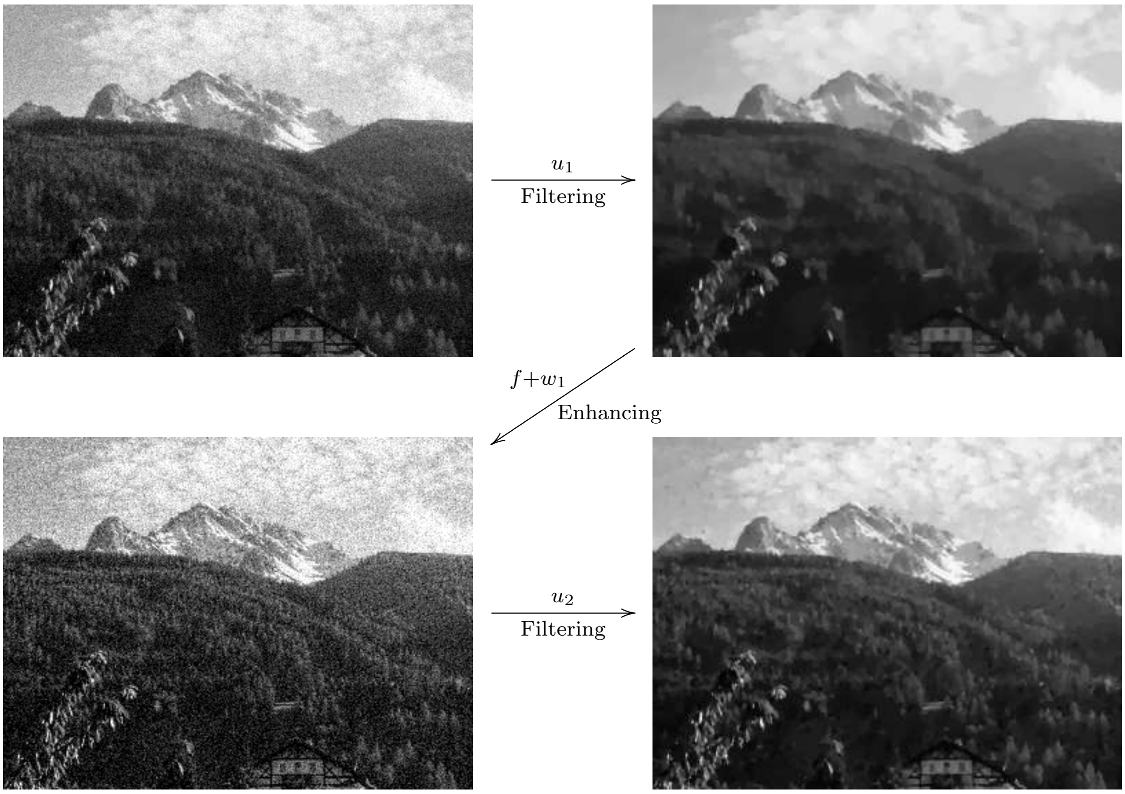

Figure 1.3: The first two steps of the iteration process (1.7) (with J and f as in Figure 1.2). From the original (noisy) data f (upper left image) a minimizer u_1 of (1.2) is computed (upper right image). Subsequently the image f is enhanced by $w_1 = f - u_1$ (lower left image) and a minimizer u_2 of (1.2) with data $f + w_1$ is computed (lower right image).

will turn out in the course of this thesis (and as it was shown by Burger et al. in [30]), one has

$$\lim_{n\to\infty} u_n = f.$$

By (formal) computation of the optimality condition of (1.7a) one finds for each $n \in \mathbb{N}$ that $v_n := \alpha^{-1} w_n = \partial J(u_n)$ and thus it follows from (1.7b) that

$$\frac{\partial J(u_{n+1}) - \partial J(u_n)}{\alpha^{-1}} = u_{n+1} - f.$$

Thus, when interpreting α^{-1} as time step size, (1.7) constitutes an implicit time scheme for the equation

$$\begin{aligned} \frac{d}{dt}\partial J(u) &= f - u, && (1.8a) \\ \nabla u \cdot \nu &= 0 \text{ on } \partial\Omega, && (1.8b) \\ u(0) &= u_0, && (1.8c) \end{aligned}$$

for an initial value u_0.

In the course of this thesis we will show that (1.8) admits a solution that can be approximated by the iteration in (1.7). If $\mathcal{R}_t(f)$ denotes a solution of (1.8) at time $t > 0$, we will show that

$$\lim_{t\to\infty} \mathcal{R}_t(f) = f.$$

This is the opposite behaviour of the scale space generated by (1.5). Thus $\{\mathcal{R}_t\}_{t\geq 0}$ is referred to as *inverse scale space*. Existence of solutions of (1.8) as well their multiscale properties were studied by Burger et al. in [30] and F. & Scherzer in [61]. It is important to note that the concept of recursivity is abandoned in the inverse scale space approach. That is, the solution $\mathcal{R}_t(f)$ can not be computed from the knowledge of $\mathcal{R}_{t-\delta t}(f)$ alone (for all $0 < \delta t < t$). Instead, no information is lost during the evolution, since $\mathcal{R}_t(f)$ is computed directly from the data f for all $t > 0$. As we will see in the upcoming paragraph, this property will allow us to generalize the inverse scale space methodology to a solution technique for ill posed problems.

We finally note, that Algorithm (1.7) can be rewritten in terms of the *Bregman distance* of u and u_n w.r.t. J and v_n, that is,

$$D_J^{v_n}(u, u_n) := J(u) - J(u_n) - \int_\Omega v_n(u - u_n)\,\mathrm{d}x.$$

Then, the minimization in (1.7a) is equivalent to

$$u_{n+1} = \operatorname*{argmin}_{u\in U} \frac{1}{2}\int_\Omega |f-u|^2\ \mathrm{d}x + \alpha D_J^{v_n}(u, u_n).$$

In other words, (1.7) can be considered as a *(generalized) proximal point method* where the distance of the iterates is measured by the Bregman distance rather than by the squared L^2-distance. Equation (1.8) then plays the role of the *associated evolution equation.*

Ill-posed Problems The analysis of this work shall not be restricted to the image denoising problem. Instead, we will formulate our results in the framework of *linear and ill-posed* operator equations. To be more precise, we assume that X, Y are Banach spaces and that $K : X \to Y$ is a linear and bounded operator. For a given $y \in Y$ we then consider the problem:

$$\text{Find } x \in X \text{ such that } Kx = y. \tag{1.9}$$

According to Hadamard [74] problem (1.9) is *well-posed* if the following conditions are satisfied

1. For each $y \in Y$ there exists a solution $x \in X$.

2. The solution x is unique.

3. The unique solution x depends continuously on the right hand side y w.r.t. a reasonable topology.

If problem (1.9) lacks one of these properties it is called *ill-posed.*

It is well known (cf. Engl et al. [53]) that for compact linear operators K, solutions x of (1.9) do not depend continuously on the right hand side y. That is, if the exact right hand side y is approximately given by (noisy) measurement data, there is no hope to approximate true solutions of (1.9) by evaluating K^{-1} at the measurment data (if K^{-1} exists).

One example of a ill posed problem is image restoration. The abberation caused by the optical system is modelled as a linear convolution operator K defined by given kernel function k (the *point spread function* of the optical system). The generalization of model (1.1) thus can be written as

$$f = k * u + n = Ku + n. \tag{1.10}$$

Convolution operators $K : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ with smoothing kernels are usually compact operators and thus retrieving the true image u from the measurment f in (1.10) in general is an ill-posed problem.

Many problems in applied mathematics and technology that can be written as (linear) operator equation turn out to be ill-posed. For further example of ill-posed operator equations, we refer to [53, Chap. 1].

In order to clarify ideas we will stay with the image restoration problem for the time being. Motivated by the variational method (1.2) we compute for $\alpha > 0$ a minimizer (provided it exists)

$$u_\alpha = \operatorname*{argmin}_{u \in U} \frac{1}{2} \int_\Omega |Ku - f|^2 \,\mathrm{d}x + \alpha J(u). \tag{1.11}$$

For the choice

$$J(u) = \frac{1}{2} \int_\Omega |u|^2 \,\mathrm{d}x \tag{1.12}$$

the resulting method in (1.11) is called *Tikhonov – Phillips regularization* (Tikhonov [119, 120] and Phillips [106]) and is probably one of the most common methods to compute *regularized solutions* of ill-posed operator equations. Regularized solutions in this context means that u_α approaches the true image u as $\|n\| \to 0$ and $\alpha = \alpha(\|n\|) \to 0$.

Straightforeward generalization of the k-th step in the proximal point method (1.6) (and thus also of Equation (1.5)) to the present situation amounts to replacing f in (1.11) by $Ku_{k-1,n}$. From an inverse problem point of view this is not very reasonable, since f is the only source which holds information of the problem (f is the measurment data).

We focus on the iteration (1.7) instead. By taking into account the equivalent formulation with the Bregman distance we find the following algorithm: Set $v_0 = 0$ and compute

$$\begin{aligned} u_{n+1} &= \operatorname*{argmin}_{u \in U} \frac{1}{2} \int_\Omega |Ku - f|^2 \,\mathrm{d}x + \alpha D_J^{K^* v_n}(u, u_n), && \text{(1.13a)} \\ v_{n+1} &= v_n + \alpha^{-1}(f - Ku_{n+1}). && \text{(1.13b)} \end{aligned}$$

As we will see in Chapter 1, this procedure is well known in the optimization community as *the augmented Lagrangian method.* It was introduced simultaneously by Hestenes [77] and Powell [107] and is designed to compute solutions of the constrained optimization problem

$$J(u) \to \inf! \quad \text{subject to} \quad Ku = f. \tag{1.14}$$

As in the image denoising case, we can (formally) argue that (1.13) forms an implicit time scheme for the evolution equation

$$\begin{aligned} \frac{d}{dt} \partial J(u) &= K^*(f - Ku), && \text{(1.15a)} \\ \nabla u \cdot \nu &= 0 \ \text{ on } \partial\Omega, && \text{(1.15b)} \\ u(0) &= u_0, && \text{(1.15c)} \end{aligned}$$

We note, that (1.15) (and thus by setting $K = \mathrm{Id}$ also (1.8)) can be considered as a gradient flow equation for the functional $\frac{1}{2}\|Ku - f\|^2$, where the time is scaled by the operator ∂J. For example, for J as in (1.12) one has that $\partial J = \mathrm{Id}$. Equation (1.15) is then called *Showalter's method* (or asymptotic regularization) (cf. [53, Ex. 4.7]).

For the general case, we recall that the considerations above are purely formal and that for nondifferentiable functionals J (such as the BV-seminorm, i.e. $g(s) = s$) the operator ∂J is *set-valued* (∂J is the *subgradient* of J). Thus Equation (1.15) has to be rewritten properly in order to cover nonsmooth functionals. To our knowledge, evolution equations as in (1.15) and their relation to the augmented Lagrangian method (1.13) have not been studied so far.

Aside to this pure theoretical motivation, we also note that the Augmented Lagrangian Algorithm 1.13 belongs to the class of *first order methods* for solving (1.14) which are known to converge rather slowly. Proving existence (and uniqueness) results for (1.15) paves the way for other approximation techniques than the first order implicit method discussed above. Moreover, we will be able to compute (at least for the image denoising case) *exact* solutions of (1.15) for a certain class of data f. This reveals the interesting behaviour of the solution trajectories of (1.15) unspoiled by errors due to numerical approximation. Computing of exact solutions for the minimization problem in Algorithm 1.13, however, in general is not as straightforward.

Main Goals and Organization of this Work In Chapter 2 the *augmented Lagrangian method* is formulated in a general Banach space setting with a *nonsmooth* regularization functional J. Moreover, it is pointed out that the dual sequence generated by the augmented Lagrangian method can be characterized by a proximal point algorithm.

With these preparations the main issue of the chapter will be addressed: regularizing properties of the augmented Lagrangian method for linear and ill-posed problems as in (1.9). We prove convergence in the general Banach space case and establish stronger results (convergence to *J-minimizing solutions*) including convergence rates (w.r.t. the Bregman distance) for equations with data in a Hilbert space. We compare our results to classical quadratic models as the iterative Tikhonov method and the Tikhonov – Morozov method.

Chapter 3 is dedicated to evolution equations and their relation to the augmented Lagrangian method viewed as implicit time scheme. We generalize (1.15) for nonsmooth operators J and prove existence of solutions. We explicitly construct solutions by considering interpolations of sequences generated by the augmented Lagrangian method.

We will make extensive use of the dual formulation derived in Chapter 2 combined with the analysis on gradient flows in the recent work by Ambrosio et al. in [9]. With these results at hand we prove convergence of discrete solutions. Moreover, special situations are studied in which stronger convergence results as well as better smoothness properties for solutions can be shown.

Finally, we will consider the solutions of the derived evolution equation as continuous regularization methods for the operator equation (1.9). As in Chapter 2 we derive stronger results in the Hilbert space case. In particular, an estimate for the approximation error of the augmented Lagrangian method as implicit time scheme is given.

The image denoising technique (1.7) is revisited in Chapter 4. We prove that the basic assumptions in Chapter 2 and Chapter 3 are satisfied and reformulate the general results developed therein. In particular, we focus on the correponding evolution equation, referred to as *inverse total variation flow* equation. Additional to the general results in Chapter 2 and Chapter 3, we will prove a maximum priniciple for the inverse total variation flow and will give a characterization of exact solutions.

We finally remark that each chapter is closed by a section called *Notes* that subsumes the main results of the preceeding analysis and gives references for further reading. Furthermore, mathematical preliminaries as well as some technical results are collected in Appendix A.

2 Iterative Regularization of Linear and Ill-posed Problems

In this chapter we study the linear and ill-posed problem

$$\text{For } y \in Y \text{ find } x \in X \text{ such that } Kx = y.$$

Here we assume that X, Y are Banach spaces and $K : X \to Y$ is linear and bounded. We recall that such a problem is called *well-posed* if each given $y \in Y$ there exists a unique solution x that depends continuously on the right hand side y. Otherwise the problem is *ill-posed*.

In case of ill-posedness, arbitrary small deviations in the right hand side y may yield to useless solutions x. *Regularization methods* are one approach in order to compute stable approximations of true solutions x from (possibly) noisy data y. In this chapter a particular regularization method is studied, which is an iterative algorithm and referred to as *augmented Lagrangian method.*

This chapter is organized as follows: After clarifying the basic assumptions and notation in Section 2.1 we will recall the definition of a regularization method for linear and ill-posed problems in Banach spaces and formulate the augmented Lagrangian method in Section 2.2. Moreover, by using the Legendre – Fenchel duality concept an alternative characterization of the algorithm is established.

With these preparations we proceed to a (preliminary) convergence result in Section 2.3, stating that a suitable parameter choice renders the augmented Lagrangian method a regularization method. We show that such a parameter choice can be realized by the *discrepancy principle*.

When the data y is assumed to be an element of a Hilbert space remarkably stronger convergence results (including convergence rates) are shown in Section 2.4. We compare our results to well known results in *quadratic regularization*, in particular to iterative Tikhonov and Tikhonov – Morozov regularization (Section 2.5). Section 2.6 concludes the chapter with a short summary (including a historical outline of the presented methods) as well as references for further reading.

2.1 Assumptions and Notation

We agree upon $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ being real Banach spaces and denote by X^* and Y^* there duals, equipped with the dual norms $\|\cdot\|_{X^*}$ and $\|\cdot\|_{Y^*}$ respectively. That is, the space $(X^*, \|\cdot\|_{X^*})$ consists of all bounded and linear functionals $x^* : X \to \mathbb{R}$ and

$$\|x^*\|_{X^*} = \sup_{x \in X \setminus \{0\}} \frac{x^*(x)}{\|x\|_X}.$$

Moreover, we will use the symbol $\langle \cdot, \cdot \rangle_{X^*,X}$ for the pairing of X^* with X, i.e.

$$\langle x^*, x \rangle_{X^*,X} := x^*(x)$$

for $x^* \in X^*$ and $x \in X$. Analogously we define $\|\cdot\|_{Y^*}$ and $\langle \cdot, \cdot \rangle_{Y^*,Y}$. We will skip the subscripts of norms and pairings as long as the situation is clear.

For $Z \in \{X, Y\}$ we introduce the symbols τ_Z^n and τ_Z^w for the strong and weak topology on Z respectively and we write $\tau_{Z^*}^{w^*}$ for the weak* topology on Z^*. If a sequence $\{z_n\}_{n\in\mathbb{N}} \subset Z$ converges weakly to some $z \in Z$ we use the notation

$$\underset{n\to\infty}{w\text{-lim}}\, z_n = z, \quad \text{or} \quad z_n \rightharpoonup z$$

and for the analogous situation for a weakly* converging sequence $\{z_n^*\}_{n\in\mathbb{N}} \subset Z^*$ with limit $z^* \in Z^*$

$$\underset{n\to\infty}{w^*\text{-lim}}\, z_n^* = z^* \quad \text{or} \quad z_n^* \rightharpoonup^* z^*.$$

For a sequence $\{z_n\}_{n\in\mathbb{N}} \subset Z$ we denote subsequences by $\left\{z_{\rho(n)}\right\}_{n\in\mathbb{N}}$, where $\rho : \mathbb{N} \to \mathbb{N}$ is strictly increasing. We call such a mapping ρ *selection*.

We further denote with $\mathcal{L}(X, Y)$ the collection of all linear and bounded operators mapping X into Y and we will use the symbol Kx instead of $K(x)$ for the evaluation of $K \in \mathcal{L}(X, Y)$ at $x \in X$ unless we do not run risk of getting confused. We define the *range* and *kernel* of $K \in \mathcal{L}(X, Y)$ as the sets

$$\begin{aligned} \operatorname{ran}(K) &= \{y \in Y \ : \ \text{there exists a } x \in X \text{ with } Kx = y\} \\ \ker(K) &= \{x \in X \ : \ Kx = 0\}. \end{aligned}$$

Further we denote by K^* the adjoint operator, that is, $K^* \in \mathcal{L}(Y^*, X^*)$ and for $p^* \in Y^*$ one has

$$\langle K^* p^*, x \rangle = \langle p^*, Kx \rangle.$$

for all $x \in X$ and $p^* \in Y^*$.

For a functional $J : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ we define the *domain of* J to be the set

$$D(J) := \{x \in X \ : \ J(x) < \infty\}.$$

and we refer to the Appendix A.2.1 for the definition of the subdifferential ∂J and its domain $D(\partial J)$. For elements $x_1 \in D(J)$ and $x_2 \in D(\partial J)$ such that $\xi^* \in \partial J(x_2)$ we recall the definition of the Bregman distance of x_1 and x_2 w.r.t. J and ξ^* (cf. Definition A.2.6)

$$D_J^{\xi^*}(x_1, x_2) = J(x_1) - J(x_2) - \langle \xi^*, x_1 - x_2 \rangle.$$

Furthermore we assume that $\phi : [0, \infty) \to [0, \infty)$ is a weight function (cf. Definition A.1.1). That is, ϕ is continuous, increasing and one has

$$\phi(0) = 0 \quad \text{and} \quad \lim_{s\to\infty} \phi(s) = \infty.$$

and we denote by ψ_ϕ its primitive, i.e.

$$\psi_\phi(s) = \int_0^s \phi(\sigma)\, \mathrm{d}\sigma.$$

We note that ψ_ϕ is increasing, strictly convex and continuously differentiable (with derivative ϕ). For every weight function ϕ, its inverse ϕ^{-1} is a weight function as well and the corresponding primitives are the duals of each other (cf. Example A.2.13), that is

$$\psi_{\phi^{-1}}(s) = \sup_{t\geq 0} \left(st - \psi_\phi(t)\right).$$

for each $s \geq 0$.

Given a weight function ϕ, a function $J : X \to \overline{\mathbb{R}}$ and a operator $K \in \mathcal{L}(X, Y)$, we will assume a list of requirements that establishes the mutual relation between these objects:

Assumption 2.1.1. R1. **(Topology)** There exist topologies τ_X on X and τ_Y on Y that are weaker than the norm topologies and stronger than the weak topologies, that is, for $Z \in \{X, Y\}$ we have that

$$\tau_Z^w \subset \tau_Z \subset \tau_Z^n.$$

Moreover, the norm $\|\cdot\|_Y$ is sequentially τ_Y-lower semicontinuous, i.e. for every τ_Y-convergent sequence $\{y_n\}_{n\in\mathbb{N}}$ with limit $y \in Y$ we have that

$$\|y\| \leq \liminf_{n\to\infty} \|y_n\| .$$

R2. **(Continuity)** The operator K is continuous w.r.t. τ_X and τ_Y, that is for all τ_X-convergent sequences $\{x_n\}_{n\in\mathbb{N}}$, the sequence $\{Kx_n\}_{n\in\mathbb{N}}$ is converging w.r.t. τ_Y.

R3. **(Lower Semicontinuity)** The functional J is convex, sequentially τ_X-lower semicontinuous and proper.

R4. **(Attainability)** The set

$$\{y \in Y \ : \ \text{there exists } x \in D(J) \text{ with } Kx = y\} \subset \operatorname{ran}(K).$$

contains at least one element, which we refer to as *attainable with respect to K and J* or simply *attainable* if the situation is clear.

R5. **(Compactness)** For each $y \in Y$ as well as for all $\alpha > 0$ and $c \in \mathbb{R}$ the sets

$$\Lambda(c) = \{x \in X \ : \ \psi_\phi (\|Kx - y\|) + \alpha J(x) \leq c\}$$

are sequentially τ_X-precompact, i.e. for every sequence $\{x_n\}_{n\in\mathbb{N}} \subset \Lambda(c)$ a selection $n \to \rho(n)$ can be chosen, such that $\left\{x_{\rho(n)}\right\}_{n\in\mathbb{N}}$ τ_X-converges to some $x \in X$.

R6. **(Range Condition)** There exists $p_0^* \in Y^*$ and $x_0 \in D(\partial J) \subset X$ such that $K^* p_0^* \in \partial J(x_0)$.

Remark 2.1.2. 1. If J is convex, lower semicontinuous and proper, then (R1) - (R3) hold, when τ_Z is chosen to be τ_Z^w for $Z = X, Y$.

2. Assume that Assumption 2.1.1 holds and that $\{x_n\}_{n\in\mathbb{N}} \subset \Lambda(c)$ for a $c \in \mathbb{R}$. Then each τ_X-cluster point of $\{x_n\}_{n\in\mathbb{N}}$ is already an element of $\Lambda(c)$, since due to (R1) - (R3) and due to the continuity of ψ_ϕ the mapping

$$x \mapsto \psi_\phi (\|Kx - y\|) + \alpha J(x)$$

is sequentially τ_X-lower semicontinuous and thus

$$\psi_\phi (\|K\hat{x} - y\|) + \alpha J(\hat{x}) \leq \liminf_{n\to\infty} \psi_\phi (\|Kx_n - y\|) + \alpha J(x_n) \leq c$$

for each τ_X-cluster point $\hat{x}$ of $\{x_n\}_{n\in\mathbb{N}}$. In other words, $\Lambda(c)$ is sequentially τ_X-compact.

If not stated differently we will throughout this chapter adopt the present notation, that is, we will always assume that a weight function ϕ, a functional J and a linear operator $K \in \mathcal{L}(X, X)$ are chosen such that Assumption 2.1.1 holds for suitable topologies τ_X and τ_Y on X and Y respectively. Moreover, we shall presume that $p_0^* \in Y^*$ and $x_0 \in X$ are chosen as in (R6), that is,

$$K^* p_0^* \in \partial J(x_0). \tag{2.1}$$

Finally, we fix a sequence of positive parameters $\boldsymbol{\alpha} := \{\alpha_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ and define for each $n \in \mathbb{N}$ the partial sum

$$t_n(\boldsymbol{\alpha}) := \sum_{j=1}^{n} \frac{1}{\alpha_j}. \tag{2.2}$$

We will assume that

$$\lim_{n\to\infty} t_n(\boldsymbol{\alpha}) = \infty. \tag{2.3}$$

2.2 The Augmented Lagrangian Method

In this section we will study the linear (and in general ill-posed) operation equation

$$Kx = y. \tag{2.4}$$

In particular we are interested in finding **regularizing operators**, that stably approximate **J-minimizing solutions** of (2.4). Before we do so, we clarify the notions in bold letters.

Definition 2.2.1. Let $y \in Y$ be attainable.

1. We call $x \in X$ a *J-minimizing solution* of (2.4), if x is a solution of the constrained minimization problem

$$J(x) \to \inf! \quad \text{subject to} \quad Kx = y. \tag{2.5}$$

2. A J-minimizing solution $x \in X$ satisfies the *source condition*, if $x \in D(\partial J)$ and there exists an element $p^* \in Y^*$ such that

$$K^* p^* \in \partial J(x). \tag{2.6}$$

   In that case, the element p^* is called a *source element*.

Remark 2.2.2. The notions *J-minimizing solution* and *source condition* as in Definition 2.2.1 have been introduced by Burger & Osher in [32]. They are motivated from Tikhonov regularization, that is, when X and Y are Hilbert spaces, $\phi(s) = s$ and $J = \frac{1}{2}\|\cdot\|^2$. In this case (2.6) reads as

$$x \in \operatorname{ran}(K^*)$$

and J-minimizing solutions are referred to as *minimal norm solutions*. It is well known (see e.g. [53] for Tikhonov regularization and [32] for the general case) that for minimum norm solutions, that satisfy the source condition qualitative error estimates for regularized solutions can be given (*convergence rates*).

Theorem 2.2.3. *Let $y \in Y$ be attainable. Then there exists a J-minimizing solution of* (2.4).

Proof. Since y is attainable we have in particular that $y \in \operatorname{ran}(K)$ and thus there exists an element x such that $Kx = y$. In other words, the strategy set

$$S = \{v \in X \ : \ Kv = y\} = \ker(K) + x$$

is non empty. Since S is closed and convex, it is also weakly closed, such that by assumption (R1) it follows that S is also closed w.r.t. the topology τ_X. Let $\{v_k\}_{k\in\mathbb{N}}$ be a sequence in S such that

$$\lim_{k\to\infty} J(v_k) = \inf \{J(v) \ : \ v \in S\} =: \mu_0 < \infty.$$

Consequently the sequence $\{J(v_k)\}_{k\in\mathbb{N}}$ is uniformly bounded and since $Kv_k = y$ for all $k \in \mathbb{N}$, one has $v_k \in \Lambda(c)$ for all $k \in \mathbb{N}$ with a suitably chosen constant c. Thus the compactness requirement (R5) and the τ_X-closedness of S imply that there exists a selection $k \mapsto \rho(k)$, such that the subsequence $\left\{v_{\rho(k)}\right\}_{k\in\mathbb{N}}$ τ_X-converges to $\hat{x} \in S$. The τ_X-lower semicontinuity of J eventually implies that

$$J(\hat{x}) \leq \liminf_{k\to\infty} J(v_{\rho(k)}) = \mu_0.$$

□

We move on to the definition of a regularization method for (2.4).

Definition 2.2.4. Let $y \in Y$ be attainable and I be an index set. A family $\{\mathcal{R}_\gamma : Y \to X\}_{\gamma\in I}$ is called a family of *regularizing operators* for (2.4) if there exists a function $\Gamma : [0,\infty) \times Y \to I$ such that for all sequences $\{\delta_n\}_{n\in\mathbb{N}} \subset [0,\infty)$ and $\{y_n\}_{n\in\mathbb{N}} \subset Y$ satisfying

$$\|y - y_n\| \leq \delta_n \quad \text{and} \quad \lim_{n\to\infty} \delta_n = 0$$

the following conditions hold:

1. There exists $M \in \mathbb{R}$ such that

$$\limsup_{n\to\infty} J\left(\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n)\right) \leq M.$$

2. There exists a τ_X-sequentially compact set $C \subset X$ such that $\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n) \in C$ for all $n \in \mathbb{N}$ and every τ_X-cluster point is a solution of (2.4)

In case the above conditions are fulfilled, one calls Γ a *parameter choice rule* and the pair $\left(\{\mathcal{R}_\gamma\}_{\gamma\in I}, \Gamma\right)$ a *regularization method for* (2.4).

Remark 2.2.5. 1. Of particular interest is the case when M in Definition 2.2.4 can be set to

$$M = \inf_{v\in X} \{J(v) \ : \ Kv = y\}.$$

In that case, every τ_X-cluster point of $\left\{\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n)\right\}_{n\in\mathbb{N}}$ is a J-minimizing solution of (2.4). In particular, if $x \in X$ is the *unique* J-minimizing solution of (2.4) one has

$$\lim_{n\to\infty} \mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n) = x \quad \text{w.r.t. } \tau_X.$$

2. In Definition 2.2.4 the notion of a regularization method is introduced in a fairly general manner, since it covers the (discrete) iterative case ($I = \mathbb{N}$) presented in this chapter as well as the continuous case ($I = [0, \infty)$) as it will be studied in Chapter 3.

In what follows we will introduce a (countable) family of regularizing operator (i.e. $I = \mathbb{N}$) for (2.4) in Algorithm 2.2.9. Before we do so, we need some preparations.

Definition 2.2.6. Let $y \in Y$ and $\alpha > 0$. The function $L_\phi(\cdot, \cdot; y, \alpha) : X \times Y^* \to \overline{\mathbb{R}}$ defined by

$$L_\phi(v, q^*; y, \alpha) = \psi_\phi\left(\|Kv - y\|\right) + \alpha\left(J(v) - \langle q^*, Kv - y\rangle\right) \tag{2.7}$$

is called *augmented Lagrangian* of the constrained problem (2.5).

Remark 2.2.7. In the optimization literature the *Lagrangian of* (2.5) usually is defined as

$$L(v, q^*; y) = J(v) - \langle q^*, Kv - y\rangle$$

The function $L_\phi(v, q^*; y, \alpha)$ then is obtained by *augmenting* the Lagrangian by $\psi_\phi\left(\|Kv - y\|\right)$. For further remarks on the notion see Section 2.6.

There exists a convenient relation between J-minimizing solutions and the augmented Lagrangian, which can be expressed in terms of *saddle points*: The saddle points of the augmented Lagrangian of (2.5) are exactly those J-minimizing solutions of (2.4) that satisfy the source condition (2.6).

Proposition 2.2.8. *Let $y \in Y$ be attainable, $\alpha > 0$ and $(x, p^*) \in X \times Y^*$. Then the following two statements are equivalent:*

1. *The element x is a J-minimizing solution of* (2.4) *and satisfies the source condition* (2.6) *with source element p^*.*
2. *The pair (x, p^*) is a saddle point of L_ϕ, that is,*

$$L_\phi(x, q^*; y, \alpha) \leq L_\phi(x, p^*; y, \alpha) \leq L_\phi(v, p^*; y, \alpha) \tag{2.8}$$

holds for all $(v, q^*) \in X \times Y^*$.

Proof. (1) $\Rightarrow$ (2): If a J-minimizing solution x satisfies (2.6) with source element p^* it follows from $K^*p^* \in \partial J(x)$ that

$$J(v) \leq J(x) - \langle p^*, Kv - y\rangle$$

for all $v \in X$ and in turn

$$L_\phi(x, q^*; y, \alpha) = \alpha J(x) = L_\phi(x, p^*; y, \alpha) \leq \alpha(J(v) - \langle p^*, Kv - y\rangle) \leq L_\phi(v, p^*; y, \alpha).$$

for all $v \in X$ and $q^* \in Y^*$.

(2) $\Rightarrow$ (1): Assume that (2.8) holds for all $v \in X$ and $q^* \in Y^*$ and observe that

$$\frac{\partial}{\partial q^*} L_\phi(x, q^*; y, \alpha) = Kx - y.$$

Therefore the first inequality immediately shows that $Kx = y$. The second inequality in (2.8) is equivalent to

$$0 \in \partial_1 L_\phi(x, p^*; y, \alpha)$$

where ∂_1 denotes the subgradient operator w.r.t. the first variable. Note that the mapping $f : v \mapsto \psi_\phi(\|Kv - y\|)$ is continuous. Thus the Moreau – Rockafellar Theorem (cf. [51, Chap. 1 Prop. 5.6]) is applicable and (together with Lemma A.2.18) this shows that

$$0 \in \partial_1 L_\phi(x, p^*; y, \alpha) = \partial f(x) + \alpha\left(\partial J(x) - K^* p^*\right) = K^* \mathfrak{J}_\phi(Kx - y) + \alpha\left(\partial J(x) - K^* p^*\right).$$

Here $\mathfrak{J}_\phi$ denotes the duality mapping on Y with weight ϕ (recall Definition A.1.1). Since $Kx - y = 0$ it follows from Remark A.1.2 that $\mathfrak{J}_\phi(Kx - y) = \{0\}$ and therefore the above inclusion implies $K^* p^* \in \partial J(x)$. Again by the second inequality we find that

$$\alpha J(x) = L_\phi(x, p^*; y, \alpha) \leq L_\phi(v, p^*; y, \alpha) = \psi_\phi(\|Kv - y\|) + \alpha(J(v) - \langle p^*, Kv - y\rangle)$$

for all $v \in X$ and in particular $J(x) \leq J(v)$ for all v such that $Kv = y$, that is x is a J-minimizing solution of (2.4). □

With these preparations we are able to formulate the fundamental algorithm of this chapter, which consists in iteratively minimizing the Lagrangian of (2.5) w.r.t to the primal variable x and a subsequent update of the dual variable p^*. This procedure is known as the *augmented Lagrangian method.*

Algorithm 2.2.9 (Augmented Lagrangian method)**.** Let $y \in Y$. For $n = 1, 2, \ldots$ compute

$$x_n \in \operatorname*{argmin}_{v \in X} L_\phi(v, p^*_{n-1}; y, \alpha_n) \tag{2.9a}$$

and choose $p^*_n \in Y^*$ such that

$$p^*_n \in p^*_{n-1} + \alpha_n^{-1} \mathfrak{J}_\phi(y - Kx_n) \quad \text{and} \quad K^* p^*_n \in \partial J(x_n). \tag{2.9b}$$

Remark 2.2.10. The minimization in (2.9a) (provided that it is well defined) is not affected by adding constants to the Lagrangian L_ϕ. Therefore we can formulate an equivalent objective function for the minimization problem in (2.9a) by adding

$$-\alpha_n\left(J(x_{n-1}) + \left\langle p^*_{n-1}, y + Kx_{n-1}\right\rangle\right).$$

to $L_\phi(v, p^*_{n-1}; y, \alpha_n)$ in the n-th step. This results in

$$\begin{aligned} x_n &\in \operatorname*{argmin}_{v \in X} \psi_\phi\left(\|Kv - y\|\right) + \alpha_n\left(J(v) - J(x_{n-1}) - \left\langle K^* p^*_{n-1}, v - x_{n-1}\right\rangle\right) \\ &= \operatorname*{argmin}_{v \in X} \psi_\phi\left(\|Kv - y\|\right) + \alpha_n D_J^{K^* p^*_{n-1}}(v, x_{n-1}), \end{aligned} \tag{2.10}$$

where $D_J^{\xi^*}(x, y)$ denotes the Bregman distance between x and y w.r.t. J and $\xi^* \in \partial J(y)$ (cf. Definition A.2.6). In other words this means that the augmented Lagrangian method consists in iteratively minimizing the (weighted) sum of $\psi_\phi\left(\|Kx - y\|\right)$ and the Bregman distance to the previous iterate.

A-priori it is not clear that Algorithm 2.2.9 — neither the minimization (2.9a) nor the choice of p^*_n in (2.9b) — is well defined. Under Assumption 2.1.1, however, well-posedness can be shown with standard techniques of variational calculus:

Theorem 2.2.11. *Let $y \in Y$. Then, the n-th step in Algorithm 2.2.9 is well defined, that is, for every $n \geq 1$ there exist elements $(x_n, p_n^*) \in X \times Y^*$, such that x_n is a minimizer of $L_\phi(\cdot, p_{n-1}^*; y, \alpha_n)$ and p_n^* satisfies* (2.9b).

Proof. Let $n = 1$. Moreover, assume that $\{v_k\}_{k\in\mathbb{N}} \subset X$ is such that

$$\lim_{k\to\infty} L_\phi(v_k, p_0^*; y, \alpha_1) = \inf_{v\in X} L_\phi(v, p_0^*; y, \alpha_1) =: \mu_0 < \infty.$$

This clearly implies that $L_\phi(v_k, p_0^*; y, \alpha_1) \leq c_1 < \infty$ for all $k \in \mathbb{N}$ and a constant $c_1 \in \mathbb{R}$. According to (R6) we have that $K^* p_0^* \in \partial J(x_0)$. The definition of the subgradient then gives that

$$J(v) - \langle p_0^*, Kv\rangle \geq J(x_0) - \langle p_0^*, Kx_0\rangle =: c_2 > -\infty$$

for all $v \in X$. For $k \in \mathbb{N}$ it follows that

$$-\alpha_1 \left(J(v_k) - \langle p_0^*, Kv_k - y\rangle\right) \leq -\alpha_1 \left(J(x_0) - \langle p_0^*, Kx_0 - y\rangle\right) =: c_2$$

and we deduce that

$$\psi_\phi\left(\|Kv_k - y\|\right) = L_\phi(v_k, p_0^*; y, \alpha_1) - \alpha_1 \left(J(v_k) - \langle p_0^*, Kv_k - y\rangle\right) \leq c_1 + c_2 < \infty. \tag{2.11}$$

Moreover, we find that for all $v \in X$

$$\alpha_1 J(v) = L_\phi(v, p_0^*; y, \alpha_1) - \psi_\phi\left(\|Kv - y\|\right) + \alpha_1 \langle p_0^*, Kv - y\rangle \leq c_1 + \|p_0^*\| \, \|Kv - y\|$$

Since ψ_ϕ is increasing it follows from (2.11) that for $k \in \mathbb{N}$

$$\alpha_1 J(v_k) \leq c_1 + \|p_0^*\| \, \|Kv_k - y\| \leq c_1 + \psi_\phi^{-1}(c_1 + c_2). \tag{2.12}$$

Combining (2.11) with (2.12) results in

$$\psi_\phi\left(\|Kv_k - y\|\right) + \alpha_1 J(v_k) \leq 2c_1 + c_2 + \psi_\phi^{-1}(c_1 + c_2) =: c_3$$

or, in other words, $v_k \in \Lambda(c_3)$ for all $k \in \mathbb{N}$. The compactness assumption (R5) thus shows existence of a selection $k \mapsto \rho(k)$ such that

$$\lim_{k\to\infty} v_{\rho(k)} = \hat{x} \in X, \qquad \text{w.r.t. } \tau_X.$$

Furthermore, the mapping

$$v \mapsto L_\phi(v, p_0^*; y, \alpha_1)$$

is sequentially τ_X-lower semicontinuous according to Remark 2.1.2. Combining these two facts we get

$$L_\phi(\hat{x}, p_0^*; y, \alpha_1) \leq \liminf_{k\to\infty} L_\phi(v_{\rho(k)}, p_0^*; y, \alpha_1) = \mu_0$$

or in other words, $\hat{x}$ is a minimizing argument of $L_\phi(\cdot, p_0^*; y, \alpha_1)$.

Let x_1 be a minimizer of $L_\phi(\cdot, p_0^*; y, \alpha_1)$. It remains to show, that $p_1^* \in Y^*$ can be chosen such that (2.9b) holds. Since the mapping $f : x \mapsto \psi_\phi\left(\|Kx - y\|\right)$ is continuous, it follows from the Moreau – Rockafellar Theorem (cf. [51, Chap. 1 Prop. 5.6]) that

$$\begin{aligned} 0 \in \partial_1 L(x_1, p_1^*; y, \alpha) &= \partial f(x_1) + \alpha_1 \partial J(x_1) - \alpha_1 K^* p_0^* \\ &= K^* \mathfrak{J}_\phi(Kx_1 - y) + \alpha_1 \partial J(x_1) - \alpha_1 K^* p_0^*, \end{aligned}$$

where the last equality follows from Lemma A.2.18. If $h \in \mathfrak{J}_\phi(Kx_1 - y)$ is such that

$$0 \in K^*h - \alpha_1 K^* p_0^* + \alpha_1 \partial J(x_1),$$

then the element $p_1^* := p_0^* - \frac{h}{\alpha_1}$ meets the conditions

$$p_1^* \in p_0^* + \frac{1}{\alpha_1}\mathfrak{J}_\phi(y - Kx_1) \quad \text{and} \quad K^* p_1^* \in \partial J(x_1)$$

as desired. By induction the Theorem is proven. □

Remark 2.2.12. Theorem 2.2.11 states that Algorithm 2.2.9 is well-defined for each $y \in Y$. We will henceforth assume that for $y \in Y$ an *arbitrary* pair of sequences $\{(x_n, p_n^*)\}_{n\in\mathbb{N}} \subset X \times Y^*$ generated by Algorithm 2.2.9 is chosen. We define

$$\mathcal{R}_n(y) := x_n \quad \text{and} \quad \mathcal{R}_n^*(y) := p_n^*$$

for each $n \in \mathbb{N}$ and end up with two families of (primal and dual) operators

$$\{\mathcal{R}_n : Y \to X\}_{n\in\mathbb{N}} \quad \text{and} \quad \{\mathcal{R}_n^* : Y \to Y^*\}_{n\in\mathbb{N}}.$$

These operators will be the basic objects of study in this chapter.

In the remainder of this section we develop a dual representation of the augmented Lagrangian method by using the Legendre – Fenchel duality concept (cf. e.g. Ekeland & Temam [51, Chap. 3]).

To this end, assume that $y \in Y$ is fixed and consider the mapping $F(\cdot; y) : Y \to \overline{\mathbb{R}}$ given by

$$F(w; y) \mapsto \begin{cases} \inf\{J(v) \ : \ v \in X,\ Kv = y + w\} & \text{if } y + w \text{ is attainable,} \\ +\infty & \text{else.} \end{cases}$$

That is, $F(\cdot; y)$ maps elements $w \in Y$ on the value of problem (2.5) with perturbed right hand side $y + w$ in Equation (2.4).

Lemma 2.2.13. Let $y \in Y$. Then

$$F^*(q^*; y) = J^*(K^* q^*) - \langle q^*, y\rangle, \qquad \text{for all } q^* \in Y^*. \tag{2.13}$$

Proof. We directly compute

$$\begin{aligned} F^*(q^*; y) &= \sup_{w\in Y} \left(\langle q^*, w\rangle - F(w; y)\right) \\ &= \sup_{w\in Y} \left(\langle q^*, w\rangle - \inf\{J(v) \ : \ v \in X,\ Kv = y + w\}\right) \\ &= \sup_{w\in Y} \sup_{Kv = y+w} \left(\langle q^*, w\rangle - J(v)\right) \\ &= \sup_{w\in Y} \sup_{Kv = y+w} \left(\langle K^* q^*, v\rangle - J(v)\right) - \langle q^*, y\rangle \\ &= \sup_{v\in X} \left(\langle K^* q^*, v\rangle - J(v)\right) - \langle q^*, y\rangle = J^*(K^* q^*) - \langle q^*, y\rangle. \end{aligned}$$

□

The functional $F^*(\cdot; y) : Y^* \to \overline{\mathbb{R}}$ will play a key role in the dual characterization of Algorithm 2.2.9. We therefore collect some basic properties

Lemma 2.2.14. Let $y \in Y$. The functional $F^*(\cdot; y) : Y^* \to \overline{\mathbb{R}}$ is convex, proper and sequentially weakly* lower semicontinuous.

Moreover, if y is attainable, then

$$\inf_{q^* \in Y^*} F^*(q^*; y) > -\infty.$$

Proof. Lower semicontinuity and convexity follow from Lemma A.2.14. Moreover, requirement (R6) says that $K^* p_0^* \in \partial J(x_0)$ which is equivalent to $i_X(x_0) \in \partial J^*(K^* p_0^*)$ according to Lemma A.2.12. Here $i_X : X \to X^{**}$ denotes the natural mapping on X. This in particular implies that $K^* p_0^* \in D(J^*)$ and it follows that

$$F^*(p_0^*; y) = J^*(K^* p_0^*) - \langle p_0^*, y \rangle < \infty.$$

Thus $F^*(\cdot; y)$ is proper.

It remains to check, that $F^*(\cdot; y)$ is bounded from below. To this end, observe from Lemma 2.2.13 that

$$F^*(q^*; y) \geq \langle q^*, w \rangle - \inf_{v \in X} \{ J(v) \;:\; Kv = y + w \}$$

for all $w \in Y$. Setting $w = 0$ and the fact that y is attainable yield the desired result. $\square$

Before we proceed, we recall that for a given weight function ϕ, the inverse ϕ^{-1} is well defined due to strict monotonicity and, moreover, a weight function in its own right (cf. Appendix Section A.1). If $\mathfrak{J}_\phi$ denotes the duality mapping on Y w.r.t. ϕ and $\mathfrak{J}_{\phi^{-1}}$ the duality mapping on Y^* w.r.t. ϕ^{-1}, then one has for all $y \in Y$ (cf. Lemma A.1.5)

$$\mathfrak{J}_{\phi^{-1}}(\mathfrak{J}_\phi(y)) = i_Y(y). \tag{2.14}$$

Here $i_Y : Y \to Y^{**}$ denotes the natural mapping on Y.

Proposition 2.2.15. *Let $y \in Y$. The dual sequence $\{\mathcal{R}_n^*(y)\}_{n \in \mathbb{N}}$ generated by Algorithm 2.2.9 satisfies*

$$\mathcal{R}_n^*(y) \in \operatorname*{argmin}_{q^* \in Y^*} \alpha_n^{-1} \psi_{\phi^{-1}} \left(\alpha_n \left\| q^* - \mathcal{R}_{n-1}^*(y) \right\| \right) + F^*(q^*; y). \tag{2.15}$$

Proof. Let $y \in Y$ and define for $n \in \mathbb{N}$

$$x_n := \mathcal{R}_n(y) \quad \text{and} \quad p_n^* := \mathcal{R}_n^*(y).$$

From the update rule (2.9b) we then have $K^* p_n^* \in \partial J(x_n)$ for all $n \in \mathbb{N}$ and since J is proper, it follows from Lemma A.2.12 that

$$i_X(x_n) \in \partial J^* \left(K^* p_n^* \right).$$

This means that

$$\langle \xi^* - K^* p_n^*, x_n \rangle + J^*(K^* p_n^*) = \langle i_X(x_n), \xi^* - K^* p_n^* \rangle_{X^{**}, X^*} + J^*(K^* p_n^*) \leq J^*(\xi^*)$$

holds for all $\xi^* \in X^*$. Considering this inequality for $\xi^* = K^* q^*$ where $q^* \in Y^*$ shows

$$\langle Kx_n, q^* - p_n^* \rangle_{Y^{**},Y^*} + J^*(K^* p_n^*) = \langle q^* - p_n^*, Kx_n \rangle + J^*(K^* p_n^*) \leq J(K^* q^*),$$

which is equivalent to $i_Y(Kx_n) \in \partial\left(J^* \circ K^*\right)(p_n^*)$.

From (2.9b) and Remark A.1.2 we find that $\alpha_n(p_{n-1}^* - p_n^*) \in \mathfrak{J}_\phi(Kx_n - y)$ and therefore according to (2.14)

$$\mathfrak{J}_{\phi^{-1}}\left(\alpha_n(p_{n-1}^* - p_n^*)\right) = i_Y(Kx_n) - i_Y(y) \in \partial\left(J^* \circ K^*\right)(p_n^*) - i_Y(y)$$

For the function $f : q^* \mapsto \psi_{\phi^{-1}}(\alpha_n \left\|q^* - p_{n-1}^*\right\|)$, this together with Asplund's theorem A.2.4 shows that

$$\partial f(p_n^*) = \alpha_n \mathfrak{J}_{\phi^{-1}}(\alpha_n(p_n^* - p_{n-1}^*)) \in -\left(\alpha_n \partial\left(J^* \circ K^*\right)(p_n^*) - i_Y(y)\right).$$

Combining the previous two equations and applying the Moreau – Rockafellar Theorem [51, Chap. 1 Prop. 5.6] (f is continuous) now shows that

$$\begin{aligned} 0 \in \partial f(p_n^*) + \alpha_n &\left(\partial\left(J^* \circ K^*\right)(p_n^*) - i_Y(y)\right) \\ &= \partial f(p_n^*) + \alpha_n \partial F^*(p_n^*; y) = \partial\left(f + \alpha_n \partial F^*(\cdot; y)\right)(p_n^*) \end{aligned}$$

which is equivalent to the fact that p_n^* minimizes $f(\cdot) + \alpha_n F^*(\cdot; y)$ over Y^* or in other words

$$p_n^* \in \operatorname*{argmin}_{q^* \in Y^*} \alpha_n^{-1} \psi_{\phi^{-1}}\left(\alpha_n \left\|q^* - p_{n-1}^*\right\|\right) + F^*(q^*; y).$$

□

Proposition 2.2.15 amounts to saying that the (dual) sequence $\{\mathcal{R}_n^*(y)\}_{n\in\mathbb{N}}$ in the augmented Lagrangian method is characterized by another iteration process. More explicitly:

Algorithm 2.2.16. Let $y \in Y$. For $n = 1, 2, \ldots$ compute

$$p_n^* \in \operatorname*{argmin}_{q^* \in Y^*} \alpha_n^{-1} \psi_{\phi^{-1}}\left(\alpha_n \left\|q^* - p_{n-1}^*\right\|\right) + F^*(q^*; y).$$

This gives rise to the following (cf. references in Section 2.6)

Definition 2.2.17. Let Z be a Banach space and assume that $G : Z \to \overline{\mathbb{R}}$ is a proper and convex functional.

1. Let $\alpha > 0$. Then the mapping $R_G^\alpha : Z \to \mathfrak{P}(Z)$ defined by

$$R_G^\alpha(z) = \operatorname*{argmin}_{w \in Z} \alpha^{-1} \psi_{\phi^{-1}}\left(\alpha \left\|w - z\right\|\right) + G(w) \tag{2.16}$$

   is called *Resolvent operator of* G.

2. Assume that $\{\alpha_n\}_{n\in\mathbb{N}}$ is a sequence of positive parameters and that $z_0 \in Z$. Then the algorithm that computes for $n = 1, 2, \ldots$

$$z_n \in R_G^{\alpha_n}(z_{n-1})$$

   (provided z_n exists) is called *proximal point algorithm w.r.t.* G.

Remark 2.2.18. Proposition 2.2.15 thus states that for each $y \in Y$ the dual sequence $\{\mathcal{R}_n^*(y)\}_{n\in\mathbb{N}}$ in the augmented Lagrangian Algorithm 2.2.9 is characterized by Algorithm 2.2.16, the proximal point algorithm w.r.t. $F^*(\cdot;y)$ (as in (2.13)).

We close this section with a complementary remark on duality: The unconstrained problem

$$J^*(K^*q^*) - \langle q^*, y\rangle \to \inf! \quad q^* \in Y^*. \tag{2.17}$$

is called the *dual problem* of problem (2.5). The two problems are linked via the *Karush – Kuhn – Tucker* conditions. To be more precise, one finds

Proposition 2.2.19. *Let $y \in Y$ be attainable and $(x, p^*) \in X \times Y^*$. Then the following two statements are equivalent*

1. *The elements x and p^* are solutions of* (2.5) *and* (2.17) *respectively and*

$$J(x) + J^*(K^*p^*) - \langle p^*, y\rangle = 0.$$

2. *The Karush – Kuhn – Tucker conditions hold:*

$$Kx = y \quad \text{and} \quad K^*p^* \in \partial J(x).$$

Proof. [51, Chap. 3 Prop. 4.1] □

Remark 2.2.20. Theorem 2.2.19 amounts to saying that as soon as an *arbitrary* solution x of Equation (2.4) satisfies the source condition (2.6), it is already a J-minimizing solution. Moreover, in that case the source element p^* is a solution of the dual problem (2.17).

In view of Proposition 2.2.8 this shows that $(x, p^*) \in X \times Y^*$ is a saddle point of the augmented Lagrangian L_ϕ if and only if x is a solution of (2.4) and

$$K^*p^* \in \partial J(x)$$

holds.

2.3 Convergence for Data in a Banach Space

In this section we will study the regularizing properties of the operator family

$$\{\mathcal{R}_n : Y \to X\}_{n\in\mathbb{N}}$$

as introduced in the previous section (cf. Remark 2.2.12). That is, the issue lies in finding a parameter choice rule $\Gamma : [0,\infty) \times Y \to \mathbb{N}$ such that $\left(\{\mathcal{R}_n\}_{n\in\mathbb{N}}, \Gamma\right)$ forms a regularization method for the operator equation (2.4) (cf. Definition 2.2.4).

For $y \in Y$ we adopt the notation from the previous section and make use of the dual functional $F^*(\cdot;y)$ as given in (2.13) and in order to keep the presentation transparent we abbreviate

$$\mu^*(y) := \inf_{q^*\in Y^*} F^*(q^*;y).$$

According to Lemma 2.2.14, $\mu^*(y)$ is finite, whenever y is attainable, i.e. if there exists $x \in D(J)$ such that $Kx = y$.

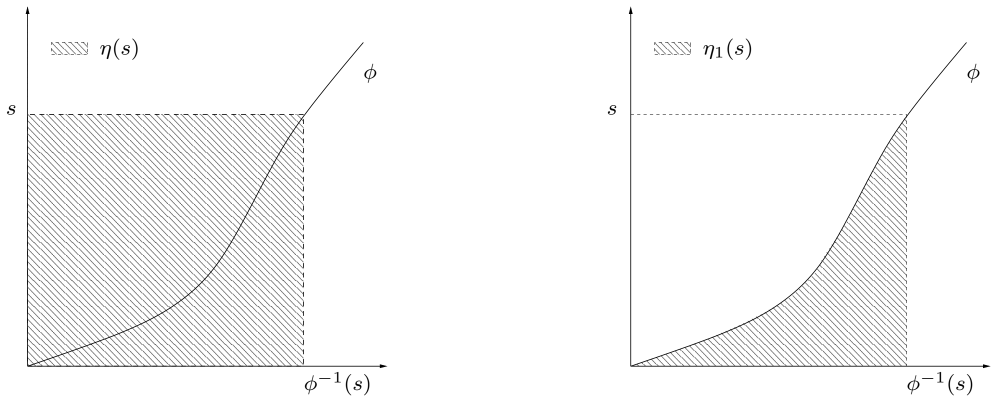

Figure 2.1: The auxiliary functions η (left) and η_1 (right).

Throughout this section we will make use of the following two auxiliary functions defined for each $s \geq 0$ by

$$\eta(s) := \phi^{-1}(s)s \quad \text{and} \quad \eta_1(s) := (\psi_\phi \circ \phi^{-1})(s).$$

The graphs of the two functions are depicted in Figure 2.1. It becomes (visually) evident that both, η and η_1, are increasing and mutually related by

$$\eta_1(s) = \eta(s) - \psi_{\phi^{-1}}(s) \tag{2.18}$$

for all $s \geq 0$.

We start our analysis with an (asymptotic) estimate for the residuals in Theorem 2.3.4. To this end we shall at first prove a result on monotonicity. We refer to Osher et al. [103, Prop. 3.2] for the corresponding result in Hilbert spaces and to Kiwiel [87, Lem. 4.1] for a finite dimensional version.

Lemma 2.3.1. Let $y \in Y$, $n \geq 1$ and set $\xi_n^* := K^* \mathcal{R}_n^*(y)$. Then

$$\|K\mathcal{R}_n(y) - y\| \leq \|K\mathcal{R}_{n-1}(y) - y\|.$$

Moreover, for all[1] $v \in X$

$$\alpha_n \Big(D_J^{\xi_n^*}(v, \mathcal{R}_n(y)) + D_J^{\xi_{n-1}^*}(\mathcal{R}_n(y), \mathcal{R}_{n-1}(y)) - D_J^{\xi_{n-1}^*}(v, \mathcal{R}_{n-1}(y)) \Big) \\ \leq \psi_\phi\left(\|Kv - y\|\right) - \psi_\phi\left(\|K\mathcal{R}_n(y) - y\|\right).$$

Proof. For the sake of simplicity, we agree upon the abbreviation

$$x_n := \mathcal{R}_n(y), \quad \text{and} \quad p_n^* := \mathcal{R}_n^*(y), \quad \text{for all } n \in \mathbb{N}.$$

[1] In case that $v \notin D(J)$ we agree upon

$$D_J^{\xi_n^*}(v, \mathcal{R}_n(y)) - D_J^{\xi_{n-1}^*}(v, \mathcal{R}_{n-1}(y)) = \infty - \infty =: 0$$

Since for each $n \geq 1$ one has $K^* p^*_{n-1} \in \partial J(x_{n-1})$ according to (2.9b) it follows from the definition of the subgradient that for an arbitrary $v \in X$

$$J(v) - J(x_{n-1}) + \left\langle K^* p^*_{n-1}, x_{n-1} - v \right\rangle \geq 0. \tag{2.19}$$

Hence we observe by setting $v = x_n$ in above inequality from optimality in (2.9a) that

$$\begin{aligned}
\psi_\phi\left(\|Kx_n - y\|\right) &\leq \psi_\phi\left(\|Kx_n - y\|\right) + \alpha_n\left(J(x_n) - J(x_{n-1}) + \left\langle \xi^*_{n-1}, x_{n-1} - x_n \right\rangle\right) \\
&= \psi_\phi\left(\|Kx_n - y\|\right) + \alpha_n\left(J(x_n) - J(x_{n-1}) + \left\langle p^*_{n-1}, Kx_{n-1} - Kx_n \right\rangle\right) \\
&= \psi_\phi\left(\|Kx_n - y\|\right) + \alpha_n\left(J(x_n) - \left\langle p^*_{n-1}, Kx_n - y \right\rangle\right) \\
&\quad - \alpha_n\left(J(x_{n-1}) - \left\langle p^*_{n-1}, Kx_{n-1} - y \right\rangle\right) \\
&\leq \psi_\phi\left(\|Kx_{n-1} - y\|\right) + \alpha_n\left(J(x_{n-1}) - \left\langle p^*_{n-1}, Kx_{n-1} - y \right\rangle\right) \\
&\quad - \alpha_n\left(J(x_{n-1}) - \left\langle p^*_{n-1}, Kx_{n-1} - y \right\rangle\right) \\
&= \psi_\phi\left(\|Kx_{n-1} - y\|\right).
\end{aligned} \tag{2.20}$$

The first assertion of the lemma follows from the monotonicity of ψ_ϕ.

In order to show the second inequality, note that for $f : v \mapsto \psi_\phi\left(\|Kv - y\|\right)$ it follows from (2.9b) and Lemma A.2.18 that

$$\alpha_n(\xi^*_{n-1} - \xi^*_n) = \alpha_n K^*(p^*_{n-1} - p^*_n) \in K^* \mathfrak{J}_\phi(Kx_n - y) = \partial f(x_n)$$

and thus for $n \geq 1$ and $v \in X$

$$\psi_\phi\left(\|Kv - y\|\right) - \psi_\phi\left(\|Kx_n - y\|\right) \geq \alpha_n \left\langle \xi^*_{n-1} - \xi^*_n, v - x_n \right\rangle.$$

Eventually, it follows from the definition of the Bregman distance (cf. Definition A.2.6) that

$$\begin{aligned}
&\alpha_n\left(D_J^{\xi^*_n}(v, x_n) + D_J^{\xi^*_{n-1}}(x_n, x_{n-1}) - D_J^{\xi^*_{n-1}}(v, x_{n-1})\right) \\
&\qquad = \alpha_n \left\langle \xi^*_{n-1} - \xi^*_n, v - x_n \right\rangle \leq \psi_\phi\left(\|Kv - y\|\right) - \psi_\phi\left(\|Kx_n - y\|\right)
\end{aligned}$$

for all $v \in X$. □

Corollary 2.3.2. If $y \in \operatorname{ran}(K)$ and x is a solution of (2.4), then for all $n \geq 1$

$$D_J^{\xi^*_n}(x, \mathcal{R}_n(y)) \leq D_J^{\xi^*_n}(x, \mathcal{R}_n(y)) + D_J^{\xi^*_{n-1}}(\mathcal{R}_n(y), \mathcal{R}_{n-1}(y)) \leq D_J^{\xi^*_{n-1}}(x, \mathcal{R}_{n-1}(y))$$

Proof. From Lemma 2.3.1 it follows for all $n \geq 1$ that

$$\begin{aligned}
&\alpha_n\left(D_J^{\xi^*_n}(x, \mathcal{R}_n(y)) + D_J^{\xi^*_{n-1}}(\mathcal{R}_n(y), \mathcal{R}_{n-1}(y)) - D_J^{\xi^*_{n-1}}(x, \mathcal{R}_{n-1}(y))\right) \\
&\qquad \leq \psi_\phi\left(\|Kx - y\|\right) - \psi_\phi\left(\|K\mathcal{R}_n(y) - y\|\right) = -\psi_\phi\left(\|K\mathcal{R}_n(y) - y\|\right) \leq 0.
\end{aligned}$$

The assertion follows from the fact that $\alpha_n > 0$ and $D_J^{\xi^*_{n-1}}(\mathcal{R}_n(y), \mathcal{R}_{n-1}(y)) \geq 0$. □

Remark 2.3.3. Let $y \in Y$. We note that for each $n \in \mathbb{N}$, it follows from (2.9b) that $\alpha_n(\mathcal{R}^*_n(y) - \mathcal{R}^*_{n-1}(y)) \in \mathfrak{J}_\phi(y - K\mathcal{R}_n(y))$ and thus in turn from the definition of $\mathfrak{J}_\phi$ in (A.1.1) that

$$\phi^{-1}\left(\alpha_n \left\|\mathcal{R}^*_n(y) - \mathcal{R}^*_{n-1}(y)\right\|\right) = \|K\mathcal{R}_n(y) - y\|.$$

Since ϕ^{-1} is monotone, Lemma 2.3.1 implies that

$$\alpha_{n+1}\left\|\mathcal{R}^*_{n+1}(y) - \mathcal{R}^*_n(y)\right\| \leq \alpha_n \left\|\mathcal{R}^*_n(y) - \mathcal{R}^*_{n-1}(y)\right\|.$$

We proceed with the announced estimate for the residuals. We also refer [103, Thm. 3.3 and Thm. 3.5] for similar results which are covered by the upcoming Theorem. Recall the definition of $t_n(\boldsymbol{\alpha})$ in (2.2).

Theorem 2.3.4. *Let $y \in Y$ be attainable and $\tilde{y} \in Y$. Then for all $n \in \mathbb{N}$*

$$\|K\mathcal{R}_n(\tilde{y}) - \tilde{y}\| \leq \psi_\phi^{-1}\left(\frac{F^*(p_0^*, y) - \mu^*(y)}{t_n(\boldsymbol{\alpha})} + \psi_\phi(\|y - \tilde{y}\|)\right). \tag{2.21}$$

Proof. We abbreviate for each $n \in \mathbb{N}$

$$p_n^* := \mathcal{R}_n^*(\tilde{y}) \quad \text{ and } \delta := \|y - \tilde{y}\|\,.$$

Then, by keeping in mind Remark 2.3.3, the assertion to prove is equivalent to

$$\phi^{-1}\left(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|\right) \leq \psi_\phi^{-1}\left(\frac{F^*(p_0^*, y) - \mu^*(y)}{t_n(\boldsymbol{\alpha})} + \psi_\phi(\delta)\right).$$

From Proposition 2.2.15 it follows that the sequence $\{p_n^*\}_{n\in\mathbb{N}}$ is characterized by the proximal point algorithm 2.2.16, that is

$$p_n^* \in \operatorname*{argmin}_{q^* \in Y^*} \alpha_n^{-1}\psi_{\phi^{-1}}\left(\alpha_n \left\|q^* - p_{n-1}^*\right\|\right) + F^*(q^*; \tilde{y})$$

for each $n \geq 1$. By setting $f : q^* \mapsto \alpha_n^{-1}\psi_{\phi^{-1}}\left(\alpha_n \left\|q^* - p_{n-1}^*\right\|\right)$, this implies that

$$0 \in \partial(f + F^*(\cdot; \tilde{y}))(p_n^*) = \partial f(p_n^*) + \partial F^*(p_n^*; \tilde{y}) = \mathfrak{J}_{\phi^{-1}}(\alpha_n(p_n^* - p_{n-1}^*)) + \partial F^*(p_n^*; \tilde{y})$$

We choose elements $g, h \in Y^{**}$ such that $g \in \mathfrak{J}_{\phi^{-1}}(\alpha_n(p_n^* - p_{n-1}^*))$, $h \in \partial F^*(p_n^*; \tilde{y})$ and $g+h = 0$. Evaluating the left and right hand side of this equation at the element $\alpha_n(p_n^* - p_{n-1}^*)$ and taking into account the definition of $\mathfrak{J}_{\phi^{-1}}$ (cf. Definition A.1.1) gives (recall that $\eta(s) = \phi^{-1}(s)s$)

$$0 = \left\langle g + h, \alpha_n(p_n^* - p_{n-1}^*)\right\rangle_{Y^{**},Y^*} = \eta(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) + \alpha_n \left\langle h, p_n^* - p_{n-1}^*\right\rangle_{Y^{**},Y^*}. \tag{2.22}$$

Since $h \in \partial F^*(p_n^*; \tilde{y})$ it follows that $F^*(p_{n-1}^*; \tilde{y}) - F^*(p_n^*; \tilde{y}) \geq \left\langle h, p_{n-1}^* - p_n^*\right\rangle_{Y^{**},Y^*}$. This together with (2.22) implies that

$$\alpha_n^{-1}\eta(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) = \left\langle h, p_{n-1}^* - p_n^*\right\rangle_{Y^{**},Y^*} \leq F^*(p_{n-1}^*; \tilde{y}) - F^*(p_n^*; \tilde{y}) \tag{2.23}$$

for all $n \geq 1$. Since $F^*(p^*; y) = J^*(K^*p^*) - \langle p^*, y\rangle$ one finds

$$\begin{aligned}
\alpha_n^{-1}\eta(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) &\leq J^*(K^*p_{n-1}^*) - \left\langle p_{n-1}^*, \tilde{y}\right\rangle - (J^*(K^*p_n^*) - \langle p_n^*, \tilde{y}\rangle) \\
&= J^*(K^*p_{n-1}^*) - \left\langle p_{n-1}^*, y\right\rangle - (J^*(K^*p_n^*) - \langle p_n^*, y\rangle) \\
&\quad + \left\langle p_{n-1}^* - p_n^*, y - \tilde{y}\right\rangle \\
&= F^*(p_{n-1}^*; y) - F^*(p_n^*; y) + \left\langle p_{n-1}^* - p_n^*, y - \tilde{y}\right\rangle \\
&\leq F^*(p_{n-1}^*; y) - F^*(p_n^*; y) + \left\|p_n^* - p_{n-1}^*\right\| \delta.
\end{aligned} \tag{2.24}$$

We recall that the primitive ψ_ϕ equals the Legendre – Fenchel conjugate of $\psi_{\phi^{-1}}$ (cf. Example A.2.13) and thus one finds for the function $\alpha^{-1}\psi_{\phi^{-1}}$ by applying Lemma A.2.15 that

$(\alpha^{-1}\psi_{\phi^{-1}})^*(s) = \alpha^{-1}\psi_\phi(\alpha s)$ for all $s, \alpha > 0$. Therefore Fenchel's inequality (A.9) shows for all $s, t > 0$ that

$$st = (\alpha^{-1}s)(\alpha t) \leq (\alpha^{-1}\psi_{\phi^{-1}})^*(\alpha^{-1}s) + \alpha^{-1}\psi_{\phi^{-1}}(\alpha t) = \alpha^{-1}\psi_\phi(s) + \alpha^{-1}\psi_{\phi^{-1}}(\alpha t) \tag{2.25}$$

Setting $\alpha = \alpha_n$, $s = \delta$ and $t = \left\|p_n^* - p_{n-1}^*\right\|$ in (2.25) gives together with (2.24)

$$\alpha_n^{-1}\eta(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) \leq F^*(p_{n-1}^*; y) - F^*(p_n^*; y) + \alpha_n^{-1}\psi_{\phi^{-1}}(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) + \alpha_n^{-1}\psi_\phi(\delta).$$

Recall the definition $\eta_1 := \psi_\phi \circ \phi^{-1}$ and relation (2.18). With this the previous estimate yields

$$\begin{aligned} \alpha_n^{-1}\eta_1(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) &= \alpha_n^{-1}\eta(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) - \alpha_n^{-1}\psi_{\phi^{-1}}(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) \\ &\leq F^*(p_{n-1}^*; y) - F^*(p_n^*; y) + \alpha_n^{-1}\psi_\phi(\delta) \end{aligned}$$

Since η_1 is increasing, we conclude from Remark 2.3.3 that $\left\{\eta_1(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|)\right\}_{n\in\mathbb{N}}$ is non-increasing. Therefore we finally get

$$\begin{aligned} t_n(\boldsymbol{\alpha})\eta_1(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) &= \sum_{j=1}^{n} \alpha_j^{-1}\eta_1(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) \\ &\leq \sum_{j=1}^{n} \alpha_j^{-1}\eta_1(\alpha_j \left\|p_j^* - p_{j-1}^*\right\|) \\ &\leq \sum_{j=1}^{n} F^*(p_j^*; y) - F^*(p_{j-1}^*; y) + \alpha_j^{-1}\psi_\phi(\delta) \\ &\leq F^*(p_0^*; y) - \mu^*(y) + t_n(\boldsymbol{\alpha})\psi_\phi(\delta). \end{aligned}$$

This shows that

$$\phi^{-1}\left(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|\right) \leq \psi_\phi^{-1}\left(\frac{F^*(p_0^*, y) - \mu^*(y)}{t_n(\boldsymbol{\alpha})} + \psi_\phi(\delta)\right).$$

and the assertion eventually follows. $\square$

Before we move on to the main result of this section, we need to impose some additional assumptions on the weight function ϕ.

Assumption 2.3.5. The weight function ϕ satisfies the following property: For every $\lambda, \varepsilon > 0$ one has

$$\sup_{\tau\geq\lambda} \phi\left(\frac{\varepsilon}{\tau}\right)\phi(\tau) = \mathcal{O}(\varepsilon) \tag{2.26}$$

as well as

$$\sup_{\tau\geq\lambda} \psi_\phi^{-1}\left(\frac{\varepsilon}{\tau}\right)\psi_\phi^{-1}(\tau) = \mathcal{O}(\varepsilon), \tag{2.27}$$

Remark 2.3.6. The above requirements at first glance seem to be quite technical. We therefore point out that in the (frequently occurring) case $\phi(s) = s^{p-1}$ for $p > 1$ the assumptions are satisfied. For the remainder of this section we will assume that Assumption 2.3.5 holds.

Lemma 2.3.7. Let $\{a_n\}_{n\in\mathbb{N}} \subset [0,\infty)$ and $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ be sequences of positive numbers such that

$$\inf_{n\in\mathbb{N}} b_n =: \delta > 0.$$

Then the following two assertions hold:

1. $a_n\psi_\phi^{-1}(b_n) = \mathcal{O}(\psi_\phi(a_n)b_n)$.
2. For every $\lambda > 0$ there exists a nonnegative and real valued function ζ such that for each $\varepsilon \geq 0$ one has that $\zeta(\varepsilon) = \mathcal{O}(\varepsilon)$ and

$$\eta_1^{-1}\left(\frac{\varepsilon}{\tau}\right) \leq \zeta(\varepsilon)\frac{d}{d\tau}\psi_\phi^{-1}(\tau), \qquad \text{for all } \tau \geq \lambda.$$

Proof. 1. Let $\varepsilon_n := \psi_\phi(a_n)b_n$. Since $\delta \leq b_n$ for all $n \in \mathbb{N}$ it follows from (2.27) that

$$a_n\psi_\phi^{-1}(b_n) = \psi_\phi^{-1}(\psi_\phi(a_n))\psi_\phi^{-1}(b_n) = \psi_\phi^{-1}\left(\frac{\varepsilon_n}{b_n}\right)\psi_\phi^{-1}(b_n) = \mathcal{O}(\varepsilon_n).$$

Hence the first part of the Lemma follows.

2. First we note, that for every diffeomorphism $f : [0,\infty) \to [0,\infty)$ the relation $\frac{d}{d\tau}f^{-1}(\tau) = \frac{1}{f'(f^{-1}(\tau))}$ holds, where $f'(\tau) = \frac{d}{d\tau}f(\tau)$. In particular, setting $f(\tau) = \psi_\phi(\tau)$ and noting that $f'(\tau) = \phi(\tau)$ shows

$$\frac{d}{d\tau}\psi_\phi^{-1}(\tau) = \frac{1}{\phi(\psi_\phi^{-1}(\tau))} = \frac{1}{\eta_1^{-1}(\tau)}. \tag{2.28}$$

Now let $\lambda, \varepsilon > 0$. Then it follows from (2.27) that there exists a nonnegative function $\tilde{\zeta}$ with $\tilde{\zeta}(\varepsilon) = \mathcal{O}(\varepsilon)$ such that for all $\tau \geq \lambda$

$$\psi_\phi^{-1}\left(\frac{\varepsilon}{\tau}\right) \leq \frac{\tilde{\zeta}(\varepsilon)}{\psi_\phi^{-1}(\tau)}.$$

Moreover, we find from (2.26) that another such function ζ satisfying $\zeta(\varepsilon) = \mathcal{O}(\tilde{\zeta}(\varepsilon)) = \mathcal{O}(\varepsilon)$ can be chosen such that

$$\phi\left(\frac{\tilde{\zeta}(\varepsilon)}{\psi_\phi^{-1}(\tau)}\right) \leq \frac{\zeta(\varepsilon)}{\phi(\psi_\phi^{-1}(\tau))}$$

for all $t \geq \lambda$. Combining the previous two estimates and taking into account the monotonicity of ϕ, we further conclude that

$$\eta_1^{-1}\left(\frac{\varepsilon}{\tau}\right) = \phi\left(\psi_\phi^{-1}\left(\frac{\varepsilon}{\tau}\right)\right) \leq \phi\left(\frac{\tilde{\zeta}(\varepsilon)}{\psi_\phi^{-1}(\tau)}\right) \leq \frac{\zeta(\varepsilon)}{\phi(\psi_\phi^{-1}(\tau))} = \frac{\zeta(\varepsilon)}{\eta_1^{-1}(\tau)}.$$

This estimate together with (2.28) finally shows that

$$\eta_1^{-1}\left(\frac{\varepsilon}{\tau}\right) \leq \frac{\zeta(\varepsilon)}{\eta_1^{-1}(\tau)} = \zeta(\varepsilon)\frac{d}{d\tau}\psi_\phi^{-1}(\tau).$$

and the assertion follows. $\square$

Lemma 2.3.8. Let $y \in Y$ be attainable and $\tilde{y} \in Y$. Then there exists a nondecreasing and continuous function $\gamma : [0, \infty) \to \mathbb{R}$ and $\{\zeta_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ such that for $n \geq 1$

$$\|\mathcal{R}_n^*(\tilde{y})\| \leq \gamma(\|\tilde{y} - y\|) + \zeta_n \psi_\phi^{-1}(t_n(\boldsymbol{\alpha})) \quad \text{and} \quad \zeta_n = \mathcal{O}(1 + t_n(\boldsymbol{\alpha})\psi_\phi(\|\tilde{y} - y\|)).$$

Proof. In this proof we set

$$\delta := \|\tilde{y} - y\| \quad \text{and} \quad c := F^*(p_0^*, y) - \mu^*(y).$$

From Theorem 2.3.4 and Remark 2.3.3 it becomes evident that for all $n \geq 1$

$$\alpha_n \left\|\mathcal{R}_n^*(\tilde{y}) - \mathcal{R}_{n-1}^*(\tilde{y})\right\| \leq \eta_1^{-1} \left(\frac{c}{t_n(\boldsymbol{\alpha})} + \psi_\phi(\delta)\right).$$

With this we find

$$\|\mathcal{R}_n^*(\tilde{y})\| \leq \|p_0^*\| + \sum_{j=1}^{n} \left\|\mathcal{R}_j^*(\tilde{y}) - \mathcal{R}_{j-1}^*(\tilde{y})\right\| \leq \|p_0^*\| + \sum_{j=1}^{n} \alpha_j^{-1}\eta_1^{-1} \left(\frac{c}{t_j(\boldsymbol{\alpha})} + \psi_\phi(\delta)\right). \tag{2.29}$$

Since η_1 is increasing, η_1^{-1} exists and is increasing as well. Therefore, the mapping

$$\tau \mapsto \eta_1^{-1} \left(\frac{c}{\tau} + \psi_\phi(\delta)\right)$$

is decreasing and hence we gain

$$\sum_{j=1}^{n} \alpha_j^{-1}\eta_1^{-1} \left(\frac{c}{t_j(\boldsymbol{\alpha})}\psi_\phi(\delta)\right) \leq \alpha_1^{-1}\eta_1^{-1}(c\alpha_1 + \psi_\phi(\delta)) + \int_{t_1(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \eta_1^{-1} \left(\frac{c}{\tau} + \psi_\phi(\delta)\right) \mathrm{d}\tau.$$

By setting $\gamma(\delta) := \|p_0^*\| + \alpha_1^{-1}\eta_1^{-1}(c\alpha_1 + \psi_\phi(\delta))$ it follows from the last inequality and (2.29) as well as from the monotonicity of η_1^{-1} that

$$\|\mathcal{R}_n^*(\tilde{y})\| \leq \gamma(\delta) + \int_{t_1(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \eta_1^{-1} \left(\frac{c}{\tau} + \psi_\phi(\delta)\right) \mathrm{d}\tau \leq \gamma(\delta) + \int_{t_1(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \eta_1^{-1} \left(\frac{c + t_n(\boldsymbol{\alpha})\psi_\phi(\delta)}{\tau}\right) \mathrm{d}\tau.$$

From Lemma 2.3.7 it follows that there exists for each $n \geq 1$ a nonnegative constant ζ_n such that for $t_1(\boldsymbol{\alpha}) \leq \tau \leq t_n(\boldsymbol{\alpha})$

$$\eta_1^{-1} \left(\frac{c + t_n(\boldsymbol{\alpha})\psi_\phi(\delta)}{\tau}\right) \leq \zeta_n \frac{d}{d\tau}\psi_\phi^{-1}(\tau)$$

and $\zeta_n = \mathcal{O}(1 + \psi_\phi(\delta)t_n(\boldsymbol{\alpha}))$. Combining the previous two estimates results in

$$\begin{aligned}\|\mathcal{R}_n^*(\tilde{y})\| &\leq \gamma(\delta) + \zeta_n \int_{t_1(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \frac{d}{d\tau}\psi_\phi^{-1}(\tau)\,\mathrm{d}\tau \\ &= \zeta_n \left(\psi_\phi^{-1}(t_n(\boldsymbol{\alpha})) - \psi_\phi^{-1}(t_1(\boldsymbol{\alpha}))\right) \leq \gamma(\delta) + \zeta_n \psi_\phi^{-1}(t_n(\boldsymbol{\alpha}))\end{aligned}$$

and the lemma is proven. □

With the preliminary results in Lemma 2.3.7 and Lemma 2.3.8 we are now able to prove that the augmented Lagrangian algorithm constitutes a regularization method for (2.4).

Theorem 2.3.9. *Let $y \in Y$ be attainable and $\{y_n\}_{n\in\mathbb{N}} \subset Y$ such that $\delta_n := \|y - y_n\| \to 0$ as $n \to \infty$ and assume further that $\Gamma : (0, \infty) \times Y \to \mathbb{N}$ satisfies*

$$\psi_\phi(\delta_n) t_{\Gamma(\delta_n, y_n)}(\boldsymbol{\alpha}) = \mathcal{O}(1) \quad \text{and} \quad \lim_{n\to\infty} t_{\Gamma(\delta_n,y_n)}(\boldsymbol{\alpha}) = \infty. \tag{2.30}$$

Then $\left(\{\mathcal{R}_n\}_{n\in\mathbb{N}}, \Gamma\right)$ is a regularization method for (2.4). If additionally

$$\lim_{n\to\infty} \psi_\phi(\delta_n) t_{\Gamma(\delta_n,y_n)}(\boldsymbol{\alpha}) = 0 \tag{2.31}$$

one has

$$\limsup_{n\to\infty} J(\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n)) \leq \inf_{v\in X} \{J(v) \ : \ Kv = y\} + F^*(p_0^*; y) - \mu^*(y). \tag{2.32}$$

Proof. Let $x \in X$ be a J-minimizing solution of (2.4) with right hand side y, which exists according to Theorem 2.2.3. Moreover, for each $n \in \mathbb{N}$ we set

$$\nu(n) := \Gamma(\delta_n, y_n), \quad x_n := \mathcal{R}_{\nu(n)}(y_n) \quad \text{and} \quad p_n^* := \mathcal{R}^*_{\nu(n)}(y_n).$$

First, note that for all $n \geq 1$ one has $K^* p_n^* \in \partial J(x_n)$ according to (2.9b). This implies that

$$\begin{aligned} J(x_n) \leq J(x) + \langle K^* p_n^*, x_n - x\rangle &\leq J(x) + \|p_n^*\| \, \|Kx_n - y\| \\ &\leq J(x) + \|p_n^*\| \, \delta_n + \|p_n^*\| \, \|Kx_n - y_n\| . \end{aligned} \tag{2.33}$$

Recall from Remark 2.3.3, that for all $j, n \in \mathbb{N}$ one has

$$\phi^{-1}\left(\alpha_j \left\|\mathcal{R}_j^*(y_n) - \mathcal{R}_{j-1}^*(y_n)\right\|\right) = \|K\mathcal{R}_j(y_n) - y_n\| .$$

From this and from the monotonicity assertion in Lemma 2.3.1 it then follows that

$$\begin{aligned} \|p_n^*\| \, \|Kx_n - y_n\| &\leq \|p_0^*\| \, \|Kx_n - y_n\| + \sum_{j=1}^{\nu(n)} \left\|\mathcal{R}_j^*(y_n) - \mathcal{R}_{j-1}^*(y_n)\right\| \left\|K\mathcal{R}_{\nu(n)}(y_n) - y_n\right\| \\ &\leq \varepsilon_n + \sum_{j=1}^{\nu(n)} \left\|\mathcal{R}_j^*(y_n) - \mathcal{R}_{j-1}^*(y_n)\right\| \|K\mathcal{R}_j(y_n) - y_n\| \\ &= \varepsilon_n + \sum_{j=1}^{\nu(n)} \left\|\mathcal{R}_j^*(y_n) - \mathcal{R}_{j-1}^*(y_n)\right\| \phi^{-1}(\alpha_j \left\|\mathcal{R}_j^*(y_n) - \mathcal{R}_{j-1}^*(y_n)\right\|) \end{aligned} \tag{2.34}$$

where according to Theorem 2.3.4 $\varepsilon_n = \|p_0^*\| \, \|Kx_n - y_n\|$ satisfies $\lim_{n\to\infty} \varepsilon_n = 0$. Combining (2.34) with (2.23) yields (note that $\eta(s) = \phi^{-1}(s)s$)

$$\begin{aligned} \|p_n^*\| \, \|Kx_n - y_n\| &\leq \varepsilon_n + \sum_{j=1}^{\nu(n)} \alpha_j^{-1} \eta(\alpha_j \left\|\mathcal{R}_j^*(y_n) - \mathcal{R}_{j-1}^*(y_n)\right\|) \\ &\leq \varepsilon_n + F^*(p_0^*; y_n) - F^*(p_n^*; y_n) \\ &= \varepsilon_n + J^*(K^* p_0^*) - \langle p_0^*, y_n\rangle - (J^*(K^* p_n^*) - \langle p_n^*, y_n\rangle) \\ &= \varepsilon_n + J^*(K^* p_0^*) - \langle p_0^*, y\rangle - (J^*(K^* p_n^*) - \langle p_n^*, y\rangle) + \langle p_0^* - p_n^*, y_n - y\rangle \\ &\leq \varepsilon_n + F^*(p_0^*; y) - \mu^*(y) + \delta_n(\|p_0^*\| + \|p_n^*\|), \end{aligned}$$

where $\mu^*(y) = \inf\{F^*(q^*;y) \; : \; q^* \in Y^*\}$. Combining this with (2.33) gives after setting $\tilde{\varepsilon}_n := \|p_0^*\| \, \delta_n + \varepsilon_n$

$$J(x_n) \leq J(x) + F^*(p_0^*;y) - \mu^*(y) + 2\delta_n \|p_n^*\| + \tilde{\varepsilon}_n. \tag{2.35}$$

From Lemma 2.3.8 it follows that

$$\delta_n \|p_n^*\| \leq \delta_n \gamma(\delta_n) + \delta_n \zeta_n \psi_\phi^{-1}(t_{\nu(n)}(\boldsymbol{\alpha})),$$

where $\gamma(\delta_n)\delta_n \to 0$ as $n \to \infty$ and $\zeta_n = \mathcal{O}(1 + t_{\nu(n)}(\boldsymbol{\alpha})\psi_\phi(\delta_n))$. From Lemma 2.3.7 (1) it therefore follows that $\delta_n \|p_n^*\| = \mathcal{O}(t_{\nu(n)}(\boldsymbol{\alpha})\psi_\phi(\delta_n))$ and hence — provided that (2.30) holds — this already implies that

$$\limsup_{n\to\infty} J(x_n) \leq M < \infty$$

for a constant $M \in \mathbb{R}$. Furthermore, if (2.31) holds, then $\delta_n \|p_n^*\| \to 0$ as $n \to \infty$ and thus it follows from (2.35) that M can be chosen as

$$M = J(x) + F^*(p_0^*;y) - \mu^*(y) = \inf_{v\in X} \{J(v) \; : \; Kv = y\} + F^*(p_0^*;y) - \mu^*(y)$$

and (2.32) follows.

In particular, this implies that the sequence $\{J(x_n)\}_{n\in\mathbb{N}}$ is uniformly bounded and due to monotonicity (cf. Theorem 2.3.1) the same holds for $\{\psi_\phi(\|Kx_n - y_n\|)\}_{n\in\mathbb{N}}$. Therefore it follows that

$$\sup_{n\in\mathbb{N}} (\psi_\phi(\|Kx_n - y\|) + J(x_n)) \leq \sup_{n\in\mathbb{N}} (\psi_\phi(\|Kx_n - y_n\| + \delta_n) + J(x_n)) =: C < \infty.$$

In other words, for all $n \in \mathbb{N}$ one has $x_n \in \Lambda(C)$, which is a sequentially τ_X-compact set according to (R5).

Let $n \mapsto \rho(n)$ be a selection such that $\{x_{\rho(n)}\}_{n\in\mathbb{N}}$ is τ_X-convergent with limit $\hat{x}$. From lower semicontinuity (cf. Remark 2.1.2) and Theorem 2.3.4 we eventually conclude (by noting that $t_{\nu(\rho(n))}(\boldsymbol{\alpha}) \to \infty$ according to (2.31))

$$\begin{aligned} \|K\hat{x} - y\| &\leq \liminf_{n\to\infty} \left\|Kx_{\rho(n)} - y_{\rho(n)}\right\| \\ &\leq \lim_{n\to\infty} \psi_\phi^{-1}\left(\frac{F^*(p_0^*;y) - \mu^*(y)}{t_{\nu(\rho(n))}(\boldsymbol{\alpha})} + \psi(\delta_{\rho(n)})\right) = 0 \end{aligned} \tag{2.36}$$

or in other words $K\hat{x} = y$. □

We close this section remarking on a particular parameter choice rule Γ obtained by the so called *discrepancy principle*. Let $\tau > 1$ be a given constant and y as well as $\{y_n\}_{n\in\mathbb{N}}$ be as in Theorem 2.3.9. We define

$$\nu(n) := \min\{k \in \mathbb{N} \; : \; \psi_\phi(\|K\mathcal{R}_k(y_n) - y_n\|) \leq \tau\psi_\phi(\delta_n)\}. \tag{2.37}$$

The number $\nu(n)$ is well defined: From Theorem 2.3.4 it becomes clear that

$$\psi_\phi(\|K\mathcal{R}_k(y_n) - y_n\|) \leq \frac{F^*(p_0^*;y) - \mu^*(y)}{t_k(\boldsymbol{\alpha})} + \psi_\phi(\delta_n).$$

for all $k \in \mathbb{N}$. Since $t_k(\boldsymbol{\alpha}) \to \infty$ as $k \to \infty$ (cf. (2.3)) there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$

$$\frac{F^*(p_0^*; y) - \mu^*(y)}{t_k(\boldsymbol{\alpha})} \leq (\tau - 1)\psi_\phi(\delta_n).$$

Hence, $\nu(n)$ is well defined. In particular, we find that

$$\tau\psi_\phi(\delta_n) < \psi_\phi(\left\|K\mathcal{R}_{\nu(n)-1}(y_n) - y_n\right\|) \leq \frac{F^*(p_0^*; y) - \mu^*(y)}{t_{\nu(n)-1}(\boldsymbol{\alpha})} + \psi_\phi(\delta_n)$$

and in turn one finds

$$(\tau - 1)\psi_\phi(\delta_n)t_{\nu(n)}(\boldsymbol{\alpha}) - (\tau - 1)\frac{\psi_\phi(\delta_n)}{\alpha_{\nu(n)}} = (\tau - 1)\psi_\phi(\delta_n)t_{\nu(n)-1}(\boldsymbol{\alpha}) \leq F^*(p_0^*; y) - \mu^*(y).$$

This results in

$$\psi_\phi(\delta_n)t_{\nu(n)}(\boldsymbol{\alpha}) \leq \frac{F^*(p_0^*; y) - \mu^*(y)}{\tau - 1} + \frac{\psi_\phi(\delta_n)}{\alpha_{\nu(n)}}. \tag{2.38}$$

Hence the right hand side of (2.38) is bounded, when e.g. $\inf_{n\in\mathbb{N}} \alpha_n > 0$. As in the proof of Theorem 2.3.9 this implies that $\left\{\mathcal{R}_{\nu(n)}(y_n)\right\}_{n\in\mathbb{N}} \subset \Lambda(C)$ for a constant $C \in \mathbb{R}$. Moreover, the construction of $\nu(n)$ implies that

$$\lim_{n\to\infty} \left\|K\mathcal{R}_{\nu(n)}(y_n) - y\right\| \leq \lim_{n\to\infty} \left(\left\|K\mathcal{R}_{\nu(n)}(y_n) - y_n\right\| + \delta_n\right) \leq \lim_{n\to\infty} \tau\psi_\phi(\delta_n) + \delta_n = 0.$$

Thus the lower semicontinuity argument (2.36) is applicable. Note, that $\nu(n)$ not necessarily tends to ∞ as $n \to \infty$.

We summarize:

Corollary 2.3.10. Let y, $\{y_n\}_{n\in\mathbb{N}}$ and $\{\delta_n\}_{n\in\mathbb{N}}$ be as in Theorem 2.3.9 and assume additionally that $\tau > 1$ and

$$\inf_{n\in\mathbb{N}} \alpha_n > 0.$$

If $\Gamma_\tau : (0, \infty) \times Y \to \mathbb{N}$ is such that $\nu(n) = \Gamma(\delta_n, y_n)$ satisfies the discrepancy principle (2.37), then $\left(\{\mathcal{R}_n\}_{n\in\mathbb{N}}, \Gamma_\tau\right)$ is a regularization method for (2.4).

Remark 2.3.11.

1. The discrepancy principle (2.37) in the context of Algorithm 2.2.9 was introduced by Osher et al. in [103] motivated from iterative Tikhonov regularization (cf. [53]). It was shown in [103, Thm. 3.6] that the parameter choice in (2.37) yields a convergent regularization method, however, under the additional assumption that $n(\nu) \to \infty$ as $\nu \to \infty$. We note, that our analysis gets by without this requirement.

2. Estimate (2.38) gives rise to a (heuristic) variant of the discrepancy principle. Let $y \in Y$ be attainable and $\tilde{y} \in Y$ be a (noisy) approximation of y and assume that one can estimate $\delta > 0$ such that

$$\|y - \tilde{y}\| \lesssim \delta.$$

Then (2.38) suggests to use $\mathcal{R}_{n^*}(\tilde{y})$ as approximation for exact J-minimizing solutions of (2.4), where

$$n^* = \max\left\{n \in \mathbb{N} \ : \ t_n(\boldsymbol{\alpha}) \leq \frac{F^*(p_0^*; y) - \mu^*(y)}{(\tau - 1)\psi_\phi(\delta)}\right\}.$$

2.4 Convergence for Data in a Hilbert Space

In this section we consider the case, when Y is a Hilbert Space with inner product $\langle \cdot, \cdot \rangle_Y$. The main focus of this section lies on coming up with an improvement of the convergence results in Theorem 2.3.9. This will be achieved in Theorem 2.4.4. Moreover, we will present a convergence rates result for the augmented Lagrangian method in Theorem 2.4.6.

Throughout this section, we assume that $\phi(s) = s$ and note, that this implies that $\mathfrak{J}_\phi = \mathrm{Id}$. By means of Riesz' isomorphism we identify Y^* with Y and likewise the pairing $\langle \cdot, \cdot \rangle_{Y^*,Y}$ with the inner product and agree upon the notation

$$Y^* = Y \quad \text{and} \quad \langle p, y \rangle_{Y^*,Y} = \langle p, y \rangle_Y =: \langle p, y \rangle .$$

Recall that $\boldsymbol{\alpha} = \{\alpha_n\}_{n\in\mathbb{N}} \subset (0, \infty)$ and that $p_0 \in Y$ is chosen according to requirement (R6), that is, there exists $x_0 \in X$ such that $K^* p_0 \in \partial J(x_0)$.

In the present setting, the augmented Lagrangian method takes a convenient shape: Let $y \in Y$ and $\alpha > 0$. The augmented Lagrangian $L(\cdot, \cdot; y, \alpha) := L_{\mathrm{Id}}(\cdot, \cdot; y, \alpha) : X \times Y \to \overline{\mathbb{R}}$ can be written as

$$L(x, p; y, \alpha) = \frac{1}{2} \|Kx - (y + \alpha p)\|^2 + \alpha J(x) - \frac{\alpha^2}{2} \|p\|^2 . \tag{2.39}$$

Thus Algorithm 2.2.9 turns into

Algorithm 2.4.1 (Augmented Lagrangian method: Hilbert space version)**.** Let $y \in Y$. For $n = 1, 2, \ldots$ compute

$$x_n \in \operatorname*{argmin}_{v\in X} \frac{1}{2} \|Kv - (y + \alpha_n p_{n-1})\|^2 + \alpha_n J(v), \tag{2.40a}$$

$$p_n = p_{n-1} - \alpha_n^{-1}(Kx_n - y). \tag{2.40b}$$

As in Section 2.3 we set for $y \in Y$

$$\mathcal{R}_n(y) = x_n \quad \text{and} \quad \mathcal{R}_n^*(y) = p_n,$$

where $\{(x_n, p_n)\}_{n\in\mathbb{N}}$ denotes an arbitrary sequence generated by Algorithm 2.4.1 w.r.t. the data y, the initial value p_0 and the parameters $\boldsymbol{\alpha}$. Moreover, we again make use of (cf. (2.13))

$$F^*(p; y) = J^*(K^* p) - \langle p, y \rangle$$

and set $\mu^*(y) = \inf \{F^*(p; y) \ : \ p \in Y\}$. We recall that $\mu^*(y)$ is finite, whenever y is attainable (cf. Lemma 2.2.14).

The proximal point method w.r.t. $F^*(\cdot; y)$ (cf. Algorithm 2.2.16) in the Hilbert space setting comes as

Algorithm 2.4.2. Let $y \in Y$. For $n = 1, 2 \ldots$ compute

$$p_n = \operatorname*{argmin}_{q\in Y} \frac{\alpha_n}{2} \|q - p_{n-1}\|^2 + F^*(q; y). \tag{2.41}$$

Let $y \in Y$ and $\{p_n\}_{n\in\mathbb{N}}$ be the unique sequence generated by Algorithm 2.4.2. Then it follows from from Proposition 2.2.15 that

$$\mathcal{R}_n^*(y) = p_n$$

for all $n \in \mathbb{N}$. In other words, the dual sequence $\{\mathcal{R}_n^*(y)\}_{n\in\mathbb{N}}$ is uniquely defined.

We will now present the main convergence result for the Hilbert space setting. Before we do so, we cite a result by Güler established in [72], which will turn out to be extremely useful, in order to prove convergence of the augmented Lagrangian algorithm (cf. Theorem 2.4.4)

Lemma 2.4.3. For all $q \in Y$ the following estimates holds

$$F^*(\mathcal{R}_n^*(y);y) - F^*(q;y) \leq \frac{\|q-p_0\|^2}{2t_n(\boldsymbol{\alpha})} - \frac{\|q-\mathcal{R}_n^*(y)\|^2}{2t_n(\boldsymbol{\alpha})} - \frac{t_n(\boldsymbol{\alpha})}{2}\left\|\alpha_n(\mathcal{R}_n^*(y)-\mathcal{R}_{n-1}^*(y))\right\|^2.$$

Proof. [72, Lem. 2.2] □

Theorem 2.4.4. *Let $y \in Y$ be attainable and $\{y_n\}_{n\in\mathbb{N}} \subset Y$ such that $\delta_n := \|y-y_n\| \to 0$ as $n \to \infty$ and assume further that $\Gamma : (0,\infty)\times Y \to \mathbb{N}$ is such that*

$$\lim_{n\to\infty} \delta_n^2 t_{\Gamma(\delta_n,y_n)}(\boldsymbol{\alpha}) = 0 \quad \textit{and} \quad \lim_{n\to\infty} t_{\Gamma(\delta_n,y_n)}(\boldsymbol{\alpha}) = \infty. \tag{2.42}$$

Then $\left(\{\mathcal{R}_n\}_{n\in\mathbb{N}},\Gamma\right)$ is a regularization method for (2.4) such that

$$\lim_{n\to\infty} J(\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n)) = \inf_{v\in X}\{J(v) \ : \ Kv = y\}. \tag{2.43}$$

Proof. Let $\tilde{y} \in Y$ and set

$$\delta := \|y-\tilde{y}\|, \quad x_n := \mathcal{R}_n(\tilde{y}) \quad \text{and} \quad p_n := \mathcal{R}_n^*(\tilde{y}).$$

Moreover, assume that x is a J-minimizing solution of (2.4).

We will start with an estimate for the sequence $\{J(x_n)\}_{n\in\mathbb{N}}$. To this end, note that from (2.9b) it follows that $K^*p_n \in \partial J(x_n)$ for each $n \in \mathbb{N}$. Thus the definition of the subgradient implies

$$J(x_n) \leq J(x) + \langle K^*p_n, x_n - x\rangle = J(x) + \langle p_n, Kx_n - y\rangle. \tag{2.44}$$

First, by applying Lemma 2.4.3 we find for all $n \in \mathbb{N}$ and an arbitrary $q \in Y$ that

$$\begin{aligned}\frac{\|q-p_n\|^2}{2t_n(\boldsymbol{\alpha})} &\leq \frac{\|q-p_0\|^2}{2t_n(\boldsymbol{\alpha})} - \frac{t_n(\boldsymbol{\alpha})}{2}\|\alpha_n(p_n-p_{n-1})\|^2 + F^*(q;\tilde{y}) - F^*(p_n;\tilde{y}) \\ &\leq \frac{\|q-p_0\|^2}{2t_n(\boldsymbol{\alpha})} - \frac{t_n(\boldsymbol{\alpha})}{2}\|\alpha_n(p_n-p_{n-1})\|^2 \\ &\quad + F^*(q;y) - F^*(p_n;y) + \langle q-p_n, y-\tilde{y}\rangle.\end{aligned}$$

From Young's inequality we find for every $\zeta > 0$ that

$$\langle q-p_n, y-\tilde{y}\rangle \leq \frac{1}{2\zeta}\|q-p_n\|^2 + \frac{\zeta}{2}\delta^2.$$

We choose $\zeta = 2t_n(\boldsymbol{\alpha})$ and combine the previous two estimates with the fact that $F^*(p_n;y) \geq \mu^*(y)$ to

$$\begin{aligned}\frac{\|q-p_n\|^2}{4t_n(\boldsymbol{\alpha})} &\leq \frac{\|q-p_0\|^2}{2t_n(\boldsymbol{\alpha})} + t_n(\boldsymbol{\alpha})\delta^2 - \frac{t_n(\boldsymbol{\alpha})}{2}\|\alpha_n(p_n-p_{n-1})\|^2 + F^*(q;y) - F^*(p_n;y) \\ &\leq \frac{\|q-p_0\|^2}{2t_n(\boldsymbol{\alpha})} + t_n(\boldsymbol{\alpha})\delta^2 + F^*(q;y) - \mu^*(y).\end{aligned} \tag{2.45}$$

Let $\varepsilon > 0$ and choose an element $p_\varepsilon \in Y$ such that $F^*(p_\varepsilon; y) \leq \mu^*(y) + \varepsilon$. Then we conclude from (2.45) by setting $q = p_\varepsilon$ that

$$\frac{\|p_\varepsilon - p_n\|}{2\sqrt{t_n(\boldsymbol{\alpha})}} \leq \sqrt{\frac{\|p_\varepsilon - p_0\|^2}{2t_n(\boldsymbol{\alpha})} + t_n(\boldsymbol{\alpha})\delta^2 + \varepsilon}. \tag{2.46}$$

Next, after setting $c := F^*(p_0; y) - \mu^*(y)$, Theorem 2.3.4 shows that for each $n \in \mathbb{N}$.

$$\|Kx_n - y\| \leq \|Kx_n - \tilde{y}\| + \|y - \tilde{y}\| \leq \delta + \sqrt{\frac{2c}{t_n(\boldsymbol{\alpha})} + \delta^2}. \tag{2.47}$$

We eventually get the desired estimate: combining (2.44) with (2.46) and (2.47) results in

$$\begin{aligned} J(x_n) &\leq J(x) + \langle p_n - p_\varepsilon, Kx_n - y\rangle + \langle p_\varepsilon, Kx_n - y\rangle \\ &\leq J(x) + \frac{\|p_n - p_\varepsilon\|}{2\sqrt{t_n(\boldsymbol{\alpha})}} 2\sqrt{t_n(\boldsymbol{\alpha})}\, \|Kx_n - y\| + \|p_\varepsilon\|\, \|Kx_n - y\| \\ &\leq J(x) + \left(2\sqrt{t_n(\boldsymbol{\alpha})}\delta + 2\sqrt{2c + t_n(\boldsymbol{\alpha})\delta^2}\right) \sqrt{\frac{\|p^\varepsilon - p_0\|^2}{2t} + t_n(\boldsymbol{\alpha})\delta^2 + \varepsilon} \\ &\quad + \|p^\varepsilon\| \left(\delta + \sqrt{\frac{2c}{t_n(\boldsymbol{\alpha})} + \delta^2}\right). \end{aligned} \tag{2.48}$$

With this preparation we return to the prove of the assertion. For $n \in \mathbb{N}$ we set $\nu(n) = \Gamma(\delta_n, y_n)(y_n)$. Then it follows from (2.42) that

$$\lim_{n\to\infty} \delta^2_{\nu(n)} t_{\nu(n)}(\boldsymbol{\alpha}) = 0.$$

In turn, we find from (2.48) that

$$\limsup_{n\to\infty} J(\mathcal{R}_{\nu(n)}(y_n)) \leq J(x) + 2\sqrt{2c\varepsilon}.$$

Since $\varepsilon > 0$ was arbitrary this finally shows that

$$\limsup_{n\to\infty} J(\mathcal{R}_{\nu(n)}(y_n)) \leq J(x). \tag{2.49}$$

From the general result in Theorem 2.3.9 we already know that the sequence $\left\{\mathcal{R}_{\nu(n)}(y_n)\right\}_{n\in\mathbb{N}}$ is contained in a sequentially τ_X-compact set and that every τ_X-cluster point is a solution of (2.4). Thus, from Definition 2.2.4 and Remark 2.2.5 (1) it follows that each τ_X-cluster point of $\left\{\mathcal{R}_{\nu(n)}(y_n)\right\}_{n\in\mathbb{N}}$ is a J-minimizing solution (2.4).

In particular, for each subsequence of $\left\{\mathcal{R}_{\nu(n)}(y_n)\right\}_{n\in\mathbb{N}}$ there exists a further subsequence that converges to a J-minimizing solution $\hat{x}$ of (2.4) w.r.t. τ_X. Assume that $n \mapsto \rho(n)$ selects this subsequence and observe from the τ_X-sequential lower semicontinuity of J and (2.49) yield

$$J(x) \leq J(\hat{x}) \leq \liminf_{n\to\infty} J(\mathcal{R}_{\nu(\rho(n))}(y_{\rho(n)})) \leq \limsup_{n\to\infty} J(\mathcal{R}_{\nu(\rho(n))}(y_{\rho(n)})) \leq J(x)$$

Thus, each subsequence of $\left\{\mathcal{R}_{\nu(n)}(y_n)\right\}_{n\in\mathbb{N}}$ has in turn a subsequence such that (2.43) holds. Therefore (2.43) already holds for the whole sequence. □

Corollary 2.4.5. Let $y \in Y$ be attainable, $\tilde{y} \in Y$ and set $\delta := \|y - \tilde{y}\|$. Moreover, assume that x is a J-minimizing solution of (2.4), that satisfies the source condition (2.6) with source element p and set $\xi^* = K^*p$. Then the estimate

$$\left(D_J^{\xi^*}(\mathcal{R}_n(\tilde{y}), x)\right)^2 \leq 4\left(\frac{\|p_0 - p\|}{2t_n(\boldsymbol{\alpha})} + t_n(\boldsymbol{\alpha})\delta^2\right)\left(2(F^*(p_0; y) - \mu^*(y)) + t_n(\boldsymbol{\alpha})\delta^2\right). \qquad (2.50)$$

holds for all $n \in \mathbb{N}$.

Proof. Let $\{x_n\}_{n\in\mathbb{N}}$ and $\{p_n\}_{n\in\mathbb{N}}$ be as in the proof of Theorem 2.4.4. If p is a source element for x, it follows from Theorem 2.2.19 that

$$\mu^*(y) = \inf_{q \in y} F^*(q; y) = F^*(p; y).$$

Thus we deduce from (2.46) with $p_\varepsilon = p$ and $\varepsilon = 0$

$$\frac{\|p - p_n\|^2}{4t_n(\boldsymbol{\alpha})} \leq \frac{\|p - p_0\|^2}{2t_n(\boldsymbol{\alpha})} + t_n(\boldsymbol{\alpha})\delta^2.$$

From the update rule (2.9b) in the augmented Lagrangian method it follows that $K^*p_n \in \partial J(x_n)$. Thus the Bregman distance between x_n and x can be estimated by

$$\begin{aligned}D_J^{\xi^*}(x_n, x) &= J(x_n) - J(x) - \langle \xi^*, x_n - x\rangle \\ &\leq \langle K^*p_n, x_n - x\rangle - \langle K^*p, x_n - x\rangle = \langle p_n - p, Kx_n - y\rangle.\end{aligned}$$

Combination of the previous two estimates with Theorem 2.3.4 results in

$$\begin{aligned}\left(D_J^{K^*p}(x_n, x)\right)^2 &\leq \|p - p_n\|^2 \|Kx_n - y\|^2 \\ &= \frac{\|p - p_n\|^2}{4t_n(\boldsymbol{\alpha})} 4t_n(\boldsymbol{\alpha}) \|Kx_n - y\|^2 \\ &\leq 4\left(\frac{\|p - p_0\|^2}{2t_n(\boldsymbol{\alpha})} + t_n(\boldsymbol{\alpha})\delta^2\right)\left(2(F^*(p_0^*; y) - \mu^*(y)) + t_n(\boldsymbol{\alpha})\delta^2\right).\end{aligned}$$

□

The previous Corollary implies a convergence rates result for the augmented Lagrangian method (Algorithm (2.4.1)). To be more precise:

Theorem 2.4.6. *Let $y \in Y$ be attainable and $\{y_n\}_{n\in\mathbb{N}} \subset Y$ such that $\delta_n := \|y - y_n\| \to 0$ as $n \to \infty$. Assume further that there exist constants $0 < b \leq B$ and $\Gamma : (0, \infty) \times Y \to \mathbb{N}$ is such that*

$$b \leq \delta_n t_{\Gamma(\delta_n, y_n)}(\boldsymbol{\alpha}) \leq B, \qquad \text{for all } n \in \mathbb{N}. \qquad (2.51)$$

If x is a J-minimizing solution of (2.4) that satisfies the source condition (2.6) with source element $p \in Y$, then

$$D_J^{\xi^*}(\mathcal{R}_{\Gamma(\delta_n, y_n)}(y_n), x) = \mathcal{O}(\delta_n^{1/2}) \quad \text{and} \quad \left\|K\mathcal{R}_{\Gamma(\delta_n, y_n)}(y_n) - y\right\| = \mathcal{O}(\delta_n^{1/2}), \qquad (2.52)$$

where $\xi^* = K^*p$.

Proof. Let us agree upon the abbreviations

$$\nu(n) := \Gamma(\delta_n, y_n) \quad \text{and} \quad c := F^*(p_0^*; y) - \mu^*(y).$$

Then it follows from Corollary 2.4.5 and

$$\begin{aligned}\left(D_J^{\xi^*}(\mathcal{R}_{\nu(n)}(y_n), x)\right)^2 &\leq 4\left(\frac{\|p_0 - p\|}{2t_{\nu(n)}(\boldsymbol{\alpha})} + t_{\nu(n)}(\boldsymbol{\alpha})\delta_n^2\right)\left(2c + t_{\nu(n)}(\boldsymbol{\alpha})\delta_n^2\right) \\ &\leq 4\left(\frac{\|p_0 - p\|\,\delta_n}{2B} + b\delta_n\right)(2c + b\delta_n) = \mathcal{O}(\delta_n).\end{aligned}$$

The second inequality in (2.52) follows from Theorem 2.3.4: we have

$$\begin{aligned}\|K\mathcal{R}_n(y_n) - y\| &= \|K\mathcal{R}_n(y_n) - y_n\| + \|y_n - y\| \\ &\leq \sqrt{\frac{2c}{t_n(\boldsymbol{\alpha})} + \delta_n^2} + \delta_n \leq \sqrt{\frac{2c\delta_n}{B} + \delta_n^2} + \delta_n = \mathcal{O}(\delta_n^{1/2}).\end{aligned}$$

□

Remark 2.4.7. We note that the convergence rate in (2.52) in general is not optimal: For iterated Tikhonov regularization (cf. Section 2.5.1), for instance, it is well known, that the optimal rate is $\mathcal{O}(\delta)$ instead of $\mathcal{O}(\delta^{1/2})$. However, there may exists choices for J such that the estimate in (2.52) is sharp.

2.5 Example: Quadratic Regularization

In this section we will study Algorithm 2.4.1 for the special case, when J is a *quadratic functional* defined on a Hilbert space. In the follow-up paragraph we will clarify this notion:

Assume that for $i \in \{1, 2\}$, H_i are Hilbert spaces with inner product $\langle \cdot, \cdot \rangle_i$ and induced norm $\|\cdot\|_i$. Assume further that $L : D(L) \subset H_1 \to H_2$ is a linear and closed operator defined on the dense subset $D(L) \subset H_1$. It is not required that L is bounded and we use the symbol $\mathrm{Gr}(L)$ for the *graph* of L, that is,

$$\mathrm{Gr}(L) = \{(x_1, x_2) \in H_1 \times H_2 \;:\; L(x_1) = x_2\}\,.$$

We further recall that since $D(L)$ is assumed to be dense, there exists a linear operator $L^* : D(L^*) \subset H_2 \to H_1$, where

$$D(L^*) := \{x_2 \in H_2 \;:\; \text{ the mapping } x_1 \mapsto \langle Lx_1, x_2 \rangle_2 \text{ is continuous }\}$$

such that

$$\langle Lx_1, x_2 \rangle_2 = \langle x_1, L^* x_2 \rangle_1 \qquad \text{for all } x_1 \in D(L),\, x_2 \in D(L^*).$$

The operator L^* is called the *adjoint operator* of L (cf. [124, Chap. VII.2]). By a quadratic functional we refer to a functional $J : H_1 \to \overline{\mathbb{R}}$ defined by

$$J(x) = \begin{cases} \frac{1}{2}\|Lx\|_2^2 & \text{if } x \in D(L) \\ +\infty & \text{else.} \end{cases} \tag{2.53}$$

The subgradient of a quadratic functional admits a convenient characterization

Lemma 2.5.1. The functional J is proper, convex and lower semicontinuous and one has $D(\partial J) = D(L^*L)$ with

$$\partial J(x) = \begin{cases} L^*Lx & \text{if } x \in D(L^*L) \\ \emptyset & \text{else.} \end{cases}$$

Proof. Since L is densely defined, J is proper. Moreover, convexity follows from the linearity of L and the convexity of $\|\cdot\|_2^2$.

We prove lower semicontinuity. Assume that $\{x_n\}_{n\in\mathbb{N}} \subset H_1$ is a convergent sequence with limit x. If Lx_n converges to an element $y \in H_2$ we conclude from the closedness of L that $Lx = y$ and nothing remains to be shown. Therefore we can assume that Lx_n does not converge in H_2 and it is not restrictive to presume that $\{Lx_n\}_{n\in\mathbb{N}}$ has a bounded subsequence (otherwise $\liminf_{n\to\infty} J(x_n) = +\infty$ and nothing remains to show).

Since $\liminf_{n\to\infty} J(x_n) < \infty$ one can choose a selection $n \mapsto \rho(n)$ and an element $y \in H_2$ such that

$$Lx_{\rho(n)} \rightharpoonup y.$$

Since $\mathrm{Gr}(L) \subset H_1 \times H_2$ is closed and convex it is weakly closed and thus $x_{\rho(n)} \to x$ implies $Lx = y$. Weak lower semicontinuity of the norm eventually gives

$$J(x) = \frac{1}{2}\|Lx\|_2^2 \le \liminf_{n\to\infty} \frac{1}{2}\left\|Lx_{\rho(n)}\right\|_2^2.$$

It remains to show that $D(\partial J) = D(L^*L)$ and $\partial J(x) = L^*L(x)$ for $x \in D(L^*L)$. It is straight forward to check that $L^*Lx \in \partial J(x)$, whenever $x \in D(L^*L)$ or in other words that $\mathrm{Gr}(L^*L) \subset \partial J \subset H_1 \times H_1$. Moreover, one has that

$$\langle L^*Lx_1 - L^*Lx_2, x_1 - x_2\rangle = \|Lx_1 - Lx_2\|^2 \ge 0$$

for all $x_i \in D(L^*L)$ $(i = 1, 2)$, which means that $\mathrm{Gr}(L^*L)$ is a monotone subset of $H_1 \times H_1$. Since $D(L)$ was assumed to be dense, L^*L is densely defined and self-adjoint ([124, Cor. VII 2.13]) and therefore closed. This, however, is already sufficient for $\mathrm{Gr}(L^*L)$ to be maximal monotone (see e.g. [80, Chap.3 Thm.1.45]), that is, $\mathrm{Gr}(L^*L)$ is not properly contained in any monotone set in $H_1 \times H_2$. Since ∂J is (maximal) monotone and due to the fact that

$$L^*L \subset \partial J$$

this shows $L^*L = \partial J$ and the Lemma is shown. □

In order to synchronize the current setting with the general assumptions in Section 2.1 we set $X = H_1$ and assume that Y is another Hilbert space and we choose $\tau_X = \tau_X^w$ and $\tau_Y = \tau_Y^w$. Moreover, we assume that $K : X \to Y$ is a bounded and linear operator such that the sets

$$\Lambda(c) = \left\{x \in X \;:\; \|Kx\|^2 + \|Lx\|^2 \le c\right\}$$

are sequentially weakly compact (for every $c > 0$) and that $x_0 \in X$ and $p_0 \in Y$ are chosen such that

$$K^*p_0 \in \partial J(x_0) = L^*Lx_0.$$

The augmented Lagrangian method in the present setting comes as

Algorithm 2.5.2 (Augmented Lagrangian method: quadratic case)**.** Let $y \in Y$.

1. Set $\mathcal{R}_0(y) = x_0$.
2. For $n = 1, 2, \dots$ compute

$$\mathcal{R}_n(y) \in \operatorname*{argmin}_{v\in X} \|Kv - y\|^2 + \alpha_n \|L(v - \mathcal{R}_{n-1}(y))\|^2 . \tag{2.54}$$

Assuming that the above assumptions hold, the analysis in Section 2.5.2 shows that the augmented Lagrangian method (Algorithm 2.2.9) is well defined. From Theorem 2.4.4 we gain

Proposition 2.5.3. *Let $y \in Y$ be attainable and $\{y_n\}_{n\in\mathbb{N}} \subset Y$ be such that $\delta_n := \|y - y_n\| \to 0$ as $n \to \infty$. Further assume that $\Gamma : (0,\infty) \times Y \to \mathbb{N}$ is such that*

$$\lim_{n\to\infty} \delta_n^2 t_{\Gamma(\delta_n,y_n)}(\boldsymbol{\alpha}) = 0 \quad \textit{and} \quad \lim_{n\to\infty} t_{\Gamma(\delta_n,y_n)}(\boldsymbol{\alpha}) = \infty.$$

Then $\left(\{\mathcal{R}_n\}_{n\in\mathbb{N}}, \Gamma\right)$ is a regularization method for (2.4) and one has

$$\lim_{n\to\infty} D_J\left(\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n), x\right) = \lim_{n\to\infty} \left\|L\left(\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n) - x\right)\right\|_2^2 = 0$$

for all J-minimizing solutions x of (2.4).

Proof. Let x be an arbitrary J-minimizing solution of (2.4) and abbreviate

$$\nu(n) = \Gamma(\delta_n, y_n) \quad \text{and} \quad \hat{x}_n := \mathcal{R}_{\nu(n)}(y_n).$$

From Theorem 2.4.4 we already know that

$$\lim_{n\to\infty} \|L\hat{x}_n\| = \|Lx\| . \tag{2.55}$$

In particular, the set $\{\|L\hat{x}_n\|\}_{n\in\mathbb{N}}$ is bounded and consequently there exists a selection $n \mapsto \rho(n)$ and $y \in H_2$ such that

$$w\text{-}\lim_{n\to\infty} L\hat{x}_{\rho(n)} = y.$$

Again from Theorem 2.4.4 we conclude that we can extract a further subsequence (which we again indicate by ρ) such that $\hat{x}_{\rho(n)} \rightharpoonup \hat{x}$ and $\hat{x}$ is a J-minimizing solution if (2.4). Since $\mathrm{Gr}(L) \subset X \times H_2$ is closed and convex it is weakly-weakly closed and we conclude that $y = L\hat{x} = Lx$. Since we can proceed with every subsequence in analogous manner, we conclude that

$$w\text{-}\lim_{n\to\infty} L\hat{x}_n = Lx. \tag{2.56}$$

Finally, note that for all $x, v \in X$

$$J'(x)(v) = \langle Lx, Lv\rangle$$

and therefore

$$\begin{aligned} D_J(\hat{x}_n, x) &= J(\hat{x}_n) - J(x) - J'(x)(\hat{x}_n - x) \\ &= \frac{1}{2}\|L\hat{x}_n\|^2 - \frac{1}{2}\|Lx\|^2 - \langle Lx, L\hat{x}_n - Lx\rangle = \frac{1}{2}\|L\hat{x}_n - Lx\|^2 . \end{aligned}$$

Taking the limit for $n \to \infty$ shows together with (2.55) and (2.56) the desired result. □

For the remainder of this Section we will study two special cases that emerge from particular choices for K and L.

2.5.1 Iterated Tikhonov Regularization

For the special choice $L \equiv \mathrm{Id}$ and $x_0 = 0$, Algorithm 2.5.2 turns to a classical regularization method often referred to as *iterated Tikhonov regularization* (for a rigorous treatment see Engl et al. [53, pp. 123]).

Let $y \in Y$. Since $J(x) = \frac{1}{2}\|x\|^2$ is strictly convex on X, there exists a unique J-minimizing solution of (2.4), whenever $y \in \mathrm{ran}(K)$ and likewise each iteration step of Algorithm 2.5.2 generates a unique element $\mathcal{R}_n(y)$. From obvious reasons, J-minimizing solutions in this context are referred to as *minimum norm solutions.*

Furthermore, it is evident that (2.54) is equivalent to

$$(\alpha_n \mathrm{Id} + K^*K)\mathcal{R}_n(y) = K^*y + \alpha_n \mathcal{R}_{n-1}(y).$$

From the spectral theorem, it follows that we can condense this equation to

$$\mathcal{R}_n(y) = (g_n(K^*K) \circ K^*)(y) \tag{2.57}$$

where $\{g_n : [0,\infty) \to \mathbb{R}\}_{n\in\mathbb{N}}$ is a sequence of real valued functions defined recursively by

$$\begin{aligned} g_0(t) &= 0, \\ g_n(t) &= \frac{1 + \alpha_n g_{n-1}(t)}{\alpha_n + t}, \qquad \text{for } n \geq 1. \end{aligned} \tag{2.58}$$

There exists a well-established theory for studying convergence for iterations of type (2.57), which is collected e.g. in the book of Engl et al. [53, Chap.4]. From Theorems 4.1 and 4.2 therein we find

Proposition 2.5.4. *Let y, $\{y_n\}_{n\in\mathbb{N}}$ and $\{\delta_n\}_{n\in\mathbb{N}}$ be as in Proposition 2.5.3 and assume that x is the unique J-minimizing solution of* (2.4) *w.r.t. a given right hand side $y \in \mathrm{ran}(K)$. For $n \in \mathbb{N}$ we assume*

$$G_n := \sup\left\{|g_n(t)| \ : \ 0 \leq t \leq \|K\|^2\right\} < \infty.$$

and

$$\lim_{n\to\infty} g_n(t) = \frac{1}{t} \quad \textit{and} \quad \lim_{n\to\infty} \delta_n = 0.$$

If $\Gamma : (0,\infty) \times Y \to \mathbb{N}$ is a parameter choice rule such that

$$\lim_{n\to\infty} G_{\Gamma(\delta_n, y_n)} \delta_n^2 = 0$$

then $\mathcal{R}_{\Gamma(\delta_n,y_n)}(y_n)$ converges to x.

We remark, that the symbol $\|K\|$ in Proposition 2.5.4 denotes the operator norm of K, that is,

$$\|K\| := \sup_{x\in X\setminus\{0\}} \frac{\|Kx\|}{\|x\|}.$$

It was shown by Hanke & Groetsch [76] (see also Brill & Schock [28]) that in the case under consideration, that is, for g_n given by (2.58)

$$\lim_{n\to\infty} g_n(t) = 1/t \quad \text{and} \quad G_n = t_n(\boldsymbol{\alpha})$$

provided that

$$\lim_{n\to\infty} t_n(\boldsymbol{\alpha}) = \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{\alpha_j} = \infty.$$

Thus the general assumptions in Proposition 2.5.4 reduce to the requirements in Proposition 2.5.3, indicating that our result is consistent with the classical theory.

2.5.2 Tikhonov – Morozov Regularization

The second special case we are focusing on results from the setting $K \equiv \mathrm{Id}$ and $p_0 = 0$. The method is referred to as *Tikhonov – Morozov regularization* and was e.g. intensively studied by Groetsch in the recent monograph [69].

Let $x \in X$. Since K equals the identity operator, we will assume that $X = Y$. Moreover, we remark that for a given $x \in X$ the mapping $v \mapsto \frac{1}{2}\|x - v\|^2$ is strictly convex, as a result of which Algorithm 2.5.2 generates a unique sequence $\{\mathcal{R}_n(x)\}_{n\in\mathbb{N}}$ that is characterized by the equation

$$(\alpha_n L^* L + \mathrm{Id})\mathcal{R}_n(y) = x + \alpha_n L^* L \mathcal{R}_{n-1}(y).$$

Similar as in the case of iterated Tikhonov regularization in the previous section we can give a explicit formula for the solution $\mathcal{R}_n(x)$ by

$$\mathcal{R}_n(x) = \tilde{L} h_n(\tilde{L}) x, \tag{2.59}$$

where $\tilde{L} = (\mathrm{Id} + L^* L)^{-1}$ and[2] h_n is defined iteratively by

$$\begin{aligned} h_0(t) &= 0, \\ h_n(t) &= \frac{1 + \alpha_n (1-t) h_{n-1}(t)}{\alpha_n + (1-\alpha_n) t}, \qquad \text{for } n \geq 1. \end{aligned} \tag{2.60}$$

Again, iterative schemes of the form (2.59) are well investigated and we refer e.g. to Groetsch's monograph [69] for an excellent treatise on the topic. Theorem 3.4 therein states

Theorem 2.5.5. *Let $x \in D(L)$, $\tilde{x} \in X$ and define $\|x - x^\delta\| =: \delta$. For $n \in \mathbb{N}$ assume that $h_n \in C([0,1])$ satisfies*

$$\lim_{n\to\infty} h_n(t) = \frac{1}{t} \qquad \textit{for all } t \in (0,1]$$

and that $|t h_n(t)|$ is uniformly bounded on $[0,1]$ for all $n \in \mathbb{N}$. If $\{r_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ can be chosen such that $|(1-t)h_n(t)| \leq r_n$ for all $n \in \mathbb{N}$, then

$$\|L\mathcal{R}_n(\tilde{x}) - Lx\| \leq \delta \mathcal{O}(\sqrt{r_n}).$$

It was shown in [69, Chap. 4.4] that, whenever $\alpha_n \in (0,1]$ for all $n \in \mathbb{N}$, the assumptions in Theorem 2.5.5 are met with

$$r_n = t_n(\boldsymbol{\alpha}).$$

We point out that this boundedness condition for $\{\alpha_n\}_{n\in\mathbb{N}}$ is superfluous according to Proposition 2.5.3.

[2] The operator $\tilde{L}$ is well defined, linear and bounded according to von Neumann's theorem and its spectrum is contained in $[0,1]$ (cf. [101, pp.301] for the pioneering work).

Remark 2.5.6. The major purpose of the Tikhonov – Morozov method is the stable evaluation of the operator L. To be more precise, assume that for an (in general unknown) element $x \in D(L)$ a (in general noisy) approximation $\tilde{x} \in H_1$ is given.

When trying to approximate Lx by $L\tilde{x}$ one faces two difficulties: either $\tilde{x} \notin D(L)$ and thus $L\tilde{x}$ is not defined, or $L\tilde{x}$ is defined but unstable due to the discontinuity of L. From the viewpoint of Proposition 2.5.3, however, $L\mathcal{R}_n(\tilde{x})$ for an appropriately chosen stopping index n (for instance given by the discrepancy principle (2.37)) is supposed to stably approximate Lx.

We note that for a linear and compact operator $K : H_2 \to H_1$ the *Moore – Penrose inverse* $K^\dagger$ (cf. [53, Def. 2.2] for a Definition) is densely defined, linear and closed and — provided $\mathrm{ran}(K)$ is of infinite dimension — unbounded. Since the stable evaluation of $K^\dagger$ corresponds to computing a best-approximate solution of (2.4), the Tikhonov – Morozov method can be considered as a regularization method in the sense of Definition 2.2.4.

2.6 Notes

The augmented Lagrangian method was originally introduced independently by Hestenes [77] and Powell [107] (who used the notion *method of multipliers*) as a solution method for the constrained optimization problem

$$F(v) \to \inf! \quad v \in V \qquad \text{subject to} \qquad \psi_j(v) = 0,\ j \in \{1, \dots, n\}\,, \tag{$*$}$$

where F and ψ_j are given functions defined on a space V $(= \mathbb{R}^N)$. It can be considered as a hybrid method of two techniques available by then: the ordinary Lagrangian method and the slacked unconstrained minimization (or penalty) technique introduced shortly before by Fiacco & McCormick in [58]. The new method, however, does not inherit the major disadvantages of the latter, that is, slow convergence for the first and numerical instability (as the penalty parameter tends to 0) for the second.

Shortly after that the method was further developed and generalized by numerous authors, among them Miele et al. [96, 97, 98], Bertsekas [22, 23], Kort & Bertsekas [88] or Rockafellar [112, 113] who generalized the technique for inequality constraints. Further research on combination of inequality and equality constraints as well as application of the augmented Lagrangian method for variational inequalities has been conducted by Ito & Kunisch in [82, 83]. For additional references we also recommend the textbook by Bertsekas [24].

Our version of the augmented Lagrangian method as it comes in Algorithm 2.2.9 fits best to the work of Glowinski & Marrocco in [67] (or alternatively Fortin [59]). Therein the authors focus on the unconstrained problem

$$(F \circ B)(v) + G(v) \to \inf! \quad v \in V \tag{$**$}$$

where B is a linear operator and F and G are functionals defined on appropriate spaces. By means of an auxiliary variable $y = B(v)$ problem $(**)$ is subsequently transformed into $(*)$[3]. This technique is referred to as the decomposition-coordination method, an exhaustive analysis of which can be found in the textbooks by Fortin & Glowinski [60] or Glowinski [66]. The method was further studied by Ito & Kunisch in [85] for solving $(**)$ with additional constraint $v \in K$, $K \subset V$ as well as in [84] for tackling a image restoration problem.

[3] Clearly we arrive at problem (2.5) by setting F to $\chi_{\{y\}}$, the characteristic function of the set $\{y\}$.

Proposition 2.2.19 shows that the constrained problem (2.5) stands in duality to (2.17) a unconstrained problem on the dual space. We refer to the excellent textbook by Ekeland & Temam [51] for a thorough discussion on duality in convex programming. The augmented Lagrangian algorithm is likewise related to the *proximal point algorithm* as stated in Proposition 2.2.15.

This method was first introduced by Martinet in [93] — using the term successive approximation (fr. *approximation successive*) — in order to solve unconstrained minimization problems. The currently used notion goes back to the paper of Moreau [100], who introduced the resolvent operator (cf. Definition 2.2.17) as proximal point (fr. *point proximal*). In [115] Rockafellar established a rigorous mathematical background and generalized the method for finding zeros of monotone operators in Hilbert spaces. In much the same spirit is the work in [114] by the same author, coping with augmented Lagrangian and proximal point methods and their mutual relations.

Aside to [115], the work of Brézis and Lions in [27] is probably the first rigorous analysis of the proximal point algorithm. More recently, Güler [72, 73] studied global convergence rates and constructed a situation where the trajectory of the proximal point algorithm converges weakly but not in norm (see also Bauschke et al. [19]).

As we will see in the upcoming chapter, there exists a strong relation between iterative algorithms of proximal point type and abstract evolution equations. Therefore we refer to Section 3.7 for a supplementary list of references on that topic.

In view of Remark 2.2.10 the augmented Lagrangian method as it is in Algorithm 2.2.9 can itself be considered as a proximal point like algorithm, where proximity is understood rather with respect to the Bregman distance than to a norm. Under the notion *proximal minimization algorithm with D-functions* algorithms of this type were introduced by Censor & Zenios [37] and thereupon studied by Chen & Teboulle [40, 41], Eckstein [50] and Kiwiel [87] to name but a few.

Shortly afterward the method of iterative minimization using the Bregman distance was adapted to the general (infinite dimensional) Banach space setting e.g. by Alber & Butnariu [3], Alber et al. [5], Burachik & Iusem [29] or Butnariu & Iusem [35]. We refer to Iusem [86] for an overview article that contains a detailed list of references.

From on inverse problem point of view, the work by Krasnosel′skiĭ in [89] is probably the first approach to solve the linear ill-posed operator equation (2.4) with iterative methods of proximal point type. A further development of this idea by Krjanev in [90] led to a convergence analysis for a class of regularization methods that are covered by the analysis in Section 2.5. We mention that Lardy [91] studied iterated Tikhonov regularization as introduced in Section 2.5.1 for the special choice $\alpha_n = 1$ for all $n \in \mathbb{N}$. Therefore iterated Tikhonov regularization with this special parameter choice is referred to as *Lardy's method*.

For a more recent treatise on *nonstationary* iterated Tikhonov regularization (that is for nonconstant α_n) we mention Brill & Schock [28] and Groetsch & Hanke [76]. Tikhonov – Morozov regularization as presented in Section 2.5.2 was studied by Groetsch & Scherzer [70] and Groetsch in [68] and [69].

We finally mention that the definition of a regularization method (as in Definition 2.2.4) is motivated from the classical textbook of Tikhonov & Arsenin [121]. For a more recent reference on regularization methods for ill-posed operator equations we refer to Engl et al. [53].

3 Evolution Equations

In Chapter 2 we reviewed the augmented Lagrangian method on Banach spaces, as a technique for computing (regularized) solutions of the constrained minimization problem (2.5), that is

$$J(x) \to \inf! \quad \text{subject to} \quad Kx = y.$$

We pointed out (cf. Remark 2.2.10) that algorithms of this type can be interpreted (by adding constant terms to the objective functional) as a generalized proximal point algorithm (cf. Definition 2.2.17), where the distance between iterates is measured by means of the Bregman distance (w.r.t. to J) rather than by the Banach space norm.

In semi group theory (see e.g. Brézis [26] or Barbu [17]) it is well known that the proximal point algorithm forms an implicit time scheme for abstract evolution equations, where the regularization parameter serves as discrete time step size. In this chapter we will address the issue whether or not the augmented Lagrangian method — when viewed as a generalized proximal point algorithm — can be considered as a numerical scheme as well. We recall the (formal) argumentation in the introductory example in Chapter 1:

Assume that $f \in \mathrm{L}^2(\Omega)$ is a noisy capture of a (unknown) image $u \in U \subset \mathrm{L}^2(\Omega)$ ($\Omega \subset \mathbb{R}^2$ is the image domain). We shall consider the model

$$f = Ku + v$$

where $K : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ is a convolution operator and $v \in \mathrm{L}^2(\Omega)$ is assumed to be i.i.d. noise. Retrieving the true image u from measurements f is in general an ill-posed problem (since K is compact). In Chapter 2 we showed that the augmented Lagrangian Algorithm 2.2.9 constitutes a method for computing a sequence of stable approximations $\{u_n\}_{n\in\mathbb{N}}$ of u.

Assume that $J : U \to \mathbb{R}$ is a given (regularizing) functional. For the particular weight function $\phi(s) = s$ and the parameters $\alpha_n = \alpha > 0$ for $n \in \mathbb{N}$ the algorithm then reads as

1. Choose $v_0 \in \mathrm{L}^2(\Omega)$.
2. For $n = 1, 2, \dots$ compute

$$\begin{aligned} u_{n+1} &= \operatorname*{argmin}_{u\in U} \frac{1}{2}\int_\Omega |Ku - f|^2 \,\mathrm{d}x + \alpha\left(J(u) - \int_\Omega v_n(Ku - f)\,\mathrm{d}x\right), && (3.1\mathrm{a})\\ v_{n+1} &= v_n + \alpha^{-1}(f - Ku_{n+1}). && (3.1\mathrm{b}) \end{aligned}$$

Assume that u_n is well defined and that J is differentiable at u_n for all $n \in \mathbb{N}$ (with derivative $\partial J(u_n)$). Then the optimality condition for (3.1a) reads as

$$0 = K^*(Ku_{n+1} - f) + \alpha\partial J(u_{n+1}) - \alpha K^* v_n$$

and thus it follows together with (3.1b) that

$$K^* v_{n+1} = K^* v_n + \alpha^{-1} K^*(f - Ku_{n+1}) = \partial J(u_{n+1}). \tag{3.2}$$

Therefore we also have $K^*v_n = \partial J(u_n)$ and we finally observe that

$$\frac{\partial J(u_{n+1}) - \partial J(u_n)}{\alpha^{-1}} = K^*(f - Ku_{n+1}).$$

By setting $\Delta t = \alpha^{-1}$ we argue as in Chapter 1 that this can be considered as an implicit time step (at time $t = n\Delta t$) of the equation

$$\begin{aligned} \frac{d}{dt}\partial J(u) &= K^*(f - Ku), \\ u(0) &= u_0 \end{aligned}$$

for a suitably chosen initial value u_0. A popular example for J that satisfies the smoothness condition is (cf. Table 1.1 and Table 1.2).

$$J(u) = \int_\Omega |\nabla u|^p \,\mathrm{d}x \quad \text{and} \quad \partial J(u) = -\mathrm{div}\left(|\nabla u|^{p-2}\,\nabla u\right), \quad \text{for } 1 < p < \infty.$$

In case that J is nonsmooth but *subdifferentiable* at u_n (cf. Definition A.2.1) above argumentation turns rather problematic for $\partial J(u_n)$ in general being a *set*. A typical example (especially in image processing) is the BV-seminorm

$$J(u) \sim \int_\Omega |\nabla u| \,\mathrm{d}x \quad \text{and} \quad \partial J(u) \sim -\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

Here $\sim$ indicates that the above relations are purely formal. We remark, that this particular choice for J will be studied in detail in Chapter 4. In the nonsmooth setting (3.2) has to be rewritten to

$$K^*v_{n-1} = K^*v_n + \alpha^{-1}K^*(f - Ku_{n+1}) \in \partial J(u_{n+1}).$$

Together with (3.1b) we obtain (again by setting $\Delta t = \alpha^{-1}$)

$$\frac{v_{n+1} - v_n}{\Delta t} = f - Ku_{n+1} \quad \text{and} \quad K^*v_{n+1} \in \partial J(u_{n+1}).$$

This is an implicit scheme for the *coupled* evolution equation

$$\frac{d}{dt}v = f - Ku \quad \text{and} \quad K^*v \in \partial J(u). \tag{3.3}$$

Equations of this type will be the basic object of study in this chapter. In particular, we are interested in existence of solutions and their regularizing properties w.r.t. the ill-posed operator equation 2.4.

This chapter is organized as follows: In Section 3.1 we will, aside to general assumptions and notation, clarify how (3.3) can be generalized in order to fit into the general Banach space used in Chapter 2 (cf. Equation (3.4)) and moreover, give the definition of the corresponding notion of solution (cf. Definition 3.1.1).

We will construct such solutions by considering the augmented Lagrangian method as (implicit) time scheme: For a given sequence of regularization parameters, the augmented Lagrangian Algorithm 2.2.9 generates a pair of sequences, in the following often imprecisely referred to as the primal and the dual sequence. We consider these sequences to be samples of curves, where the sampling points are determined by the regularization parameter.

Keeping in mind the dual representation of the augmented Lagrangian method (cf. Proposition 2.2.15) we shall first focus on the dual sequence generated by the proximal point Algorithm 2.2.16. Using the analysis in the recent book by Ambrosio et al. [9], we will prove in Section 3.3 that *piecewise affine interpolations* of the *dual variables* pointwise converge to an absolute continuous function on the real halfline with values in Y^*, as the maximal step size tends to zero. It will turn out that this function already solves a gradient flow equation (cf. (3.18)) w.r.t. to the function F^* defined as in (2.13).

Likewise, we will prove convergence of *piecewise constant interpolations* of the *primal variable* to a function on $[0, \infty)$ as the time discretization is getting finer. With the dual solution already at hand, we will show that this function is a desired solution (Section 3.3).

In Section 3.4 we clarify how solutions of (3.4) can be considered as regularizing operators for (2.4) and we investigate preliminary asymptotic properties. These results are improved in Section 3.5, where we assume that the right hand side of the operator equation (2.4) lies in a Hilbert space. For this particular case we will also give an estimate for the approximation error of the augmented Lagrangian method interpreted as implicit time scheme.

In Section 3.6 we revisit the example of quadratic regularization studied previously in Section 2.5 and we close this chapter with some notes and suggestions for further reading in Section 3.7.

3.1 Assumptions and Notation

We assume that X and Y are Banach spaces, ϕ is a weight function and $K : X \to Y$ is a bounded and linear operator with adjoint K^*. Further we assume that $J : X \to \overline{\mathbb{R}}$ is a proper and convex functional. Given $y \in Y$ and $p_0^* \in Y^*$, we will focus on differential inclusions of the following type:

$$dp^*(t) \in \mathfrak{J}_\phi\,(y - Kx(t)) \qquad\qquad K^*p^*(t) \in \partial J(x(t)) \tag{3.4a}$$

$$p^*(0) = p_0^*. \tag{3.4b}$$

We note that $x : [0, \infty) \to X$ and $p^* : [0, \infty) \to Y^*$ are curves with trajectories in X and Y^* respectively and that $dp^* : [0, \infty) \to Y^*$ denotes the strong derivative of p^*, that is

$$dp^*(t) = \lim_{h \to 0} \frac{p^*(t + h) - p^*(t)}{h}.$$

Before we introduce the corresponding notion of solution, recall the definition of the Bregman topology τ_X^J on X in Definition A.2.7. Moreover, for $T > 0$ we denote by $C_J([0, T], X)$ the collection of all sequentially τ_X^J-continuous mappings $x : [0, T] \to X$.

Definition 3.1.1. A pair of functions $(x, p^*) : [0, \infty) \to X \times Y^*$ is called *a solution of* (3.4) if $x(t)$ and $p^*(t)$ satisfy (3.4a) for a.e. $t \geq 0$ and if

$$x \in C_J([0, T]; X) \quad \text{ and } \quad p^* \in C([0, T]; Y^*) \cap \mathrm{W}^{1,1}(0, T; Y^*)$$

for every $T > 0$. Moreover, p^* satisfies (3.4b) in the sense that

$$\lim_{t \to 0^+} p^*(t) = p_0^*.$$

Additionally to the Assumption 2.1.1 in Chapter 2, we will force two more assumptions, one of which is based on the following

Definition 3.1.2. Let Z be a topological space equipped with a topology τ and assume that $\Phi : Z \to \overline{\mathbb{R}}$ is a convex and proper functional. We say, that the subgradient $\partial\Phi \subset Z \times Z^*$ is τ-weakly* closed, if the following implication holds

$$\left.\begin{array}{l} z_n^* \in \partial\Phi(z_n), \quad \sup_{n\in\mathbb{N}} \Phi(z_n) < \infty, \\ \lim_{n\to\infty} z_n = z \ \text{ w.r.t. } \tau, \quad w^*\text{-}\lim_{n\to\infty} z_n^* = z^* \end{array}\right\} \Rightarrow z^* \in \partial\Phi(z).$$

Remark 3.1.3. Let Z and Φ be as in Definition 3.1.2. If $\partial\Phi \subset Z \times Z^*$ is closed w.r.t. the product topology of τ and $\tau_Z^{w^*}$, then it is already τ-weakly* closed. However, the converse implication in general fails to hold.

In particular, if τ coincides with the strong topology on Z, then ∂Z is τ-weakly* closed (cf. Lemma A.2.2 (1)).

Let us assume that Assumption 2.1.1 holds and that topologies τ_X and τ_Y on X and Y are chosen accordingly.

Assumption 3.1.4. R7. **(Reflexivity)** Y is reflexive.

R8. **(Closedness)**

a) The set $\partial J \subset X \times X^*$ is τ_X-weakly* closed.

b) The set $\partial(J^* \circ K^*) \subset Y^* \times Y^{**}$ is weakly*-weakly*.

Unless stated differently, we will from now on assume that Assumptions 2.1.1 and 3.1.4 hold.

Remark 3.1.5. We summarize some implications of (R7). A collection of these (and related) results can be found in [94, Chap. 1.11, Chap. 1.13].

We recall that Y is said to be *reflexive* if the natural mapping $i_Y : Y \to Y^{**}$ is an isometric isomorphism. We will therefore identify Y with its bidual Y^{**} and define

$$\langle p^*, y\rangle := \langle p^*, y\rangle_{Y^*,Y} = \langle i_Y(y), p^*\rangle_{Y^{**},Y^*}\,.$$

Moreover, one has that Y is reflexive if and only if Y^* is reflexive and that the weak and weak* topologies on Y^* coincide. From this point of view we can rewrite (R8b) to

The set $\partial(J^* \circ K^*) \subset Y^* \times Y$ is weakly-weakly closed.

Finally we note that every norm bounded sequence in a reflexive Banach space has a weakly convergent subsequence or in other words, norm bounded sets in reflexive Banach spaces are sequentially weakly precompact.

3.2 Interpolation of Discrete Solutions

In this section we will consider interpolations of the primal and dual sequences generated by the augmented Lagrangian method (Algorithm 2.2.9) over the real halfline $[0, \infty)$. Here the interpolation nodes (or sampling points) are determined by the sequence of regularization parameters used in Algorithm 2.2.9. The interpolation functions introduced below will be used in Section 3.3 in order to construct solutions of Equation (3.4) (by successively refining the density of sampling points).

Throughout this section, let $y \in Y$ be fixed and recall that $p_0^* \in Y^*$ was chosen according to assumption (R6), that is, there exists $x_0 \in X$ such that

$$K^* p_0^* \in \partial J(x_0).$$

Further, we assume that $\boldsymbol{\alpha} := \{\alpha_n\}_{n\in\mathbb{N}}$ is a given sequence of positive parameters and that

$$\boldsymbol{x} := \{x_n\}_{n\in\mathbb{N}} \subset X \quad \text{ and } \quad \boldsymbol{p}^* := \{p_n^*\}_{n\in\mathbb{N}} \subset Y^*$$

are sequences generated by the augmented Lagrangian Algorithm 2.2.9 w.r.t. the data y, the initial value p_0^* and the parameters $\boldsymbol{\alpha}$.

We will study interpolations of the values $\{x_0, x_1, \ldots\}$ and $\{p_0^*, p_1^*, \ldots\}$ over the half line $[0, \infty)$, where we assume that for each $n \in \mathbb{N}$ the values x_n and p_n^* are given at the sampling point

$$t_n(\boldsymbol{\alpha}) := \sum_{i=1}^{n} \frac{1}{\alpha_i}. \tag{3.5}$$

Special emphasis will be put on interpolation of the dual sequence $\boldsymbol{p}^*$. Therefore we recall the definition of $F^*(\cdot; y) : Y^* \to \overline{\mathbb{R}}$ (as introduced in (2.13)), that is

$$F^*(q^*; y) = J^*(K^* q^*) - \langle q^*, y \rangle .$$

Then, according to Proposition 2.2.15, the sequence $\boldsymbol{p}^*$ is characterized by the proximal point algorithm w.r.t. $F^*(\cdot; y)$. (cf. Algorithm 2.2.16), that is, for all $n \in \mathbb{N}$

$$p_n^* \in \operatorname*{argmin}_{q^* \in Y^*} \alpha_n^{-1} \psi_{\phi^{-1}} \left(\alpha_n \left\| q^* - p_{n-1}^* \right\| \right) + F^*(q^*; y). \tag{3.6}$$

Since the dependence of $F^*(\cdot; y)$ on y is not relevant in this Section (y was assumed to be fixed) we will agree upon the abbreviation $F^*(q^*) = F^*(q^*; y)$ for $q^* \in Y^*$.

Finally, we note that Appendix A.3 collects some technical results concerning the proximal point algorithm, which will frequently be used in this Section. For example, it follows from Lemma A.3.2 that for $\alpha > 0$ the resolvent operator $R_{F^*}^{\alpha}$ (as defined in Remark 2.41) is well defined, that is, for each given element $p^* \in Y^*$ the sets

$$R_{F^*}^{\alpha}(p^*) = \operatorname*{argmin}_{q^* \in Y^*} \alpha^{-1} \psi_{\phi^{-1}} \left(\alpha \left\| q^* - p^* \right\| \right) + F^*(q^*)$$

are nonempty. With this notation it follows from (3.6) that $p_n^* \in R_{F^*}^{\alpha_n}(p_{n-1}^*)$ for $n \geq 1$.

Definition 3.2.1. A sequence of positive numbers $\boldsymbol{\alpha} := \{\alpha_n\}_{n\in\mathbb{N}} \subset (0, \infty)$ is called *a partition of* $[0, \infty)$ if the sequence $\{t_n(\boldsymbol{\alpha})\}_{n\in\mathbb{N}}$ as defined in (3.5) is unbounded. For $n \geq 0$ we call the elements $t_n(\boldsymbol{\alpha})$ *sampling points* (subordinate to the partition $\boldsymbol{\alpha}$), where we set $t_0(\boldsymbol{\alpha}) := 0$. The quantity

$$|\boldsymbol{\alpha}| := \inf \{\alpha_j \ : \ j \in \mathbb{N}\} . \tag{3.7}$$

is called *density of the partition* $\boldsymbol{\alpha}$.

From now on we shall assume that $\boldsymbol{\alpha}$ is a partition of $[0,\infty)$. We move on to the definition of the announced interpolation techniques. We use piecewise constant interpolation of the primal sequence $\boldsymbol{x}$ and piecewise affine interpolation of the dual sequence $\boldsymbol{p}^*$. Moreover, we introduce an additional interpolation technique for $\boldsymbol{p}^*$ based on a variational formulation.

Definition 3.2.2. 1. The *piecewise constant interpolation of* $\boldsymbol{x}$ is the function $c(\boldsymbol{\alpha},\boldsymbol{x}) : [0,\infty) \to X$ defined by $c(\boldsymbol{\alpha},\boldsymbol{x})(0) = x_0$ and

$$c(\boldsymbol{\alpha},\boldsymbol{x})(t) = x_n, \qquad \text{for } t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})].$$

2. The *piecewise affine interpolation of* $\boldsymbol{p}^*$ is the function $l(\boldsymbol{\alpha},\boldsymbol{p}^*) : [0,\infty) \to Y^*$ defined by $l(\boldsymbol{\alpha},\boldsymbol{p}^*)(0) = p_0^*$ and

$$l(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) = \alpha_n \left((t_n(\boldsymbol{\alpha}) - t)p_{n-1}^* + (t - t_{n-1}(\boldsymbol{\alpha}))p_n^*\right), \qquad \text{for } t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]. \tag{3.8}$$

3. A *variational interpolation of* $\boldsymbol{p}^*$ is a function $v(\boldsymbol{\alpha},\boldsymbol{p}^*) : [0,\infty) \to Y^*$ such that $v(\boldsymbol{\alpha},\boldsymbol{p}^*)(0) = p_0^*$ and

$$v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) \in R_{F^*}^{\omega}(p_{n-1}^*), \qquad \text{for } t = t_{n-1}(\boldsymbol{\alpha}) + \omega^{-1} \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]. \tag{3.9}$$

For each variational interpolation $v(\boldsymbol{\alpha},\boldsymbol{p}^*)$ of $\boldsymbol{p}^*$ we define the *backward difference quotient* of v as

$$\delta v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) = \frac{v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) - p_{n-1}^*}{t - t_{n-1}(\boldsymbol{\alpha})} \tag{3.10}$$

for $t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]$.

Remark 3.2.3. 1. From Definition 3.2.2 (2) it becomes clear, that the linear interpolation function $l(\boldsymbol{\alpha},\boldsymbol{p}^*)$ is differentiable almost everywhere in $[0,\infty)$ and right-differentiable everywhere an $[0,\infty)$. Risking a slight abuse of notation we identify

$$dl(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) \equiv \frac{\mathrm{d}^+}{\mathrm{d}t} l(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) = \alpha_n(p_n^* - p_{n-1}^*), \qquad \text{for } t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]. \tag{3.11}$$

2. For given $t > 0$ and a variational interpolation $v(\boldsymbol{\alpha},\boldsymbol{p}^*)$ of $\boldsymbol{p}^*$, we have that $v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) \in R_{F^*}^{\omega}(p_{n-1}^*)$, where $\omega = (t - t_{n-1}(\boldsymbol{\alpha}))^{-1}$. After recalling the definition of the slope $|\partial F^*|$ (cf. Definition A.2.1 (3)), it thus follows from Lemma A.3.10 and (3.10) that

$$|\partial F^*| \left(v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t)\right) \le \phi^{-1}\left(\omega \left\|v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t) - p_{n-1}^*\right\|\right) = \phi^{-1}\left(\|\delta v(\boldsymbol{\alpha},\boldsymbol{p}^*)(t)\|\right).$$

Due to the particular choice of the initial data $p_0^* \in Y^*$, the piecewise affine interpolation of the dual sequence $\boldsymbol{p}^*$ exhibits convenient properties. To be more precise:

Lemma 3.2.4. The function $t \mapsto \|dl(\boldsymbol{\alpha},\boldsymbol{p}^*)(t)\|$ is nonincreasing and for all $t \ge 0$ one has

$$\phi\left(\|Kc(\boldsymbol{\alpha},\boldsymbol{x})(t) - y\|\right) = \|dl(\boldsymbol{\alpha},\boldsymbol{p}^*)(t)\| \le \phi\left(\|Kx_0 - y\|\right). \tag{3.12}$$

In particular, the function $t \mapsto l(\boldsymbol{\alpha};\boldsymbol{p}^*)$ is Lipschitz continuous with Lipschitz constant $c_L = \phi\left(\|Kx_0 - y\|\right)$, i.e.

$$\|l(\boldsymbol{\alpha},\boldsymbol{p}^*)(s) - l(\boldsymbol{\alpha},\boldsymbol{p}^*)(t)\| \le \phi\left(\|Kx_0 - y\|\right)|s - t|. \tag{3.13}$$

for all $s, t \ge 0$.

Proof. From Lemma 2.3.1 and Remark 2.3.3 it follows (recall that $\mathcal{R}_n(y) = x_n$ and $\mathcal{R}_n^*(y) = p_n^*$) that for all $n \geq 1$

$$\|Kx_n - y\| \leq \|Kx_{n-1} - y\| \qquad \text{and} \qquad \phi\left(\|Kx_n - y\|\right) = \left\|\alpha_n(p_n^* - p_{n-1}^*)\right\|.$$

Thus it follows from Definition 3.2.2 and from the monotonicity of ϕ that

$$\phi\left(\|Kc(\boldsymbol{\alpha}, \boldsymbol{x})(t) - y\|\right) = \|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\|$$

for all $t \geq 0$ and that $t \mapsto \|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\|$ is nonincreasing.

Now let $n \in \mathbb{N}$ be such that $t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]$. Then it follows from Definition 3.2.2 and (3.12) that

$$\|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| = \phi\left(\|Kc(\boldsymbol{\alpha}, \boldsymbol{x}) - y\|\right) = \phi\left(\|Kx_n - y\|\right) \leq \phi\left(\|Kx_0 - y\|\right).$$

With this we eventually verify estimate (3.13):

$$\begin{aligned}\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(s)\| &= \left\|\int_s^t dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\,\mathrm{d}\tau\right\| \\ &\leq \int_s^t \|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|\,\mathrm{d}\tau \leq \phi\left(\|Kx_0 - y\|\right)|t - s|.\end{aligned}$$

□

At first glance, piecewise affine and variational interpolation seem to be of different nature: for the first being explicitly defined, whereas the latter is based on a variational formulation. In the remainder of this section we shed some light on the mutual relation of these two concepts.

Lemma 3.2.5. Assume that $t \geq 0$.

1. For all variational interpolations $v(\boldsymbol{\alpha}, \boldsymbol{p}^*)$ on has

$$\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| \leq \frac{2\phi\left(\|Kx_0 - y\|\right)}{|\boldsymbol{\alpha}|}.$$

2. If $t > 0$ then let $n \in \mathbb{N}$ be such that $t \in (t_{n-1}(\boldsymbol{\alpha}_\nu), t_n(\boldsymbol{\alpha}_\nu)]$ and set $n = 0$ otherwise. Then

$$\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - p_n^*\| \leq \frac{\phi\left(\|Kx_0 - y\|\right)}{|\boldsymbol{\alpha}|}.$$

Proof. Let $n \in \mathbb{N}$ be arbitrary. First we note, that from the definition of $\|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)\|$ and Lemma 3.2.4 it follows that

$$\alpha_n\left\|p_n^* - p_{n-1}^*\right\| = \|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| \leq \phi\left(\|Kx_0 - y\|\right).$$

Since $\alpha_n \geq |\boldsymbol{\alpha}|$ for all $n \in \mathbb{N}$ this gives

$$\left\|p_n^* - p_{n-1}^*\right\| \leq \frac{\phi\left(\|Kx_0 - y\|\right)}{\alpha_n} \leq \frac{\phi\left(\|Kx_0 - y\|\right)}{|\boldsymbol{\alpha}|}. \tag{3.14}$$

We prove (1). Since $l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(0) = p_0^* = v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(0)$ there is nothing to show for $t = 0$. Let $t > 0$ and find $n \in \mathbb{N}$ such that $t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]$. Then, there exists an element $p_\omega^* \in R_{F^*}^{\omega}(p_{n-1}^*)$ such that $t = t_{n-1}(\boldsymbol{\alpha}) + \omega^{-1}$ and $v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) = p_\omega^*$. Observe that

$$t_{n-1}(\boldsymbol{\alpha}) + \omega^{-1} = t \leq t_n(\boldsymbol{\alpha}) = t_{n-1}(\boldsymbol{\alpha}) + \alpha_n^{-1}$$

implies that $\alpha_n \leq \omega$. Since $p_n^* \in R_{F^*}^{\alpha_n}(p_{n-1}^*)$ it follows from Lemma A.3.3 that

$$\left\|p_\omega^* - p_{n-1}^*\right\| \leq \left\|p_n^* - p_{n-1}^*\right\|.$$

Note that $\alpha_n(t_n(\boldsymbol{\alpha}) - t) \leq 1$. Thus the definition of the piecewise affine interpolation and above considerations give

$$\begin{aligned}\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| &\leq \left\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - p_{n-1}^*\right\| + \left\|p_{n-1}^* - v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\right\| \\ &= \alpha_n(t_n(\boldsymbol{\alpha}) - t)\left\|p_n^* - p_{n-1}^*\right\| + \left\|p_\omega^* - p_{n-1}^*\right\| \leq 2\left\|p_n^* - p_{n-1}^*\right\|.\end{aligned}$$

This together with (3.14) gives

$$\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| \leq \frac{2\phi\left(\|Kx_0 - y\|\right)}{|\boldsymbol{\alpha}|}.$$

It remains to verify (2). Again, if $t = 0$ there is nothing to show. Hence let $t > 0$ and assume that n is chosen as above, i.e. $t \in (t_{n-1}(\boldsymbol{\alpha}_\nu), t_n(\boldsymbol{\alpha}_\nu)]$. From the definition of the piecewise affine interpolation it follows that

$$\begin{aligned}\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - p_n^*\| &= \left\|\alpha_n\left((t_n(\boldsymbol{\alpha}) - t)p_{n-1}^* + (t - t_{n-1}(\boldsymbol{\alpha}))p_n^*\right) - p_n^*\right\| \\ &= \left\|\alpha_n(t_n(\boldsymbol{\alpha}) - t)(p_{n-1}^* - p_n^*)\right\| \\ &= \alpha_n(t_n(\boldsymbol{\alpha}) - t)\left\|p_n^* - p_{n-1}^*\right\|,\end{aligned} \tag{3.15}$$

where the second equality follows from the identity

$$1 = \frac{\alpha_n}{\alpha_n} = \alpha_n(t_n(\boldsymbol{\alpha}) - t_{n-1}(\boldsymbol{\alpha})).$$

Since $\alpha_n(t_n(\boldsymbol{\alpha}) - t) \leq 1$ it follows from (3.15) that

$$\|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - p_n^*\| \leq \left\|p_n^* - p_{n-1}^*\right\|.$$

This together with (3.14) shows the assertion. □

Remark 3.2.6. Lemma 3.2.5 in particular implies that for all $t \geq 0$

$$\lim_{|\boldsymbol{\alpha}| \to \infty} \|l(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) - v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| = 0.$$

We close this section with an important energy identity: It states that the functional F^* decreases along the sequence $\{p_n^*\}_{n\in\mathbb{N}}$ and relates the derivative $dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)$ and the backward difference quotient $\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)$ with the decay of F^*. This result is based on Proposition A.3.9 proven in Appendix A.3.

Proposition 3.2.7. *For each variational interpolation $v(\boldsymbol{\alpha}, \boldsymbol{p}^*)$ the estimate*

$$\int_{t_n(\boldsymbol{\alpha})}^{t_m(\boldsymbol{\alpha})} \psi_{\phi^{-1}}(\|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|) + \psi_\phi(\phi^{-1}(\|\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|))\, d\tau = F^*(p_n^*) - F^*(p_m^*).$$

holds for all $m \geq n \geq 0$.

Proof. Let $n \geq 1$. Since $p_n^* \in R_{F^*}^{\alpha_n}(p_{n-1}^*)$ it follows from (A.30) in Corollary A.3.9 (recall the definition of $\mathrm{d}^\pm$ in (A.20)) that

$$F^*(p_{n-1}^*) - F^*(p_n^*) = \alpha_n^{-1}\psi_{\phi^{-1}}(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) + \int_{\alpha_n}^{\infty} \frac{1}{\omega^2}\psi_\phi(\phi^{-1}(\omega\, \mathrm{d}_\omega^\pm(p_{n-1}^*)))\, \mathrm{d}\omega. \tag{3.16}$$

Let $t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]$ and set $\omega(t) := (t - t_{n-1}(\boldsymbol{\alpha}))^{-1}$. Then there exists $p_{\omega(t)}^* \in R_{F^*}^{\omega(t)}(p_{n-1}^*)$ such that $v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t) = p_{\omega(t)}^*$. Consequently it follows from the definition of the backward difference quotient $\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)$ and Lemma A.3.4 that

$$\|\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\| = \frac{\left\|p_{n-1}^* - v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(t)\right\|}{t - t_{n-1}(\boldsymbol{\alpha})} = \frac{\left\|p_{n-1}^* - p_{\omega(t)}^*\right\|}{t - t_{n-1}(\boldsymbol{\alpha})} = \omega(t)\, \mathrm{d}_{\omega(t)}^\pm(p_{n-1}^*)$$

for a.e. $t \in (t_{n-1}(\boldsymbol{\alpha}), t_n(\boldsymbol{\alpha})]$. Performing the substitution $\omega(t) = (t - t_{n-1}(\boldsymbol{\alpha}))^{-1}$ in the integral in (3.16) hence shows

$$\begin{aligned}\int_{\alpha_n}^{\infty} \frac{1}{\omega^2}\psi_\phi(\phi^{-1}(\omega\, \mathrm{d}_\omega^\pm(p_{n-1}^*)))\, \mathrm{d}\omega &= \int_{t_n(\boldsymbol{\alpha})}^{t_{n-1}(\boldsymbol{\alpha})} \frac{1}{\omega^2(\tau)}\psi_\phi\left(\phi^{-1}\left(\omega(\tau)\, \mathrm{d}_{\omega(\tau)}^\pm(p_{n-1}^*)\right)\right) \frac{-\,\mathrm{d}\tau}{(\tau - t_{n-1}(\boldsymbol{\alpha}))^2} \\ &= \int_{t_{n-1}(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \psi_\phi\left(\phi^{-1}\left(\|\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|\right)\right) \mathrm{d}\tau.\end{aligned}$$

This together with (3.11) and (3.16) (note that $t_n(\boldsymbol{\alpha}) - t_{n-1}(\boldsymbol{\alpha}) = \alpha_n^{-1}$) results in

$$\begin{aligned}F^*(p_{n-1}^*) - F^*(p_n^*) &= \alpha_n^{-1}\psi_{\phi^{-1}}(\alpha_n \left\|p_n^* - p_{n-1}^*\right\|) + \int_{t_{n-1}(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \psi_\phi(\phi^{-1}(\|\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|))\, \mathrm{d}\tau \\ &= \int_{t_{n-1}(\boldsymbol{\alpha})}^{t_n(\boldsymbol{\alpha})} \psi_{\phi^{-1}}(\|dl(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|) + \psi_\phi(\phi^{-1}(\|\delta v(\boldsymbol{\alpha}, \boldsymbol{p}^*)(\tau)\|))\, \mathrm{d}\tau.\end{aligned}$$

The assertion follows by induction. □

3.3 Construction of Solutions

In this section we will prove that there exist solutions of Equation (3.4) in the sense of Definition 3.1.1. Moreover, we will explicitly construct solutions using the interpolation techniques introduced in Section 3.2. That is, we consider sequences of regularization parameters with increasing density and study the asymptotic behavior of the (interpolations of the) corresponding sequences generated by the augmented Lagrangian method (Algorithm 2.2.9).

First, we clarify some notation: Throughout this section we assume that $y \in Y$ and $p_0^* \in Y^*$ are fixed and that $x_0 \in X$ is such that

$$K^* p_0^* \in \partial J(x_0).$$

By $\{\boldsymbol{\alpha}_\nu\}_{\nu\in\mathbb{N}}$ we denote a *sequence of partitions* of $[0,\infty)$, that is, for each $\nu \in \mathbb{N}$

$$\boldsymbol{\alpha}_\nu = \{\alpha_{\nu,1}, \alpha_{\nu,2}, \ldots\} \subset (0,\infty)$$

is a partition of $[0,\infty)$ (cf. Definition 3.2.1). We shall assume for the densities $|\boldsymbol{\alpha}_\nu|$ that

$$\lim_{\nu\to\infty} |\boldsymbol{\alpha}_\nu| = \infty. \tag{3.17}$$

We further assume, that for given $\nu \in \mathbb{N}$, $\boldsymbol{x}_\nu \subset X$ and $\boldsymbol{p}^*_\nu \subset Y^*$ are arbitrary sequences generated by Algorithm 2.2.9 with respect to the data y, the initial value p^*_0 and the parameters $\boldsymbol{\alpha}_\nu$. Similar as with $\{\boldsymbol{\alpha}_\nu\}_{\nu\in\mathbb{N}}$ we will use the notation

$$\boldsymbol{x}_\nu = \{x_{\nu,1}, x_{\nu,2}, \ldots\} \subset X \quad \text{and} \quad \boldsymbol{p}^*_\nu = \left\{p^*_{\nu,1}, p^*_{\nu,2}, \ldots\right\} \subset Y^*.$$

For the sake of simplicity we define abbreviations for the interpolations of $\boldsymbol{x}_\nu$ and $\boldsymbol{p}^*_\nu$ discussed in the previous section. For $\nu \in \mathbb{N}$ and $t \geq 0$ we set

$x_\nu(t) := c(\boldsymbol{\alpha}_\nu, \boldsymbol{x}_\nu)(t)$	the piecewise constant interpolation of the primal sequence $\boldsymbol{x}_\nu$ (Definition 3.2.2 (1)).
$p^*_\nu(t) := l(\boldsymbol{\alpha}_\nu, \boldsymbol{p}^*_\nu)(t)$	the piecewise affine interpolation of the dual sequence $\boldsymbol{p}^*_\nu$ (Definition 3.2.2 (2)).
$\tilde{p}^*_\nu(t) := v(\boldsymbol{\alpha}_\nu, \boldsymbol{p}^*_\nu)(t)$	a variational interpolation of the dual sequence $\boldsymbol{p}^*_\nu$ (Definition 3.2.2 (3)).
$dp^*_\nu(t) = dl(\boldsymbol{\alpha}_\nu, \boldsymbol{p}^*_\nu)(t)$	the (right) derivative of the piecewise affine interpolation p^*_ν (Remark 3.2.3 (1)).
$\delta\tilde{p}^*_\nu(t) = \delta v(\boldsymbol{\alpha}_\nu, \boldsymbol{p}^*_\nu)(t)$	the backward difference quotient of the variational interpolation $\tilde{p}^*_\nu$ (Definition 3.2.2 (3)).

Table 3.1: Interpolations and related derivatives used in Section 3.3

Noting the abbreviations in Table 3.1, we summarize the results derived in Section 3.2.

Remark 3.3.1. Let $\nu \in \mathbb{N}$ and $t \geq 0$.

1. (cf. Remark 3.2.3 (2)). For the slope $|\partial F^*(\cdot; y)|$ we find

$$|\partial F^*(\cdot; y)|\,(\tilde{p}^*_\nu(t)) \leq \phi^{-1}\left(\|\delta\tilde{p}^*_\nu(t)\|\right).$$

2. (cf. Lemma 3.2.4) The function $t \mapsto \|dp^*_\nu(t)\|$ is nonincreasing and one has

$$\phi\left(\|Kx_\nu(t) - y\|\right) = \|dp^*_\nu(t)\| \leq \phi\left(\|Kx_0 - y\|\right).$$

In particular, this implies that $t \mapsto p^*_\nu(t)$ is Lipschitz continuous with Lipschitz constant $c_L = \phi\left(\|Kx_0 - y\|\right)$.

3. (cf. Lemma 3.2.5 (2)) If $n \in \mathbb{N}$ is such that $t \in (t_{n-1}(\boldsymbol{\alpha}_\nu), t_n(\boldsymbol{\alpha}_\nu)]$ (and $n = 0$ for $t = 0$), then
$$\left\|p_\nu^*(t) - p_{\nu,n}^*\right\| \leq \frac{\phi\left(\|Kx_0 - y\|\right)}{|\boldsymbol{\alpha}_\nu|}.$$
4. (cf. Remark 3.2.6)
$$\lim_{\nu\to\infty} \|p_\nu^*(t) - \tilde{p}_\nu^*(t)\| = 0.$$
5. (cf. Proposition 3.2.7) For all $m \geq n \geq 0$ one has
$$\int_{t_n(\boldsymbol{\alpha}_\nu)}^{t_m(\boldsymbol{\alpha}_\nu)} \psi_{\phi^{-1}}\left(\|dp_\nu^*(\tau)\|\right) + \psi_\phi\left(\phi^{-1}\left(\|\delta\tilde{p}_\nu^*(\tau)\|\right)\right) \,\mathrm{d}\tau = F^*(p_{\nu,n}^*; y) - F^*(p_{\nu,m}^*; y).$$

We will show that the sequence $\{(x_\nu, p_\nu^*)\}_{\nu\in\mathbb{N}}$ converges to a solution of (3.4) as $\nu \to \infty$ (in a sense yet to be defined). To this end we will pursue the following strategy:

First (Section 3.3.1), we shall show that as $\nu \to \infty$ the dual functions p_ν^* converge to a *strong solution*[1] p^* of the differential inclusion

$$\begin{aligned} \mathfrak{J}_{\phi^{-1}}\left(dp^*(t)\right) &\supset -\partial^0 F^*(p^*(t); y), &(3.18a)\\ p^*(0) &= p_0^*. &(3.18b)\end{aligned}$$

Here, the functional $F^*(\cdot; y)$ is as in (2.13), i.e. for all $q^* \in Y^*$
$$F^*(q^*; y) = J^*(K^*q^*) - \langle q^*, y\rangle$$
and $\partial^0 F^*(\cdot; y)$ denotes the minimal section of the subdifferential $\partial F^*(\cdot; y)$ (cf. Definition A.2.1). With the preparations in Section 3.2 (summarized in Remark 3.3.1), this result will follow from the analysis in [9, Chap. 3].

Next (Section 3.3.2), with a strong solution p^* of (3.18) at hand, we will prove that also the piecewise constant functions x_ν converge (in a weaker sense) to a function $x : [0, \infty) \to X$, such that (x, p^*) is a solution of Equation (3.4).

3.3.1 Convergence of Dual Discrete Solutions

In this paragraph we prove the pointwise (weak) convergence of the piecewise affine functions p_ν^* (cf. Table 3.1) to a solution of (3.18). Hereafter we investigate, in which effect convergence properties can be improved by imposing additional restrictions on the spaces Y and Y^*.

We start our analysis with convergence result for the sequence $\left\{\left\|dp_\nu^*\right\|\right\}_{\nu\in\mathbb{N}}$.

Lemma 3.3.2. There exists a selection $\nu \mapsto \rho(\nu)$ and a nonincreasing function $d \in \mathrm{L}^\infty(0, \infty)$ such that for all $t > 0$
$$\lim_{\nu\to\infty} \left\|dp_{\rho(\nu)}^*(t)\right\| = d(t) \quad \text{ and } \quad d(t) \leq \phi\left(\|Kx_0 - y\|\right). \tag{3.19}$$
In particular, one has for all $T > 0$ and $p \in [1, \infty)$ that
$$\lim_{\nu\to\infty} \|dp_\nu^*\|_{\mathrm{L}^p(0,T;Y^*)} = \lim_{\nu\to\infty} \int_0^T \left\|dp_{\rho(\nu)}^*(\tau)\right\|^p \mathrm{d}\tau = \lim_{\nu\to\infty} \int_0^T |d(\tau)|^p \,\mathrm{d}\tau. \tag{3.20}$$

[1] We recall that p^* is a strong solution of (3.18) if it is absolutely continuous on each compact interval and satisfies (3.18b) as well as (3.18a) for a.e. $t \geq 0$.

Proof. According to Remark 3.3.1 (2), the functions $\left\|dp_\nu^*\right\| : [0,\infty) \to \mathbb{R}$ are nonincreasing for all $\nu \in \mathbb{N}$ and it becomes evident that

$$\left\|dp_\nu^*(t)\right\| \le \phi(\|Kx_0 - y\|) =: c, \qquad \text{for all } \nu \in \mathbb{N} \text{ and } t \in [0,\infty).$$

From Helly's Lemma A.1.20 we hence conclude that there exists a selection $\nu \mapsto \rho(\nu)$ and a nonincreasing function $d : [0,\infty) \to [0,\infty]$ such that (3.19) holds with

$$\sup_{t\ge 0} d(t) = \sup_{t\ge 0} \left(\lim_{\nu\to\infty} \left\|dp_{\rho(\nu)}^*(t)\right\| \right) \le c.$$

We summarize that $d : [0,\infty) \to [0,\infty)$ is bounded (by c) and nonincreasing and therefore measurable, i.e. $d \in \mathrm{L}^\infty(0,\infty)$.

Finally, choose $1 \le p < \infty$ and observe that for all $t \ge 0$

$$\left\|dp_{\rho(\nu)}^*(t)\right\|^p \le c^p.$$

Thus the dominated convergence theorem (for the finite measure space $[0,T]$) is applicable and (3.20) follows. □

We move on to the fundamental result of this subsection that proves pointwise weak convergence of the piecewise affine interpolations p_ν^* to a strong solution p^* of the dual equation (3.18). We refer to [9, Chap. 3.3, Chap. 3.4] for the original proof (for the case $\phi(s) = s^{p-1}$).

Theorem 3.3.3. *There exists a selection $\nu \mapsto \rho(\nu)$ and a Lipschitz continuous function $p^* : [0,\infty) \to Y^*$ with Lipschitz constant $c_L = \phi(\|Kx_0 - y\|)$ that satisfies*

$$\operatorname*{w\text{-}lim}_{\nu\to\infty} p_{\rho(\nu)}^*(t) = p^*(t) \qquad \text{for all } t \in [0,\infty). \tag{3.21}$$

Moreover, p^* is a strong solution of (3.18).

Proof. Since the dependence of $F^*(\cdot; y)$ on $y \in Y$ is not significant for this proof, we will use the notation $F^*(q^*)$ instead of $F^*(q^*; y)$ (for $q^* \in Y^*$). In order to keep the presentation as lucid as possible, we divide the proof in several steps.

Claim 1. Existence of the limit in (3.21): Let $T > 0$. From Remark 3.3.1 (2) we learn that the piecewise affine interpolations p_ν^* are Lipschitz continuous with Lipschitz constant $c_L = \phi(\|Kx_0 - y\|)$. This shows that

$$\|p_\nu^*(t) - p_0^*\| = \|p_\nu^*(t) - p_\nu^*(0)\| \le \phi(\|Kx_0 - y\|)t \le \phi(\|Kx_0 - y\|)T$$

or in other words

$$\|p_\nu^*(t)\| \le \|p_0^*\| + \phi(\|Kx_0 - y\|)T =: C(T) \qquad \text{for } 0 \le t \le T.$$

This means that $p_\nu^*(t)$ lies in the ball with radius $C(T)$ in Y^*, which is a sequentially weakly compact set, for Y^* being reflexive according to requirement (R7).

Moreover, we have for $0 \le s < t \le T$ that

$$\|p_\nu^*(t) - p_\nu^*(s)\| = \left\| \int_s^t dp_\nu^*(\tau)\, \mathrm{d}\tau \right\| \le \int_s^t \|dp_\nu^*(\tau)\|\, \mathrm{d}\tau.$$

From this estimate and Lemma 3.3.2 we conclude that there exists a function $d \in \mathrm{L}^\infty(0,\infty)$, such that (possibly after extracting a subsequence)

$$\limsup_{\nu\to\infty} \|p_\nu^*(t) - p_\nu^*(s)\| \le \limsup_{\nu\to\infty} \int_s^t \|dp_\nu^*(\tau)\| \,\mathrm{d}\tau = \int_s^t d(\tau)\,\mathrm{d}\tau. \tag{3.22}$$

Thus we can apply the Arzelà – Ascoli Theorem A.1.19 and observe that there exists a absolutely continuous function $p^* : [0,T] \to Y^*$ and a selection $\nu \mapsto \rho(\nu)$, such that

$$w\text{-}\lim_{\nu\to\infty} p^*_{\rho(\nu)}(t) = p^*(t)$$

for all $t \in [0,T]$.

We note, that due to Theorem A.1.16 the function p^* is differentiable a.e. in $[0,T]$. By a standard diagonalization argument we extend p^* to the whole halfline $[0,\infty)$ such that (3.21) holds for all $t \ge 0$ and a selection $\nu \mapsto \rho(\nu)$.

Finally we note that from Lemma 3.3.2 and (3.22) (and the weak lower semicontinuity of $\|\cdot\|$) it becomes evident that

$$\|p^*(t) - p^*(s)\| \le \int_s^t d(\tau)\,\mathrm{d}\tau \le \int_s^t \phi\left(\|Kx_0 - y\|\right)\,\mathrm{d}\tau = \phi\left(\|Kx_0 - y\|\right)|t-s| \tag{3.23}$$

for all $s, t \ge 0$. This shows that p^* is Lipschitz continuous (with Lipschitz constant $c_L = \phi\left(\|Kx_0 - y\|\right)$).

It remains to prove, that p^* is a strong solution of (3.18). We shall assume (by re-indexing of the selection $\nu \mapsto \rho(\nu)$) that for all $t \ge 0$

$$\lim_{\nu\to\infty} \|dp_\nu^*(t)\| = d(t) \quad \text{and} \quad w\text{-}\lim_{\nu\to\infty} p_\nu^*(t) = p^*(t).$$

Claim 2. $\|dp^*(t)\| = \phi(|\partial F^*|\,(p^*(t)))$: From continuity of the norm on Y^* and (3.23) we conclude for every Lebesgue point $t \in [0,\infty)$ of d

$$\|dp^*(t)\| = \lim_{h\to 0} \frac{\|p^*(t+h) - p^*(t)\|}{|h|} \le \lim_{h\to 0} \frac{1}{|h|} \int_t^{t+h} d(\tau)\,\mathrm{d}\tau = d(t) = \lim_{\nu\to\infty} \|dp_\nu^*(t)\|\,. \tag{3.24}$$

Recall, that $\tilde{p}_\nu^*$ denotes a variational interpolation of the sequence $\boldsymbol{p}_\nu^*$ (cf. Table 3.1). Since $p_\nu(t) \rightharpoonup p^*(t)$ it follows from Remark 3.3.1 (4) that

$$w\text{-}\lim_{\nu\to\infty} \tilde{p}_\nu^*(t) = w\text{-}\lim_{\nu\to\infty} (\tilde{p}_\nu^*(t) - p_\nu^*(t)) + w\text{-}\lim_{\nu\to\infty} p_\nu^*(t) = p^*(t).$$

Moreover, according to Remark 3.3.1 (1) one has

$$|\partial F^*|\,(\tilde{p}_\nu^*(t)) \le \phi^{-1}(\|\delta\tilde{p}_\nu^*(t)\|) \quad \text{for all } t \ge 0.$$

Since $\partial F^*(\cdot;y) = \partial(J^* \circ K^*) - y$ is weakly-weakly closed according to requirement (R8b) it follows with Lemma A.2.3 that the slope $|\partial F^*(\cdot;y)|$ is sequentially weakly lower semicontinuous. Summarizing this fact and the last two estimates, we find

$$|\partial F^*|\,(p^*(t)) = |\partial F^*|\,(w\text{-}\lim_{\nu\to\infty} \tilde{p}_\nu^*(t)) \le \liminf_{\nu\to\infty} |\partial F^*|\,(\tilde{p}_\nu^*(t)) \le \liminf_{\nu\to\infty} \phi^{-1}(\|\delta\tilde{p}_\nu^*(t)\|). \tag{3.25}$$

For each $\nu \in \mathbb{N}$ we choose $n(\nu) \in \mathbb{N}$ such that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$ (we agree upon $n(\nu) = 0$ for $t = 0$). Then it follows from Remark 3.3.1 (3) and from the fact that $p^*_\nu(t) \rightharpoonup p^*(t)$ that

$$w\text{-}\lim_{\nu\to\infty} p^*_{\nu,n(\nu)} = w\text{-}\lim_{\nu\to\infty}(p^*_{\nu,n(\nu)} - p^*_\nu(t)) + w\text{-}\lim p^*_\nu(t) = p^*(t).$$

From the weak lower semicontinuity of F^* it hence follows that

$$F^*(p^*(t)) = F^*(w\text{-}\lim p^*_{\nu,n(\nu)}) \leq \liminf_{\nu\to\infty} F^*(p^*_{\nu,n(\nu)}). \tag{3.26}$$

Combination of the estimates (3.24), (3.25) and (3.26) gives

$$\begin{aligned}
&\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|) + \psi_\phi(|\partial F^*|\,(p^*(\tau))\,\mathrm{d}\tau + F^*(p^*(t)) \\
&\quad\leq \int_0^t \liminf_{\nu\to\infty} \psi_{\phi^{-1}}(\|dp^*_\nu(\tau)\|) + \liminf_{\nu\to\infty} \psi_\phi\left(\phi^{-1}(\|\delta p^*_\nu(\tau)\|)\right)\,\mathrm{d}\tau + \liminf_{\nu\to\infty} F^*(p^*_{\nu,n(\nu)}).
\end{aligned}\tag{3.27}$$

Moreover, the energy estimate in Remark 3.3.1 (5) gives (note that $t \leq t_{n(\nu)}(\boldsymbol{\alpha}_\nu)$ for all $\nu \in \mathbb{N}$)

$$\begin{aligned}
&\int_0^t \psi_{\phi^{-1}}\left(\|dp^*_\nu(\tau)\|\right) + \psi_\phi\left(\phi^{-1}\left(\|\delta\tilde{p}^*_\nu(\tau)\|\right)\right)\,\mathrm{d}\tau \\
&\quad\leq \int_0^{t_{n(\nu)}(\boldsymbol{\alpha}_\nu)} \psi_{\phi^{-1}}\left(\|dp^*_\nu(\tau)\|\right) + \psi_\phi\left(\phi^{-1}\left(\|\delta\tilde{p}^*_\nu(\tau)\|\right)\right)\,\mathrm{d}\tau = F^*(p^*_0) - F^*(p^*_{\nu,n(\nu)}).
\end{aligned}\tag{3.28}$$

This estimate together with (3.27) and Fatou's Lemma shows

$$\begin{aligned}
&\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|) + \psi_\phi(|\partial F^*|\,(p^*(\tau))\,\mathrm{d}\tau + F^*(p^*(t)) \\
&\quad\leq \liminf_{\nu\to\infty}\left(\int_0^t \psi_{\phi^{-1}}(\|dp^*_\nu(\tau)\|) + \psi_\phi\left(\phi^{-1}(\|\delta p^*_\nu(\tau)\|)\right)\,\mathrm{d}\tau + F^*(p^*_{\nu,n(\nu)})\right) \leq F^*(p^*_0).
\end{aligned}\tag{3.29}$$

Furthermore, we find from Lemma A.2.2 (3) that for all $t \geq 0$

$$0 \leq F^*(p^*_0) - F^*(p^*(t)) \leq \int_0^t |\partial F^*|\,(p^*(\tau))\,\|dp^*(\tau)\|\,\mathrm{d}\tau.$$

Combining the previous two estimates with Fenchel's inequality (A.11) gives

$$\begin{aligned}
\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|)\,\mathrm{d}\tau + \int_0^t \psi_\phi(|\partial F^*|(p^*(\tau))\,\mathrm{d}\tau
&\underset{(3.29)}{\leq} F^*(p^*_0) - F^*(p^*(t)) \\
&\leq \int_0^t |\partial F^*|\,(p^*(\tau))\,\|dp^*(\tau)\|\,\mathrm{d}\tau \\
&\underset{(A.11)}{\leq} \int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|)\,\mathrm{d}\tau + \int_0^t \psi_\phi(|\partial F^*|\,(p^*(\tau))\,\mathrm{d}\tau.
\end{aligned}\tag{3.30}$$

This implies that

$$\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|) + \psi_\phi(|\partial F^*|\,(p^*(\tau)) - |\partial F^*|\,(p^*(\tau))\,\|dp^*(\tau)\|\,\mathrm{d}\tau = 0 \tag{3.31}$$

and

$$F^*(p_0^*) - F^*(p^*(t)) = \int_0^t |\partial F^*| (p^*(\tau)) \|dp^*(\tau)\| \, \mathrm{d}\tau. \tag{3.32}$$

Again, from Fenchel's inequality we conclude that the integrand in (3.31) is nonnegative and therefore vanishes a.e. in $[0, t]$, that is

$$\psi_{\phi^{-1}}(\|dp^*(t)\|) + \psi_\phi(|\partial F^*| (p^*(t)) = |\partial F^*| (p^*(t)) \|dp^*(t)\|$$

for a.e. $t \geq 0$. Combining this with the relation

$$\psi_{\phi^{-1}}(t) + \psi_\phi(s) = st \;\Leftrightarrow\; t = \phi(s)$$

(cf. (A.10) in Example A.2.13) implies that for a.e. $t \in [0, \infty)$

$$\|dp^*(t)\| = \phi\left(|\partial F^*| (p^*(t))\right) \tag{3.33}$$

as desired.

Claim 3. $\frac{\mathrm{d}}{\mathrm{d}t} F^*(p^*(t)) = - |\partial F^*| (p^*(t)) \|dp^*(t)\|$: Let $s < t$. Then it follows from (3.32) that

$$\begin{aligned} F^*(p^*(s)) - F^*(p^*(t)) &= (F^*(p_0^*) - F^*(p^*(t))) - (F^*(p_0^*) - F^*(p^*(s))) \\ &= \int_0^t |\partial F^*| (p^*(\tau)) \|dp^*(\tau)\| \, \mathrm{d}\tau - \int_0^s |\partial F^*| (p^*(\tau)) \|dp^*(\tau)\| \, \mathrm{d}\tau \\ &= \int_s^t |\partial F^*| (p^*(\tau)) \|dp^*(\tau)\| \, \mathrm{d}\tau. \end{aligned}$$

and consequently it is evident that for a.e. $t \geq 0$

$$\frac{\mathrm{d}}{\mathrm{d}t} F^*(p^*(t)) = - |\partial F^*| (p^*(t)) \|dp^*(t)\| \,. \tag{3.34}$$

Claim 4. $d(t) = \|dp^*(t)\|$: From (3.27), (3.29) and (3.30) it follows that

$$\begin{aligned} &\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|) \, \mathrm{d}\tau + \int_0^t \psi_\phi(|\partial F^*| (p^*(\tau)) \, \mathrm{d}\tau + F^*(p^*(t)). \\ &= \int_0^t \liminf_{\nu \to \infty} \psi_{\phi^{-1}}(\|dp_\nu^*(\tau)\|) + \liminf_{\nu \to \infty} \psi_\phi\left(\phi^{-1}(\|\delta p_\nu^*(\tau)\|)\right) \, \mathrm{d}\tau + \liminf_{\nu \to \infty} F^*(p^*_{\nu, n(\nu)}). \end{aligned} \tag{3.35}$$

From the estimates (3.25) and (3.26) it follows that $\liminf_{\nu \to \infty} F^*(p^*_{\nu,n(\nu)}) \geq F^*(p^*(t))$ and

$$\int_0^t \liminf_{\nu \to \infty} \psi_\phi\left(\phi^{-1}(\|\delta p_\nu^*(\tau)\|)\right) \, \mathrm{d}\tau \geq \int_0^t \psi_\phi(|\partial F^*| (p^*(\tau)) \, \mathrm{d}\tau.$$

Therefore we conclude from (3.35) that

$$\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|) - \liminf_{\nu \to \infty} \psi_{\phi^{-1}}(\|dp_\nu^*(\tau)\|) \, \mathrm{d}\tau \geq 0.$$

Consequently we find from Lemma 3.3.2 (keeping in mind the continuity of $\psi_{\phi^{-1}}$) that

$$\int_0^t \psi_{\phi^{-1}}(\|dp^*(\tau)\|) - \psi_{\phi^{-1}}(d(\tau)) \, \mathrm{d}\tau \geq 0.$$

Since $\|dp^*(t)\| \leq d(t)$ for a.e. $t > 0$ according to (3.24) this clearly amounts to saying that (note that $\psi_{\phi^{-1}}$ is nondecreasing)

$$\lim_{\nu\to\infty} \psi_{\phi^{-1}}(\|dp_\nu^*(t)\|) = \psi_{\phi^{-1}}(d(t)) = \psi_{\phi^{-1}}(\|dp^*(t)\|) \quad \text{a.e. in } [0,\infty). \tag{3.36}$$

Injectivity of $\psi_{\phi^{-1}}$ finally shows that $d(t) = \|dp^*(t)\|$.

Claim 5. Conclusion of the proof: For an arbitrary element $\eta \in \partial^0 F^*(p(t))$ it follows from the definition of the subgradient and the fact that $\|\eta\| = |\partial F^*|(p^*(t))$ that

$$-\left(\frac{F^*(p^*(t+h)) - F^*(p^*(t))}{h}\right) \leq \frac{\langle p^*(t) - p^*(t+h), \eta\rangle}{h} \leq |\partial F^*|(p^*(t)) \frac{\|p^*(t+h) - p^*(t)\|}{h}.$$

After letting $h \to 0$ is above estimate it follows together with Claim 3 and 4 that

$$-\frac{d}{dt}F^*(p^*(t)) \leq -\langle dp^*(t), \eta\rangle \leq |\partial F^*|(p^*(t))\|dp^*(t)\| = -\frac{d}{dt}F^*(p^*(t))$$

and consequently

$$-\langle dp^*(t), \eta\rangle = |\partial F^*|(p^*(t))\|dp^*(t)\|.$$

Since $|\partial F^*|(p^*(t)) = \phi^{-1}(\|dp^*(t)\|)$ (according to Claim 2) it follows from the definition of the duality mapping $\mathfrak{J}_{\phi^{-1}}$ (cf. Definition A.1.1) that $-\eta \in \mathfrak{J}_{\phi^{-1}}(dp^*(t))$. Thus we eventually find

$$-\partial^0 F^*(p^*(t)) \subset \mathfrak{J}_{\phi^{-1}}(dp^*(t))$$

as desired. □

For the remainder of this section let us assume that p^* is a strong solution of (3.18) and (after dropping a suitable subsequence) that $p_\nu^*(t) \rightharpoonup p^*(t)$ for all $t \geq 0$. From the proof of Theorem 3.3.3 we collect some important relations used in the course of the upcoming analysis:

Remark 3.3.4. Let $t \geq 0$.

1. (cf. Claim 2) For the slope we find
$$|\partial F^*(\cdot; y)|(p^*(t)) = \phi^{-1}(\|dp^*(t)\|).$$
2. (cf. Claim 3) For all $s \leq t$ one has
$$F^*(p^*(s); y) - F^*(p^*(t); y) = \int_s^t |\partial F^*(\cdot; y)|(p^*(\tau))\|dp^*(\tau)\| \,\mathrm{d}\tau.$$
In particular, the function $t \mapsto F^*(p^*(t); y)$ is nonincreasing and
$$\frac{d}{dt}F^*(p^*(t); y) = -|\partial F^*(\cdot; y)|(p^*(t))\|dp^*(t)\|.$$
3. (cf. Lemma 3.3.2 and Claim 4) The function $t \mapsto \|dp^*(t)\|$ is nonincreasing and for each $T > 0$ and $p \in [1,\infty)$ one has
$$\lim_{\nu\to\infty} \|dp_\nu^*(t)\| = \|dp^*(t)\| \quad \text{and} \quad \lim_{\nu\to\infty} \|dp_\nu^*\|_{\mathrm{L}^p(0,T;Y^*)} = \|dp^*\|_{\mathrm{L}^p(0,T;Y^*)}.$$
Moreover, one has for all $t \geq 0$ and $\nu \in \mathbb{N}$ that
$$\max(\|dp_\nu^*(t)\|, \|dp^*(t)\|) \leq \phi(\|Kx_0 - y\|).$$

From Remark 3.3.4 (3) we already find that the norms of the derivatives of the piecewise affine interpolations $\{p_\nu^*\}_{\nu\in\mathbb{N}}$ converge to $\|dp^*(t)\|$ for all $t \geq 0$. When Y is assumed to by a *separable* Banach space we additionally get

Theorem 3.3.5. *Assume that Y is separable. Then for each $T > 0$ one has that*

$$w^*\text{-}\lim_{\nu\to\infty} dp_\nu^* = dp^*$$

in $L^\infty(0, T; Y^*)$.

Proof. Let $T > 0$. From Remark 3.3.1 (2) it follows that

$$\|dp_\nu^*(t)\| \leq \phi\left(\|K(x_0) - y\|\right) =: r < \infty, \qquad \text{for all } 0 \leq t \leq T,\ \nu \in \mathbb{N}.$$

This implies that

$$\|dp^*\|_{\mathrm{L}^\infty(0,T;Y^*)} = \sup_{0\leq t\leq T} \|dp^*(t)\| \leq r.$$

In other words, this means that the sequence $\{dp_\nu^*\}$ lies in the ball with radius r in the space $\mathrm{L}^\infty(0, T; Y^*)$, which is sequentially weakly compact for Y being separable (cf. Corollary A.1.12). Thus there exists $q^* \in \mathrm{L}^\infty(0, T; Y^*)$ and a selection $\nu \mapsto \rho(\nu)$ such that

$$dp_{\rho(\nu)}^* \rightharpoonup^* q^*$$

We remark that this is equivalent to

$$\lim_{\nu\to\infty} \int_0^T \left\langle dp_{\rho(\nu)}^*(\tau), g(\tau)\right\rangle \,\mathrm{d}\tau = \int_0^T \langle q^*, g(\tau)\rangle \,\mathrm{d}\tau, \quad \text{for all } g \in \mathrm{L}^1(0, T; X). \tag{3.37}$$

We show that $dp^* = q^*$ λ^1-a.e. in $(0, T)$. To this end we first note that for every $\nu \in \mathbb{N}$ and $0 \leq t \leq T$ we have that

$$p_{\rho(\nu)}^*(t) = p_0^* + \int_0^t dp_{\rho(\nu)}^*(\tau)\,\mathrm{d}\tau.$$

From Theorem 3.3.3 we see that the left hand side of above equation weakly converges to $p^*(t)$, that is for all $h \in Y$ we have that

$$\lim_{\nu\to\infty} \left\langle \int_0^t dp_{\rho(\nu)}^*(\tau)\,\mathrm{d}\tau, h\right\rangle = \langle p^*(t) - p^*(0), h\rangle\,.$$

By applying Lemma A.1.15 (note that $Y = Y^{**}$) and keeping in mind (3.37) (with the constant function $g(t) = h$) we can evaluate the left hand side of the previous equation to

$$\begin{aligned}\lim_{\nu\to\infty} \left\langle \int_0^t dp_{\rho(\nu)}^*(\tau)\,\mathrm{d}\tau, h\right\rangle &\underset{\text{Lem. A.1.15}}{=} \lim_{\nu\to\infty} \int_0^t \left\langle dp_{\rho(\nu)}^*(\tau), g(\tau)\right\rangle \,\mathrm{d}\tau \\ &\underset{(3.37)}{=} \int_0^t \langle q^*(\tau), g(\tau)\rangle \,\mathrm{d}\tau \underset{\text{Lem. A.1.15}}{=} \left\langle \int_0^t q^*(\tau)\,\mathrm{d}\tau, h\right\rangle.\end{aligned}$$

Combining the last two formulas gives

$$\left\langle \int_0^t q^*(\tau)\,\mathrm{d}\tau, h\right\rangle = \langle p^*(t) - p_0^*, h\rangle \qquad \text{for all } h \in Y \text{ and } 0 \leq t \leq T,$$

which in turn implies

$$p^*(t) = p_0^* + \int_0^t q^*(\tau)\,\mathrm{d}\tau \qquad \text{for all } 0 \le t \le T.$$

From the fundamental theorem of calculus A.1.16 it finally follows that $q^*(t) = dp^*(t)$ a.e. in $[0,T]$, as desired. Therefore, for every subsequence of $\{dp_\nu^*\}_{\nu\in\mathbb{N}}$ one can drop another subsequence, that weakly* converges to dp^* in $\mathrm{L}^\infty(0,T;Y^*)$. This already shows that

$$dp_\nu^* \rightharpoonup^* dp^*$$

and the theorem is proven. □

By imposing additional restrictions on the space Y and Y^*, we can establish a stronger convergence behavior of the sequences $\{p_\nu^*\}_{\nu\in\mathbb{N}}$ and $\{dp_\nu^*\}_{\nu\in\mathbb{N}}$. In particular, we can formulate sufficient smoothness conditions on Y^*, such that the sequence $\{p_\nu^*\}_{\nu\in\mathbb{N}}$ converges *locally uniformly* in the *strong topology* of Y^*.

Recall that an *E-space* is a reflexive and strictly convex Banach space, that has the Radon – Riesz property (see Definition A.1.14). Standard examples for E-spaces are the spaces $\mathrm{L}^p(\Omega)$ where $\Omega \subset \mathbb{R}^N$ is sufficiently smooth and $1 < p < \infty$ (cf. Chapter 4).

Corollary 3.3.6. Assume that Y is separable and that Y^* is an E-space. Then for each $T > 0$ and $1 < p < \infty$ one has that

$$\lim_{\nu\to\infty} p_\nu^* = p^*$$

in $\mathrm{W}^{1,p}(0,T;Y^*)$. In particular, this implies that

$$\lim_{\nu\to\infty} \sup_{0\le t\le T} \|p_\nu^*(t) - p^*(t)\| = 0.$$

Proof. Let $T > 0$ and fix $p \in (1,\infty)$. Since Y is separable, it follows from Theorem 3.3.5 that $\{dp_\nu^*\}_{\nu\in\mathbb{N}}$ weakly* converges to dp^* in $\mathrm{L}^\infty(0,T;Y^*)$. This implies that

$$w\text{-}\lim_{\nu\to\infty} dp_\nu^* = dp^*$$

in $\mathrm{L}^p(0,T;Y^*)$. Moreover, from Remark 3.3.4 (3) it follows that

$$\lim_{\nu\to\infty} \|dp_\nu^*\|^p_{\mathrm{L}^p(0,T;Y^*)} = \lim_{\nu\to\infty} \int_0^T \|dp_\nu^*(\tau)\|^p \,\mathrm{d}\tau = \int_0^T \|dp^*(\tau)\|^p \,\mathrm{d}\tau = \|dp^*\|^p_{\mathrm{L}^p(0,T;Y^*)}\,.$$

Thus $\{dp_\nu^*\}_{\nu\in\mathbb{N}}$ converges strongly in $\mathrm{L}^p(0,T;Y^*)$ due to the Radon – Riesz property of $\mathrm{L}^p(0,T;Y^*)$ (cf. Theorem A.1.13).

We now verify that $\{p_\nu^*\}_{\nu\in\mathbb{N}}$ converges in $\mathrm{L}^p(0,T;Y^*)$. To this end, let $t \ge 0$ and observe from Theorem A.1.16 and from the strong convergence of dp_ν^* in $\mathrm{L}^p(0,T;Y^*)$

$$\begin{aligned} p^*(t) = p_0^* + \int_0^t dp^*(\tau)\,\mathrm{d}\tau = p_0^* + \int_0^t dp^*(\tau) - dp_\nu^*(\tau)\,\mathrm{d}\tau + \int_0^t dp_\nu^*(\tau)\,\mathrm{d}\tau \\ = p_0^* + \mathcal{O}\left(\|dp^* - dp_\nu^*\|_{\mathrm{L}^p(0,t;Y^*)}\right) + p_\nu^*(t) - p_0^*. \end{aligned}$$

This proves that $p_\nu^*(t) \to p^*(t)$ strongly, as $\nu \to \infty$. From Remark 3.3.1 (2) it follows that the function $t \mapsto p_\nu^*(t)$ is Lipschitz continuous with Lipschitz constant $c_L = \phi(\|Kx_0 - y\|)$. This gives that for all $0 \le t \le T$

$$\begin{aligned}\|p_\nu^*(t)\| &\le \|p_0^*\| + \|p_\nu^*(t) - p_\nu^*(0)\| \\ &\le \|p_0^*\| + \phi(\|Kx_0 - y\|)\, t \le \|p_0^*\| + \phi(\|Kx_0 - y\|)\, T =: C(T).\end{aligned}$$

From the continuity of the norm on Y^* it follows for $0 \le t \le T$ that

$$\|p^*(t)\| = \left\| \lim_{\nu\to\infty} p_\nu^*(t) \right\| = \lim_{\nu\to\infty} \|p_\nu^*(t)\| \le C(T).$$

From the last two estimates we conclude that

$$\|p^*(t) - p_\nu(t)\|^p \le 2^p \left(\|p^*(t)\|^p + \|p_\nu(t)\|^p\right) \le (2C(T))^p.$$

Therefore, by dominated convergence, we find that $\{p_\nu^*\}_{\nu\in\mathbb{N}}$ converges strongly in $\mathrm{L}^p(0,T;Y^*)$. Summarizing, we find

$$\lim_{\nu\to\infty} p_\nu^* = p^*, \quad \text{in } \mathrm{W}^{1,p}(0,T;Y^*) \tag{3.38}$$

In order to conclude the proof, we note that (3.38) is equivalent to (cf. Definition A.1.17)

$$\lim_{\nu\to\infty} \int_0^T \|p_\nu^*(\tau) - p^*(\tau)\|^p \, \mathrm{d}\tau + \int_0^T \|dp_\nu^*(\tau) - dp^*(\tau)\|^p \, \mathrm{d}\tau = 0.$$

This implies that the sequence $\{t \mapsto \|p_\nu^*(t) - p^*(t)\|\}_{\nu\in\mathbb{N}}$ converges strongly in $\mathrm{W}^{1,p}(0,T)$ and hence it follows from the continuous embedding $\mathrm{W}^{1,p}(0,T) \hookrightarrow C(0,T)$ (cf. [2, Thm. 5.4] with $n = 1$, $m = 1$ and $p > 1$) that

$$\lim_{\nu\to\infty} \sup_{0\le t\le T} \|p_\nu^*(t) - p^*(t)\| = 0.$$

□

We summarize the results of this section: Let $y \in Y$ and assume that $p_0^* \in Y^*$ is chosen according to (R6), that is, $K^* p_0^* \in \partial J(x_0)$ for an element $x_0 \in X$.

Condition	Convergence of p_ν^*	Convergence of $dp_\nu^*(t)$
—	$p_\nu^*(t) \rightharpoonup^* p^*(t)$ for $t \ge 0$	$\|dp_\nu^*(t)\| \to \|dp^*(t)\|$ for all $t \ge 0$
Y separable	$p_\nu^*(t) \rightharpoonup^* p^*(t)$ for $t \ge 0$	$dp_\nu^* \rightharpoonup^* dp^*$ in $\mathrm{L}^\infty(0,T;Y^*)$ for all $T > 0$
Y separable and Y^* E-space	$p_\nu^* \to p^*$ locally uniformly	$dp_\nu^* \to dp^*$ in $\mathrm{L}^p(0,T;Y^*)$ for all $T > 0$ and $1 < p < \infty$

Table 3.2: Convergence properties of strong solutions of (3.18)

Then, there exists a Lipschitz continuous function $p^* : [0,\infty) \to Y$ with Lipschitz constant $c_L = \phi(\|Kx_0 - y\|)$ such that p^* is a strong solution of (3.18). Moreover, p^* can be written (possibly after dropping a subsequence) as the pointwise weak* limit of the piecewise affine interpolations $\{p_\nu^*\}_{\nu\in\mathbb{N}}$ as defined in Table 3.1.

By imposing additional restrictions on the spaces Y and Y^*, we derived stronger convergence results of the piecewise affine interpolations. Table 3.2 displays the derived results at one glance.

3.3.2 Convergence of Primal Discrete Solutions

In this section we shall prove that the piecewise constant functions x_ν (as defined in Table 3.1) converge to a function $x : [0,\infty) \to X$ such that (x, p^*) is a solution of (3.4), where p^* is a strong solution of the dual equation (3.18).

It will turn out that the functions $x_\nu(t)$ converge in a fairly weak sense and that in general only little can be said about smoothness of the limit function x. Therefore we will also study particular cases — that are motivated from applications — where stronger convergence and better smoothness properties for x can be achieved. Recall the definition of the piecewise affine interpolations $p_\nu^*(t)$ in Table 3.1.

Theorem 3.3.7. *Assume that p^* is a strong solution of* (3.18) *and that $p_\nu^*(t) \rightharpoonup^* p^*(t)$ for all $t \geq 0$ as in Theorem 3.3.3.*

Then, there exists a mapping $x : [0,\infty) \to X$ such that for each $t \geq 0$ a selection $\nu \mapsto \rho(\nu)$ can be found with

$$\lim_{\nu\to\infty} x_{\rho(\nu)}(t) = x(t) \qquad \textit{w.r.t. } \tau_X. \tag{3.39}$$

Moreover, (x, p^*) is a solution of (3.4) *in the sense of Definition 3.1.1.*

Proof. We will proceed as follows: First (Claim 1), we prove that for each $t \geq 0$ the sequence $\{x_\nu(t)\}_{\nu\in\mathbb{N}}$ is contained in a sequentially τ_X-compact set and construct the function x as pointwise limit of τ_X-convergent subsequences.

Hereafter, we show that the pair (x, p^*) is a solution of (3.4), that is (3.4a) is satisfied for all $t \geq 0$ (Claims 2 and 3) and x is continuous w.r.t. the Bregman-topology τ_X^J on X (Claim 4).

We note that for $t = 0$ there is nothing to show. Thus we assume throughout the proof that $t > 0$ is fixed and we choose $n(\nu) \in \mathbb{N}$ such that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$. In particular, the definition of x_ν (cf. Table 3.1) implies that

$$x_\nu(t) = x_{\nu,n(\nu)}.$$

Claim 1. Existence of the limit in (3.39): From the update rule (2.9b) in Algorithm 2.2.9 we find that

$$K^* p^*_{\nu,n(\nu)} \in \partial J(x_{\nu,n(\nu)}) = \partial J(x_\nu(t)). \tag{3.40}$$

From the definition of the subgradient it hence follows

$$J(x_\nu(t)) \leq J(x) + \left\langle p^*_{\nu,n(\nu)}, K(x_\nu(t) - x)\right\rangle \qquad \text{for all } x \in X \text{ and } \nu \in \mathbb{N}. \tag{3.41}$$

First, Remark 3.3.1 (2) gives for all $\nu \in \mathbb{N}$ and $t \geq 0$

$$\phi\left(\|Kx_\nu(t) - y\|\right) = \|dp_\nu^*(t)\| \leq \phi\left(\|Kx_0 - y\|\right).$$

Thus, due to the monotonicity of ϕ, we end up with

$$\|Kx_\nu(t) - y\| \leq \|Kx_0 - y\| \qquad \text{for all } \nu \in \mathbb{N},\ t \geq 0. \tag{3.42}$$

Further, we point out that due to the fact that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$, it follows that

$$\left|t_{n(\nu)}(\boldsymbol{\alpha}_\nu) - t\right| = t_{n(\nu)}(\boldsymbol{\alpha}_\nu) - t \leq t_{n(\nu)}(\boldsymbol{\alpha}_\nu) - t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu) = \frac{1}{\alpha_{\nu,n(\nu)}} \leq \frac{1}{|\boldsymbol{\alpha}_\nu|}.$$

Since $|\boldsymbol{\alpha}_\nu| \to \infty$, this implies that $t_{n(\nu)}(\boldsymbol{\alpha}_\nu) \to t$ as $\nu \to \infty$ and in particular that $\left\{t_{n(\nu)}(\boldsymbol{\alpha}_\nu)\right\}_{\nu\in\mathbb{N}}$ is bounded, say by $c > 0$. From the Lipschitz continuity of p_ν^* (Remark 3.3.1 (2)) we hence find that

$$\begin{aligned}\left\|p^*_{\nu,n(\nu)}\right\| = \left\|p_\nu^*(t_{n(\nu)}(\boldsymbol{\alpha}_\nu))\right\| &\le \phi\left(\|Kx_0 - y\|\right) t_{n(\nu)}(\boldsymbol{\alpha}_\nu) + \|p_0^*\| \\ &\le c\phi\left(\|Kx_0 - y\|\right) + \|p_0^*\| =: c_1. \end{aligned} \tag{3.43}$$

Combining (3.42) and (3.43) with (3.41) shows that

$$\begin{aligned} J(x_\nu(t)) \le J(x) + \left\langle p^*_{\nu,n(t,\nu)}, K(x_\nu(t) - x)\right\rangle &\le J(x) + c_1 \left\|K(x_\nu(t) - x)\right\| \\ &\le J(x) + c_1\phi\left(\|Kx_0 - y\|\right) + c_1 \left\|Kx - y\right\|. \end{aligned} \tag{3.44}$$

for all $x \in X$. Choose an arbitrary $\bar{x} \in D(J)$ and observe that the previous estimate and (3.42) (note that ψ_ϕ is nondecreasing) imply that

$$\begin{aligned} &\psi_\phi(\|Kx_\nu(t) - y\|) + J(x_\nu(t)) \\ &\qquad \le \psi_\phi(\|Kx_0 - y\|) + J(\bar{x}) + c_1\phi\left(\|Kx_0 - y\|\right) + c_1 \left\|K\bar{x} - y\right\| =: c_2 \end{aligned} \tag{3.45}$$

for all $\nu \in \mathbb{N}$. In other words, this means that $x_\nu(t) \in \Lambda(c_2)$ for all $\nu \in \mathbb{N}$.

According to the compactness requirement (R5), there exists a selection $\nu \mapsto \rho(\nu)$ and an element $x(t) \in X$ such that

$$\lim_{\nu\to\infty} x_{\rho(\nu)}(t) = x(t)$$

w.r.t the topology τ_X. Similarly one proceeds for arbitrary $t \ge 0$ and ends up with a function $x : [0, \infty) \to X$, where we define $x(0) := x_0$.

Since we considered $t > 0$ to be fixed we we shall assume (for the remainder of the proof) that

$$x_\nu(t) \to x(t) \qquad \text{w.r.t. } \tau_X. \tag{3.46}$$

Claim 2. *$dp^*(t) \in \mathfrak{J}_\phi(y - Kx(t))$*: First, we recall from (3.40) that $K^*p^*_{\nu,n(\nu)} \in \partial J(x_{\nu,n(\nu)})$ for all $\nu \in \mathbb{N}$. Thus it follows from Lemma A.2.12 implies that $i_X(x_{\nu,n(\nu)}) \in \partial J^*(K^*p^*_{\nu,n(\nu)})$, where $i_X : X \to X^{**}$ is the natural mapping on X. As in the proof of Proposition 2.2.15 one shows that (note that $i_Y \equiv \mathrm{Id}$)

$$Kx_\nu(t) - y \in \partial F^*(p^*_{\nu,n(\nu)}; y). \tag{3.47}$$

We intend to pass to the limit $\nu \to \infty$ in (3.47) and show that $Kx(t) - y \in \partial F^*(p^*(t); y)$.

To this end, recall that $n(\nu) \in \mathbb{N}$ was chosen such that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$ for all $\nu \in \mathbb{N}$. Therefore it follows from Remark 3.3.1 (3) that

$$\left\|p_\nu^*(t) - p^*_{\nu,n(\nu)}\right\| \le \frac{\phi(\|Kx_0 - y\|)}{|\boldsymbol{\alpha}_\nu|}.$$

Since $|\boldsymbol{\alpha}_\nu| \to \infty$ as $\nu \to \infty$ we conclude together with Theorem 3.3.3 that

$$\underset{\nu\to\infty}{w\text{-lim}}\, p^*_{\nu,n(\nu)} = \underset{\nu\to\infty}{w\text{-lim}}(p^*_{\nu,n(\nu)} - p_\nu^*(t)) + \underset{\nu\to\infty}{w\text{-lim}}\, p_\nu^*(t) = p^*(t). \tag{3.48}$$

Furthermore, the τ_X-τ_Y continuity of K and (3.46) implies that the left hand side of (3.47) converges to $Kx(t)-y$ w.r.t. the topology τ_Y and hence also weakly according to requirement (R1). Finally, we observe (by recursive application of Corollary A.3.4)

$$F^*(p^*_{\nu,n(\nu)};y) \le F^*(p^*_{\nu,n(\nu)-1};y) \le \cdots \le F^*(p^*_{\nu,1};y) \le F^*(p^*_0;y) \quad \text{for all } \nu \in \mathbb{N}.$$

Summarizing, we have that

$$\begin{aligned} &Kx_\nu(t)-y \in \partial F^*(p^*_{\nu,n(\nu)};y), && \sup_{\nu\in\mathbb{N}} F^*(p^*_{\nu,n(\nu)};y) < \infty, \\ &w\text{-}\lim_{\nu\to\infty}(Kx_\nu(t)-y) = Kx(t)-y, && w\text{-}\lim_{\nu\to\infty} p^*_{\nu,n(\nu)} = p^*(t). \end{aligned} \tag{3.49}$$

Since $\partial F^*(p^*;y) = \partial(J^* \circ K^*) - y \subset Y^* \times Y$ is assumed to be weakly-weakly closed (cf. requirement (R8b)) we conclude from (3.49) that

$$Kx(t)-y \in \partial F^*(p^*(t);y). \tag{3.50}$$

Next, we recall that according to the assumptions (R1) and (R2), the norm on Y is sequentially τ_Y-lower semicontinuous and K is τ_X-τ_Y continuous respectively. This implies that

$$\|Kx(t)-y\| \le \liminf_{\nu\to\infty} \|Kx_\nu(t)-y\|.$$

Further Remark 3.3.1 (2) gives that

$$\|Kx_\nu(t)-y\| = \phi^{-1}\left(\|dp^*_\nu(t)\|\right).$$

Combining these two facts with Remark 3.3.4 (1) shows

$$\|Kx(t)-y\| \le \liminf_{\nu\to\infty} \phi^{-1}\left(\|dp^*_\nu(t)\|\right) = \phi^{-1}(\|dp^*(t)\|) = |\partial F^*(\cdot;y)|\,(p^*(t)).$$

This estimate together with (3.50) gives that

$$Kx(t)-y \in \partial^0 F^*(p^*(t);y).$$

Since p^* is a strong solution of (3.18) it finally follows that

$$y - Kx(t) \in -\partial^0 F^*(p^*(t);y) \subset \mathfrak{J}_{\phi^{-1}}(dp^*(t)).$$

We remark, that on reflexive spaces one always has $\mathfrak{J}_\phi^{-1} = \mathfrak{J}_{\phi^{-1}}$ (see e.g. [42, Cor. 3.5]). Thus we end up with

$$dp^*(t) \in \mathfrak{J}_\phi(y - Kx(t))$$

as desired.

Claim 3. $K^*p^*(t) \in \partial J(x(t))$: As already pointed out in (3.40) we have that

$$K^*p^*_{\nu,n(\nu)} \in \partial J(x_\nu(t)) \quad \text{for all } \nu \in \mathbb{N}.$$

As $K^* \in \mathcal{L}(Y^*, X^*)$ is weakly-weakly* continuous (see e.g. [94, Prop. 3.1.15]) we further find (noting (3.48)) that

$$w^*\text{-}\lim_{\nu\to\infty} K^*p^*_{\nu,n(\nu)} = K^*p^*(t).$$

We collect the previous two statements and (3.44) and (3.46):

$$\begin{aligned}
&K^* p^*_{\nu,n(\nu)} \in \partial J(x_\nu(t)), && \sup_{\nu \in \mathbb{N}} J(x_\nu(t)) < \infty, \\
&\lim_{\nu\to\infty} x_\nu(t) = x(t) \ \text{ w.r.t. } \ \tau_X, && w^*\text{-}\lim_{\nu\to\infty} K^* p^*_{\nu,n(\nu)} = K^* p^*(t).
\end{aligned} \tag{3.51}$$

Since ∂J is τ_X-weakly* closed according to requirement (R8a) we eventually get

$$K^* p^*(t) \in \partial J(x(t)).$$

Claim 4. x is continuous w.r.t. τ_X^J: Then it follows from Claim 2 and the definition of $\mathfrak{J}_\phi$ that for all $t \geq 0$

$$\|dp^*(t)\| = \phi(\|Kx(t) - y\|).$$

Thus it follows from Remark 3.3.4 (3) that

$$\phi(\|Kx(t) - y\|) = \|dp^*(t)\| \leq \phi(\|Kx_0 - y\|)$$

Now let $s, t, \geq 0$. Then it follows from the previous estimate that

$$\|K(x(s) - x(t))\| \leq \|Kx(s) - y\| + \|Kx(t) - y\| \leq 2\,\|Kx_0 - y\|\,.$$

and since $K^*p^*(s) \in \partial J^*(x(s))$ (cf. Claim 2) the definition of the subgradient gives

$$J(x(s)) - J(x(t)) \leq \langle K^*p^*(s), x(s) - x(t)\rangle\,.$$

Combining these the last two estimates results in

$$\begin{aligned}
D_J(x(s), x(t)) &\leq D_J^{K^*p^*(t)}(x(s), x(t)) \\
&= J(x(s)) - J(x(t)) - \langle K^*p^*(t), x(s) - x(t)\rangle \\
&\leq \langle K^*p^*(s) - K^*p^*(t), x(s) - x(t)\rangle \\
&\leq \|p^*(s) - p^*(t)\|\,\|K(x(s) - x(t))\| \\
&\leq 2\,\|p^*(s) - p^*(t)\|\,\|Kx_0 - y\| \\
&\leq 2\phi(\|Kx_0 - y\|)\,\|Kx_0 - y\|\,|s - t|\,,
\end{aligned}$$

where the last inequality follows from the Lipschitz continuity of p^* (cf. Theorem 3.3.3). □

Theorem 3.3.7 proves existence of solutions of Equation (3.4). However, the notion of solution in Definition 3.1.1 in general (that is, for general functionals J and operators K) is very weak.

In the remainder of this section we will investigate special cases for which the assertion of Theorem 3.3.7 can be improved. We focus on two special situations: Strict (total) convex functionals J (Theorem 3.3.8) and injective operators K with closed range (Theorem 3.3.10). We note that the first case occurs e.g. in regularization theory (e.g. $J = \|\cdot\|_X$, X strictly convex), whereas the latter case for instance is standard in image denoising (i.e. $K = \mathrm{Id}$; cf. Chapter 4).

Theorem 3.3.8. *Assume that J is strictly convex and that p^* is a strong solution of* (3.18) *with $p^*_\nu(t) \rightharpoonup^* p^*(t)$ for all $t \geq 0$ as in Theorem 3.3.3.*

Then, there exists a unique, weakly measurable function $x : [0,\infty) \to X$ satisfying

$$\lim_{\nu\to\infty} x_\nu(t) = x(t), \qquad \text{w.r.t. } \tau_X$$

for all $t \geq 0$. Moreover, the pair (x, p^*) is a solution of (3.4).

Additionally, if J is totally convex (cf. Definition A.2.9) then the function $t \mapsto x(t)$ is strongly continuous.

Proof. If J is strictly convex, then every $\xi^* \in X^*$ is contained in the subdifferential of at most one element $x \in D(\partial J)$: Assume, by contradiction, that for $x \neq \tilde{x}$ one has $\xi^* \in \partial J(x)$ and $\xi^* \in \partial J(\tilde{x})$. The definition of the subgradient and the strict convexity of J imply

$$J(x) > J(\tilde{x}) + \langle \xi^*, x - \tilde{x}\rangle > J(x) + \langle \xi^*, x - \tilde{x}\rangle + \langle \xi^*, \tilde{x} - x\rangle = J(x),$$

which is of course contradictory.

Now assume, that $x_1, x_2 : [0,\infty) \to X$ are such that (x_1, p^*) and (x_2, p^*) are solutions of (3.4). Then it follows that

$$K^*p^*(t) \in \partial J(x_1(t)) \quad \text{ and } \quad K^*p^*(t) \in \partial J(x_2(t))$$

and thus, according to the considerations above, $x_1(t) = x_2(t)$. From now on we will assume that $x : [0,\infty) \to X$ is the unique function such that (x, p^*) solves (3.4).

We recall from Claim 1 in the proof of Theorem 3.3.7 that for all $t \geq 0$ the sequence $\{x_\nu(t)\}_{\nu\in\mathbb{N}}$ is contained in a τ_X-sequentially precompact set. Moreover, it follow from Claim 3 in the proof of Theorem 3.3.7 that each τ_X-cluster point $\bar{x}$ of $\{x_\nu(t)\}_{\nu\in\mathbb{N}}$ satisfies

$$K^*p^*(t) \in \partial J(\bar{x}).$$

Keeping in mind the argumentation above, this already implies that $\bar{x} = x(t)$ or, in other words, that the set of τ_X cluster points of $\{x_\nu(t)\}_{\nu\in\mathbb{N}}$ is the singleton $\{x(t)\}$ and hence $x_\nu(t) \to x(t)$.

We prove weak measurability of x (recall Definition A.1.6 (2)): Since τ_X is stronger than the weak topology on X by (R1) it follows that for $\xi^* \in X^*$

$$\lim_{\nu\to\infty} \langle \xi^*, x_\nu(t)\rangle = \langle \xi^*, x(t)\rangle \qquad \text{for all } t \geq 0.$$

The numerical functions $t \mapsto \langle \xi^*, x_\nu(t)\rangle$ are simple, real valued functions for all $\nu \in \mathbb{N}$ and thus measurable. Consequently $\langle \xi^*, x\rangle$ is measurable as pointwise limit of measurable, real valued functions (cf. [57, Thm 2.3.2.(6)]) and hence weak measurability of x follows.

Finally, if J is totally convex, it follows from Lemma A.2.10 (2) that continuity of x w.r.t. the Bregman topology already implies continuity in norm. $\square$

We proceed with the second special case announced above, i.e. when the operator K is injective with closed range. Before we do so, we prove

Lemma 3.3.9. Let $T > 0$. For each $n \in \mathbb{N}$ assume that $g_n : [0,T] \to [0,\infty)$ is such

$$\sup_{0\leq t\leq T}\sup_{n\in\mathbb{N}} g_n(t) < \infty \quad \text{ and } \quad \lim_{n\to\infty}\int_0^T g_n(\tau)\,\mathrm{d}\tau = 0.$$

Then one has for all continuous $\omega : [0,\infty) \to [0,\infty)$ with $\omega(0) = 0$ that

$$\lim_{n\to\infty} \int_0^T (\omega \circ g_n)(\tau)\,\mathrm{d}\tau = 0.$$

Proof. Since $\lim_{n\to\infty} \int_0^T g_n(\tau)\,\mathrm{d}\tau = 0$ it follows from [55, Chap. 1.3 Thm. 5] that there exists a selection $n \mapsto \rho(n)$ such that

$$\lim_{n\to\infty} g_{\rho(n)}(t) = 0, \qquad \text{for a.e. } t \in [0,T].$$

Since ω is continuous, this implies that

$$\lim_{n\to\infty} \omega(g_{\rho(n)}(t)) = \omega(0) = 0 \qquad \text{for a.e. } t \in [0,T].$$

Moreover, since there exists a constant $c \geq 0$ such that $g_n(t) \leq c$ for all $n \in \mathbb{N}$ and $0 \leq t \leq T$ we have (again due to continuity) that

$$\sup_{0\leq t\leq T} \sup_{n\in\mathbb{N}} \omega(g_{\rho(n)}(t)) < \infty.$$

Thus, the dominated convergence theorem shows that

$$\lim_{n\to\infty} \int_0^T (\omega \circ g_{\rho(n)})(\tau)\,\mathrm{d}\tau = \int_0^T \lim_{n\to\infty} \omega(g_{\rho(n)}(\tau))\,\mathrm{d}\tau = 0.$$

The assertion follows from a standard sub-subsequence argument. □

Theorem 3.3.10. *Assume that Y is uniformly convex and separable and that Y^* is an E-space and let $1 < p < \infty$. Moreover, let p^* be a strong solution of* (3.18) *with $p^*_\nu(t) \to p^*(t)$ for all $t \geq 0$ as in Corollary 3.3.6.*

If $\ker(K) = \{0\}$ and $\operatorname{ran}(K)$ is closed, there exists a function $x : [0,\infty) \to X$ such that for all $T > 0$, $x \in L^p(0,T;X)$ and

$$\lim_{\nu\to\infty} x_\nu = x$$

in $L^p(0,T;X)$. Moreover, the pair (x,p^*) is a solution of (3.4).

Proof. First, we note that from [42, Thm.2.13] it follows that uniform convexity of Y implies uniform smoothness of Y^* and consequently [42, Thm.2.16] says that the duality mapping $\mathfrak{J}_{\phi^{-1}}$ is single valued and uniformly continuous on bounded sets of Y^* .

In other words, this means that for each $R > 0$ and $p^*_1, p^*_2 \in Y^*$ with $\|p^*_1\| \leq R$ and $\|p^*_2\| \leq R$, there exists a continuous function $\omega_R : [0,\infty) \to [0,\infty)$ such that $\omega_R(s) > 0$ if $s > 0$ and $\omega_R(0) = 0$, for which the equality

$$\left\|\mathfrak{J}_{\phi^{-1}}(p^*_1) - \mathfrak{J}_{\phi^{-1}}(p^*_2)\right\| \leq \omega_R\left(\|p^*_1 - p^*_2\|\right) \tag{3.52}$$

holds (cf. also [4, Chap. 1.6]).

Next, recall from Remark 3.3.4 (3) that

$$\max\left(\|dp^*_\nu(t)\|, \|dp^*(t)\|\right) \leq \phi(\|Kx_0 - y\|) \text{ for all } t \geq 0,\ \nu \in \mathbb{N}.$$

Setting $R := \phi(\|Kx_0 - y\|)$ we get from (3.52) the estimate

$$\left\|\mathfrak{J}_{\phi^{-1}}(dp^*(t)) - \mathfrak{J}_{\phi^{-1}}(dp^*_\nu(t))\right\| \leq \omega_R\left(\|dp^*(t) - dp^*_\nu(t)\|\right) \text{ for all } t \geq 0. \tag{3.53}$$

Furthermore, recall that for each $\nu \in \mathbb{N}$ the sequences $\boldsymbol{x}_\nu = \{x_{\nu,1}, x_{\nu,2}, \ldots\}$ and $\boldsymbol{p}^*_\nu = \{p^*_{\nu,1}, p^*_{\nu,2}, \ldots\}$ are generated by the augmented Lagrangian method (Algorithm 2.2.9) w.r.t. the data y, the initial value p_0^* and the parameters $\boldsymbol{\alpha}_\nu = \{\alpha_{\nu,1}, \alpha_{\nu,2}, \ldots\}$. From the update step (2.9b) in the algorithm it hence follows that for each $n \in \mathbb{N}$

$$Kx_{\nu,n} - y = \mathfrak{J}_{\phi^{-1}}(\alpha_{\nu,n}(p^*_{\nu,n-1} - p^*_{\nu,n})).$$

Keeping in mind the definitions of $x_\nu(t)$ and $dp^*_\nu(t)$ (cf. Table 3.1) this gives

$$Kx_\nu(t) = y - \mathfrak{J}_{\phi^{-1}}(dp^*_\nu(t)) =: z_\nu(t). \tag{3.54}$$

for all $t \geq 0$.

Finally, the injectivity of K and the fact that $\operatorname{ran}(K)$ is closed imply that $K^{-1} : \operatorname{ran}(K) \to X$ is continuous (closed graph theorem; cf. [124, Thm. IV.4.4]). Moreover, from Theorem 3.3.7 it follows that there exists at least one function $x : [0, \infty) \to X$ such that (x, p^*) is a solution of (3.4), which implies that (note that $\mathfrak{J}_{\phi^{-1}} = \mathfrak{J}_\phi^{-1}$)

$$Kx(t) = y - \mathfrak{J}_{\phi^{-1}}(dp^*(t)) =: z(t). \tag{3.55}$$

for all $t > 0$.

So far, it becomes evident from combining (3.53), (3.54) and (3.55) that

$$\begin{aligned}\|x_\nu(t) - x(t)\| &= \left\|K^{-1}(z_\nu(t) - z(t))\right\| \leq \left\|K^{-1}\right\| \|z_\nu(t) - z(t)\| \\ &= \left\|K^{-1}\right\| \left\|\mathfrak{J}_{\phi^{-1}}(dp^*_\nu(t)) - \mathfrak{J}_{\phi^{-1}}(dp^*(t))\right\| \leq \left\|K^{-1}\right\| \omega_R\left(\|dp^*(t) - dp^*_\nu(t)\|\right).\end{aligned} \tag{3.56}$$

Now, let $1 < p < \infty$. From Corollary 3.3.6 (under the present assumptions on Y and Y^*) it follows that

$$\lim_{\nu\to\infty} \int_0^T \|dp^*_\nu(\tau) - dp^*(\tau)\| \,\mathrm{d}\tau = 0.$$

Since the integrand in above equation satisfies

$$\|dp^*_\nu(t) - dp^*(t)\| \leq 2R$$

for $0 \leq t \leq T$ and $\nu \in \mathbb{N}$, we can combine (3.56) with Lemma 3.3.9 and observe (by setting $g_\nu(t) = \|dp^*_\nu(t) - dp^*(t)\|$ and $\omega = \omega_R^p$) that

$$\lim_{\nu\to\infty} \int_0^T \|x_\nu(\tau) - x(\tau)\|^p \,\mathrm{d}\tau = \left\|K^{-1}\right\|^p \int_0^T \omega_R\left(\|dp^*(t) - dp^*_\nu(t)\|\right)^p \,\mathrm{d}\tau = 0. \tag{3.57}$$

□

We summarize the results of this section: Let $y \in Y$ and assume that $p_0^* \in Y^*$ is chosen according to (R6), that is, $K^*p_0^* \in \partial J(x_0)$ for an element $x_0 \in X$. Moreover, let p^* be a strong solution of (3.18) such that $p^*_\nu(t) \rightharpoonup^* p^*(t)$ for all $t \geq 0$ as in Theorem 3.3.3.

Then, there exists a function $x : [0, \infty) \to X$ that is continuous w.r.t. the Bregman topology τ_X^J (cf. Definition A.2.7) such that (x, p^*) is a solution of (3.4). Moreover, x can be chosen as the pointwise limit of subsequences of x_ν.

By imposing additional restrictions on the spaces Y and Y^* as well as on J and K, we derived stronger convergence results of the piecewise constant interpolations and better smoothness properties for x. Table 3.3 displays the derived results at one glance.

Condition	Convergence of x_ν	Properties of x
—	$x_\nu(t) \rightharpoonup x(t)$ for all $t \geq 0$, up to a subsequence	continuous w.r.t. τ_X^J.
J strictly convex	$x_\nu(t) \rightharpoonup x(t)$ for all $t \geq 0$	weakly measurable
J totally convex	$x_\nu(t) \to x(t)$ for all $t \geq 0$	strongly continuous
Y uniformly convex, Y^* E-space, K injective with $\operatorname{ran}(K)$ closed	$x_\nu \to x$ in $\mathrm{L}^p(0,T;X)$ for all $T > 0$ and $1 < p < \infty$	$x \in \mathrm{L}^p(0,T;X)$ for all $T > 0$ and $1 < p < \infty$.

Table 3.3: Convergence and smoothness properties of primal solutions of (3.4)

3.4 Continuous Regularizing Operators

In this section we will introduce the continuous counterparts of the operators $\mathcal{R}_n$ and $\mathcal{R}_n^*$ as defined in Remark 2.2.12. With these operators at hand, we will prove a continuous version of Theorem 2.3.4, that is, an estimate on the asymptotic behavior of the residuals.

Let $p_0^* \in Y^*$ be chosen according to (R6), that is, there exists $x_0 \in X$ such that $K^* p_0^* \in \partial J(x_0)$. Moreover, we shall assume that for each $\nu \in \mathbb{N}$ the sequence $\boldsymbol{\alpha}_\nu = \{\alpha_{\nu,1}, \alpha_{\nu,2}, \ldots\} \subset (0,\infty)$ is a partition of $[0,\infty)$ such that

$$\lim_{\nu\to\infty} |\boldsymbol{\alpha}_\nu| = \infty.$$

Then, it follows from Theorem 3.3.3 that for each $y \in Y$ there exists a Lipschitz continuous function $p^* : [0,\infty) \to Y^*$ with Lipschitz constant $c_L = \phi(\|Kx_0 - y\|)$ such that p^* is a strong solution of (3.18). Assume that for $\nu \in \mathbb{N}$

$$\boldsymbol{p}_\nu^* = \{p_{\nu,1}^*, p_{\nu,2}^*, \ldots\} \subset Y^*$$

is a dual sequence generated by the augmented Lagrangian method (Algorithm 2.2.9) w.r.t. the data y, the initial value p_0^* and the parameters $\boldsymbol{\alpha}_\nu$ and let $p_\nu^* : [0,\infty) \to Y^*$ denote the piecewise affine interpolation of $\boldsymbol{p}_\nu^*$ (cf. Table 3.1). Then, according to Theorem 3.3.3, it is not restrictive to assume that p^* can be written as

$$p^*(t) = w\text{-}\lim_{\nu\to\infty} p_{\rho(\nu)}^*(t)$$

for all $t \geq 0$ and a suitable selection $\nu \to \rho(\nu)$.

Moreover, it follows from Theorem 3.3.7 that there exist at least one function $x : [0,\infty) \to X$ such that the pair (x, p^*) is a solution of (3.4) (in the sense of Definition 3.1.1). These considerations give rise to

Definition 3.4.1. For each $y \in Y$ choose a solution $p^* : [0,\infty) \to Y^*$ of (3.18) as in Theorem 3.3.3 and $x : [0,\infty) \to X$ such that (x, p^*) solves (3.4). Then we define for $t \geq 0$

$$\mathcal{R}_t(y) := x(t) \quad \text{and} \quad \mathcal{R}_t^*(y) := p^*(t).$$

Note that in Definition 3.4.1 we did not necessarily assume that the primal solution x can be approximated by piecewise constant functions as in Theorem 3.3.7. The following result shows that — from a regularization point of view — this additional assumption, however, is not restrictive.

Proposition 3.4.2. *Let $y \in Y$ and assume that $p^*(t) = \mathcal{R}_t^*(y)$ is a strong solution of* (3.18).
If $x_1, x_2 : [0,\infty) \to X$ are such that (x_1, p^*) and (x_2, p^*) are solutions of (3.4) *it follows that*

$$D_J(x_1(t), x_2(t)) = D_J(x_2(t), x_1(t)) = 0 \quad \text{and} \quad \|Kx_1(t) - y\| = \|Kx_2(t) - y\|\,.$$

Additionally, if Y is strictly convex, one has

$$Kx_1(t) = Kx_2(t) \quad \text{and} \quad J(x_1(t)) = J(x_2(t)).$$

Proof. Let $t \geq 0$ and set $\xi^*(t) = K^*p^*(t)$. Then it follows from (3.4) that $\xi^*(t) \in \partial J(x_i(t))$ for $i = 1, 2$ and therefore

$$\begin{aligned} D_J^{\xi^*(t)}(x_1(t), x_2(t)) &= J(x_1(t)) - J(x_2(t)) - \langle \xi^*(t), x_1(t) - x_2(t)\rangle \\ D_J^{\xi^*(t)}(x_2(t), x_1(t)) &= J(x_2(t)) - J(x_1(t)) - \langle \xi^*(t), x_2(t) - x_1(t)\rangle\,. \end{aligned}$$

By adding up this two equations we gain

$$D_J(x_1(t), x_2(t)) + D_J(x_2(t)), x_1(t)) \leq D_J^{\xi^*(t)}(x_1(t), x_2(t)) + D_J^{\xi^*(t)}(x_2(t), x_1(t)) = 0 \qquad (3.58)$$

and therefore both expressions on the left vanish for being nonnegative. The second assertion follows from Equation (3.4) and the definition of $\mathfrak{J}_{\phi^{-1}}$:

$$\|Kx_1(t) - y\| = \phi^{-1}(\|dp^*(t)\|) = \|Kx_2(t) - y\|\,.$$

Assume now, that Y is strictly convex. Since Y is presumed to be reflexive (according to requirement (R7)) we can apply [42, Chap. 2 Cor. 1.5] and conclude that the duality mapping $\mathfrak{J}_\phi$ is single valued. Thus we have that

$$y - Kx_1(t) = \mathfrak{J}_\phi(dp^*(t)) = y - Kx_2(t)$$

and therefore $Kx_1(t) = Kx_2(t)$.
Further, it follows from (3.58) that

$$\begin{aligned} 0 = D_J^{\xi^*(t)}(x_1(t), x_2(t)) &= J(x_1(t)) - J(x_2(t)) - \langle \xi^*(t), x_1(t) - x_2(t)\rangle \\ &= J(x_1(t)) - J(x_2(t)) - \langle p^*(t), Kx_1(t) - Kx_2(t)\rangle\,. \end{aligned}$$

Since $Kx_1(t) = Kx_2(t)$ it follows that $J(x_1(t)) = J(x_2(t))$ for all $t > 0$. □

We move on to a continuous version of Theorem 2.3.4, i.e. we study the asymptotic behavior of the residuals

$$\|K\mathcal{R}_t(y) - y\|$$

for large t. Before we do so, recall that for $F^*(q^*; y) = J^*(K^*q^*) - \langle q^*, y\rangle$ one has that

$$\mu^*(y) := \inf_{q^* \in Y^*} F^*(q^*; y)$$

is finite, whenever y is attainable (cf. Lemma 2.2.14).

Theorem 3.4.3. *Assume that $y \in Y$ is attainable and $\tilde{y} \in Y$. Then for all $t > 0$*

$$\|K\mathcal{R}_t(\tilde{y}) - \tilde{y}\| \le \psi_\phi^{-1}\left(\frac{F^*(p_0^*, y) - \mu^*(y)}{t} + \psi_\phi(\|\tilde{y} - y\|)\right).$$

Proof. Let $t > 0$. For each $\nu \in \mathbb{N}$, assume that $x_\nu : [0, \infty) \to X$ denotes the piecewise constant interpolation of a sequence

$$\boldsymbol{x}_\nu = \{x_{\nu,1}, x_{\nu,2}, \ldots\} \subset X$$

generated by the augmented Lagrangian method w.r.t. the data $\tilde{y}$, the initial value p_0^* and the parameters $\boldsymbol{\alpha}_\nu$. According to Proposition 3.4.2, it is not restrictive to assume that $\mathcal{R}_t(\tilde{y})$ is constructed as in the proof of Theorem 3.3.7; after dropping a subsequence this means

$$\lim_{\nu \to \infty} x_\nu(t) = \mathcal{R}_t(\tilde{y}) \quad \text{w.r.t. } \tau_X.$$

For every $\nu \in \mathbb{N}$ there exists $n(\nu) \in \mathbb{N}$ such that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$. From the definition of x_ν (cf. Table 3.1) it follows that $x_\nu(t) = x_{\nu,n(\nu)}$. Moreover, $t < t_{n(\nu)}(\boldsymbol{\alpha}_\nu)$ for all $\nu \in \mathbb{N}$ and from Theorem 2.3.4 (as well as from the monotonicity of ψ_ϕ^{-1}) it follows (after setting $\delta = \|\tilde{y} - y\|$)

$$\begin{aligned}\|Kx_\nu(t) - \tilde{y}\| &= \left\|Kx_{\nu,n(\nu)} - \tilde{y}\right\| \\ &\le \psi_\phi^{-1}\left(\frac{F^*(p_0^*, y) - \mu^*(y)}{t_{n(\nu)}(\boldsymbol{\alpha}_\nu)} + \psi_\phi(\delta)\right) \le \psi_\phi^{-1}\left(\frac{F^*(p_0^*, y) - \mu^*(y)}{t} + \psi_\phi(\delta)\right).\end{aligned}$$

The assertion follows from the τ_X-lower semicontinuity of the mapping $x \mapsto \|Kx - \tilde{y}\|$. □

We close this section with a characterization of the residuals by means of the variation of $F^*(\cdot; y)$ along the path $\mathcal{R}_t^*(y)$.

Proposition 3.4.4. *Let $y \in Y$. Then the function $t \mapsto \|K\mathcal{R}_t(y) - y\|$ is nonincreasing and bounded by $\|Kx_0 - y\|$. Moreover, for $\eta(s) = s\phi(s)$ the identity*

$$\eta(\|K\mathcal{R}_t(y) - y\|) = -\frac{d}{dt}F^*(\mathcal{R}_t^*(y); y)$$

holds for each $t \ge 0$.

Proof. For $t \ge 0$ we abbreviate

$$x(t) := \mathcal{R}_t(y) \quad \text{and} \quad p^*(t) := \mathcal{R}_t^*(y).$$

Noting that (x, p^*) is a solution of (3.4) it follows that

$$dp^*(t) \in \mathfrak{J}_\phi(y - Kx(t)), \qquad \text{for all } t \ge 0.$$

From the definition of the duality mapping $\mathfrak{J}_\phi$ (cf. Definition A.1.1) we hence find that $\|dp^*(t)\| = \phi(\|Kx(t) - y\|)$. Since ϕ is increasing it follows from Remark 3.3.4 (3) that $t \mapsto \|Kx(t) - y\|$ is nonincreasing and that for all $t \ge 0$

$$\|Kx(t) - y\| \le \|Kx_0 - y\|.$$

Finally, it follows from Remark 3.3.4 (1) and (2) that

$$\begin{aligned}-\frac{d}{dt}F^*(p^*(t); y) &= \phi^{-1}(\|dp^*(t)\|)\,\|dp^*(t)\| \\ &= \|Kx(t) - y\|\,\phi(\|Kx(t) - y\|) = \eta(\|Kx(t) - y\|).\end{aligned}$$

□

3.5 Equations for Data in a Hilbert Space

In this section the case, when Y is a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle_Y$ is studied. Throughout this section we shall assume that $\phi(s) = s$ and note that in this case one has $\mathfrak{J}_\phi = \mathrm{Id}$. Further, we identify Y^* with Y via Riesz' isomorphism and agree on using the notation

$$Y^* = Y \quad \text{ and } \quad \langle p, y\rangle_{Y^*,Y} = \langle p, y\rangle_Y =: \langle p, y\rangle\,.$$

We assume that $p_0 \in Y$ is chosen such that there exists $x_0 \in X$ with $K^*p_0 \in \partial J(x_0)$. For given data $y \in Y$, the evolution equation (3.4) in the current setting takes the following form:

$$dp(t) = y - Kx(t) \qquad\qquad K^*p(t) \in \partial J(x(t)) \tag{3.59a}$$

$$p(0) = p_0 \tag{3.59b}$$

and the corresponding equation for the dual variable (3.18) comes as

$$dp(t) = \partial^0 F^*(p(t); y), \quad p(0) = p_0. \tag{3.60}$$

We recall that for $F^*(p; y) = J^*(K^*p) - \langle p, y\rangle$, the symbol $\partial^0 F^*(p; y)$ denotes the minimal section of the subdifferential $\partial F^*(\cdot; y)$ at $p \in Y$ (cf. Definition A.2.1). That is, it contains those subgradients of $F^*(p; y)$ with minimal norm. Since the norm on Y is strictly convex, it follows that $\partial^0 F^*(p; y)$ is a singleton for all $p \in Y$.

This section is organized as follows: We start with a classical existence and uniqueness result for (3.60) that refines Theorem 3.3.3. Moreover, we shall present optimal approximation estimate for the implicit Euler scheme of (3.60), that is, the proximal point method (Algorithm 2.4.2).

With this result at hand, a similar estimate for the scheme generated by the augmented Lagrangian method for (3.59) follows. Finally, a continuous counterpart to Theorem 2.4.4 is given, stating that solutions of (3.59) constitute a regularization method for the constrained problem (2.5).

We start with the following fundamental result (compare with Theorem 3.3.3)

Theorem 3.5.1. *Let $y \in Y$. Then, there exists a unique strong solution $p : [0,\infty) \to Y$ of (3.60) that is Lipschitz continuous with Lipschitz constant $c_L = \|Kx_0 - y\|$.*

Proof. Since we have that $K^*p_0 \in \partial J(x_0)$ it follows that $x_0 \in \partial J^*(K^*p_0)$ and consequently that

$$Kx_0 - y \in \partial(J^* \circ K^*)(p_0) - y = \partial F^*(p_0; y) \tag{3.61}$$

according to Lemma A.2.5. In other words, this means $p_0 \in D(\partial F^*(\cdot; y))$ and it thus follows from [26, Thm 3.1] that there exists a unique strong solution p of (3.60) that is Lipschitz on $[0,\infty)$. From Theorem 3.3.3 it follows that the Lipschitz constant can be chosen as $c_L = \phi(\|Kx_0 - y\|) = \|Kx_0 - y\|$. □

Remark 3.5.2. Recall from (3.61) that $Kx_0 - y \in \partial F^*(p_0; y)$. This in particular implies that the slope $|\partial F^*(\cdot; y)|$ at p_0 is finite:

$$|\partial F^*(\cdot; y)|\,(p_0) = \left\|\partial^0 F^*(p_0; y)\right\| \le \|Kx_0 - y\|\,.$$

Corollary 3.5.3. Let $y \in Y$ and p be the unique solution of (3.60). Then there exists at least one function $x : [0, \infty) \to X$ such that (x, p) solves (3.59) in the sense of Definition 3.1.1. If $(\tilde{x}, p)$ is another solution of (3.59), then

$$x(t) - \tilde{x}(t) \in \ker(K) \quad \text{and} \quad J(x(t)) = J(\tilde{x}(t))$$

for all $t \geq 0$.

Proof. From Theorem 3.3.7 that a solution (x, p) of (3.59) exists and since Y is strictly convex, it follows from Theorem 3.4.2 that $Kx(t) = K\tilde{x}(t)$ and $J(x(t)) = J(\tilde{x}(t))$. □

Remark 3.5.4. From Theorem 3.5.1 it follows that for all $t \geq 0$ the operator $\mathcal{R}_t^* : Y \to Y$ (as in Definition 3.4.1) is uniquely determined.

Moreover, according to Corollary 3.5.3, we find that the operator $\mathcal{R}_t : Y \to X$ is uniquely defined up to elements in $\ker(K)$.

We now focus on the approximation error of the augmented Lagrangian method (cf. Algorithm 2.4.1 for the Hilbert version) when considered as an implicit time scheme for (3.59). To this end, assume that for all $\nu \in \mathbb{N}$ a sequence of positive parameters

$$\boldsymbol{\alpha}_\nu = \{\alpha_{\nu,1}, \alpha_{\nu,2}, \ldots\} \subset (0, \infty)$$

is given, such that

$$\lim_{\nu \to \infty} |\boldsymbol{\alpha}_\nu| = \lim_{\nu \to \infty} \inf_{n \in \mathbb{N}} \alpha_{\nu,n} = \infty.$$

We start with an error estimate for the dual solution of (3.60). Since $Y = Y^*$ is assumed to be a separable Hilbert space, it follows that Y^* is already an E-space. Thus the assumptions of Corollary 3.3.6 are met and we already conclude, that the piecewise affine interpolations of dual sequences generated by Algorithm 2.4.1 converge locally uniformly to the unique solution of (3.60). The following result, recently established by Nochetto et al. in [102], provides an error estimate for this limit.

Theorem 3.5.5. *Let $y \in Y$ and assume that for each $\nu \in \mathbb{N}$*

$$\boldsymbol{p}_\nu = \{p_{\nu,1}, p_{\nu,2}, \ldots\} \subset Y$$

is the dual sequence generated by the augmented Lagrangian method (Algorithm 2.4.1) w.r.t. the data y, the initial value p_0 and the parameters $\boldsymbol{\alpha}_\nu$. Moreover, assume that $p_\nu : [0, \infty) \to Y$ denotes that piecewise affine interpolation of the sequence $\boldsymbol{p}_\nu$ (cf. Table 3.1). Then

$$\|\mathcal{R}_t^*(y) - p_\nu(t)\| \leq \frac{|\partial F^*(\cdot; y)|(p_0)}{\sqrt{2}\,|\boldsymbol{\alpha}_\nu|}, \quad t \geq 0, \ \nu \in \mathbb{N}.$$

Proof. [102, Thm. 3.20] □

As it becomes evident from Theorem 3.3.7, convergence of the piecewise constant interpolations of primal sequence generated by the augmented Lagrangian method in general (i.e. without any restrictions on J or K) has to be understood in a fairly weak sense. Hence, it is not realistic to expect norm-error estimates for the primal variables in the spirit of the previous Theorem.

However, with Theorem 3.5.5 at hand, a convergence estimate for the primal variable in terms of the distance measure

$$\Delta_J(x_1, x_2) := D_J(x_1, x_2) + D_J(x_2, x_1). \tag{3.62}$$

follows:

Proposition 3.5.6. *Let $y \in Y$ and assume that for each $\nu \in \mathbb{N}$*

$$\boldsymbol{x}_\nu = \{x_{\nu,1}, x_{\nu,2}, \ldots\} \subset X$$

is the primal sequence generated by the augmented Lagrangian method (Algorithm 2.4.1) w.r.t. the data y, the initial value p_0 and the parameters $\boldsymbol{\alpha}_\nu$. Moreover, assume that $x_\nu : [0,\infty) \to Y$ denotes that piecewise constant interpolation of the sequence $\boldsymbol{x}_\nu$ (cf. Table 3.1). Then

$$\Delta_J(\mathcal{R}_t(y), x_\nu(t)) \leq \frac{\sqrt{2}\, \|Kx_0 - y\|^2}{|\boldsymbol{\alpha}_\nu|}, \quad t \geq 0,\ \nu \in \mathbb{N}. \tag{3.63}$$

Proof. Assume that $y \in Y$ and let $t \geq 0$. We agree upon the abbreviations

$$x(t) := \mathcal{R}_t(y) \quad \text{and} \quad p(t) := \mathcal{R}_t^*(y).$$

Moreover, we assume that for each $\nu \in \mathbb{N}$

$$\boldsymbol{p}_\nu = \{p_{\nu,1}, p_{\nu,2}, \ldots\} \subset Y$$

is the dual sequence generated by the augmented Lagrangian method (Algorithm 2.4.1) w.r.t. the data y, the initial value p_0 and the parameters $\boldsymbol{\alpha}_\nu$.

For $\nu \in \mathbb{N}$ assume that $n(\nu) \in \mathbb{N}$ is such that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$. From the augmented Lagrangian Algorithm 2.4.1 and from the definition of the piecewise constant interpolation x_ν (cf. Table 3.1) it follows that for all $\nu \in \mathbb{N}$.

$$K^* p_{\nu,n(\nu)} \in \partial J(x_{\nu,n(\nu)}) = \partial J(x_\nu(t)).$$

Moreover, since (x, p) is a solution of (3.59) we find that $K^*p(t) \in \partial J(x(t))$. Combining these two facts yields

$$\begin{aligned}
\Delta(x(t), x_\nu(t)) &= D_J(x(t), x_\nu(t)) + D_J(x_\nu(t), x(t)) \\
&\leq D_J^{K^* p_{\nu,n(\nu)}}(x(t), x_\nu(t)) + D_J^{K^* p(t)}(x_\nu(t), x(t)) \\
&= \left\langle K^* p_{\nu,n(\nu)}, x(t) - x_\nu(t) \right\rangle + \langle K^* p(t), x_\nu(t) - x(t) \rangle \\
&= \left\langle p_{\nu,n(\nu)} - p(t), Kx(t) - y \right\rangle - \left\langle p_{\nu,n(\nu)} - p(t), Kx_\nu(t) - y \right\rangle \\
&\leq \left\| p_{\nu,n(\nu)} - p(t) \right\| (\|Kx(t) - y\| + \|Kx_\nu(t) - y\|).
\end{aligned} \tag{3.64}$$

Further, note that from Remark 3.3.1 (2) and from Proposition 3.4.4 it follows that

$$\|Kx_\nu(t) - y\| \leq \|Kx_0 - y\| \quad \text{and} \quad \|Kx(t) - y\| \leq \|Kx_0 - y\|$$

for all $\nu \in \mathbb{N}$ and $t \geq 0$. Combining this with (3.64) results in

$$\Delta(x(t), x_\nu(t)) \leq 2 \|p_{\nu,n} - p(t)\| \, \|Kx_0 - y\| \,.$$

Assume that for $\nu \in \mathbb{N}$, the function $p_\nu(t)$ denotes the piecewise affine interpolation of $\boldsymbol{p}_\nu$. Then it follows from the construction of p_ν that $p_\nu(t_{n(\nu)}(\boldsymbol{\alpha})) = p_{\nu,n(\nu)}$. Therefore the previous estimate together with Theorem 3.5.5 shows

$$\Delta(x(t), x_\nu(t)) \leq 2 \left\| p_\nu(t_{n(\nu)}(\boldsymbol{\alpha})) - p(t) \right\| \|Kx_0 - y\| \leq \frac{\sqrt{2}\,\|Kx_0 - y\| \, |\partial F^*(\cdot; y)|\,(p_0)}{|\boldsymbol{\alpha}_\nu|}.$$

Finally, we find from Remark 3.5.2 that $Kx_0 - y \in \partial F^*(p_0; y)$ and therefore

$$|\partial F^*(\cdot; y)|\,(p_0) \leq \|Kx_0 - y\|$$

Combination of the last two estimates shows the assertion. □

We close this section with the continuous counterpart of Theorem 2.4.4, stating that under a suitable parameter choice $\{\mathcal{R}_t\}_{t\geq 0}$ constitutes a family of regularization operators for (2.4) (c.f. Definition 2.2.4).

Theorem 3.5.7. *Let $y \in Y$ be attainable and $\{y_n\}_{n\in\mathbb{N}} \subset Y$ such that $\delta_n := \|y - y_n\| \to 0$ as $n \to \infty$ and assume further that $\Gamma : (0, \infty) \times Y \to [0, \infty)$ is such that*

$$\lim_{n\to\infty} \delta_n^2 \Gamma(\delta_n, y_n) = 0 \quad \text{and} \quad \lim_{n\to\infty} \Gamma(\delta_n, y_n) = \infty. \tag{3.65}$$

Then $\left(\{\mathcal{R}_t\}_{t\geq 0}, \Gamma\right)$ is a regularization method for (2.4) such that

$$\lim_{n\to\infty} J(\mathcal{R}_{\Gamma(\delta_n, y_n)}(y_n)) = \inf_{v\in X} \{J(v) \ : \ Kv = y\}.$$

Proof. Let $\tilde{y} \in Y$. We define $\delta := \|y - \tilde{y}\|$ and set

$$x(t) := \mathcal{R}_t(\tilde{y}) \quad \text{and} \quad p(t) := \mathcal{R}_t^*(\tilde{y}).$$

We will essentially follow the proof of Theorem 2.4.4 and start with a preliminary estimate for the values of J along the curve $x(t)$. To this end, assume that $\bar{x} \in X$ is a J-minimizing solution of (2.4) (with right hand side y) and choose $t \geq 0$.

Since (x, p) solve (3.59) it follows that $K^*p(t) \in \partial J(x(t))$ and therefore the definition of the subgradient implies that

$$J(x(t)) \leq J(\bar{x}) + \langle K^*p(t), x(t) - \bar{x}\rangle = J(\bar{x}) + \langle p(t), Kx(t) - y\rangle. \tag{3.66}$$

Next, after introducing the abbreviation

$$c := F^*(p_0^*; y) - \mu^*(y) = F^*(p_0^*; y) - \inf_{p\in Y} F^*(p; y)$$

we find from Theorem 3.4.3 (note that $\psi_\phi^{-1}(s) = \sqrt{2s}$ and $\mathcal{R}_t(\tilde{y}) = x(t)$) that

$$\|Kx(t) - y\| \leq \|y - \tilde{y}\| + \|Kx(t) - \tilde{y}\| \leq \delta + \sqrt{\frac{2c}{t} + \delta^2}. \tag{3.67}$$

For each $\nu \in \mathbb{N}$ let

$$\boldsymbol{p}_\nu = \{p_{\nu,1}, p_{\nu,2}, \ldots\} \subset Y$$

denote the dual sequence generated by the augmented Lagrangian method (Algorithm 2.4.1) w.r.t. the data $\tilde{y}$, the initial value p_0 and the positive parameters $\boldsymbol{\alpha}_\nu$ and denote with $p_\nu(t)$ the piecewise affine interpolation of $\boldsymbol{p}_\nu$ at time t.

Then we find from Theorem 3.5.1 that $p_\nu(t) \to p(t)$ strongly. If $n(\nu) \in \mathbb{N}$ is such that $t \in (t_{n(\nu)-1}(\boldsymbol{\alpha}_\nu), t_{n(\nu)}(\boldsymbol{\alpha}_\nu)]$, it consequently follows from Remark 3.3.1 (3) that

$$\lim_{\nu\to\infty} p_{\nu,n(\nu)} = p(t). \tag{3.68}$$

For $\varepsilon > 0$ choose $p^\varepsilon \in Y$ such that $F^*(p^\varepsilon; y) \le \mu^*(y) + \varepsilon$. Then we conclude from (2.46) that

$$\frac{\left\|p^\varepsilon - p_{\nu,n(\nu)}\right\|}{2\sqrt{t_{n(\nu)}(\alpha_\nu)}} \le \sqrt{\frac{\|p^\varepsilon - p_0\|^2}{2t_{n(\nu)}(\alpha_\nu)} + t_{n(\nu)}(\alpha_\nu)\delta^2 + \varepsilon},$$

This estimate together with (3.68) and the fact that $t_{n(\nu)}(\alpha_\nu) \to t$ as $\nu \to \infty$ gives

$$\frac{\|p^\varepsilon - p(t)\|}{2\sqrt{t}} = \lim_{\nu\to\infty} \frac{\left\|p^\varepsilon - p_{\nu,n(\nu)}\right\|}{2\sqrt{t_{n(\nu)}(\alpha_\nu)}} \le \sqrt{\frac{\|p^\varepsilon - p_0\|^2}{2t} + t\delta^2 + \varepsilon}.$$

From (3.66) we find by using (3.67) and the previous inequality

$$\begin{aligned} J(x(t)) &\le J(\bar{x}) + \langle p(t) - p^\varepsilon, Kx(t) - y\rangle + \langle p^\varepsilon, Kx(t) - y\rangle \\ &\le J(\bar{x}) + \frac{\|p(t) - p^\varepsilon\|}{2\sqrt{t}} 2\sqrt{t}\,\|Kx(t) - y\| + \|p^\varepsilon\|\,\|Kx(t) - y\| \\ &\le J(\bar{x}) + \left(2\sqrt{t}\delta + 2\sqrt{2c + t\delta^2}\right)\sqrt{\frac{\|p^\varepsilon - p_0\|^2}{2t} + t\delta^2 + \varepsilon} \\ &\quad + \|p^\varepsilon\|\left(\delta + \sqrt{\frac{2c}{t} + \delta^2}\right). \end{aligned} \tag{3.69}$$

With these preparations we will now conclude the proof: set $\tau(n) = \Gamma(\delta_n, y_n)$ for each $n \in \mathbb{N}$. Then, according to (3.65) one has $\tau(n) \to \infty$ and $\delta_n^2\tau(n) \to 0$ as $n \to \infty$. This shows together with (3.69) that

$$\limsup_{n\to\infty} J(\mathcal{R}_{\tau(n)}(y_n)) \le J(\bar{x}) + 2\sqrt{2c\varepsilon}.$$

Since $\varepsilon > 0$ is arbitrary, we eventually find

$$\limsup_{n\to\infty} J(\mathcal{R}_{\tau(n)}(y_n)) \le J(\bar{x}). \tag{3.70}$$

In particular, this implies that $J(\mathcal{R}_{\tau(n)}(y_n))$ is uniformly bounded. Moreover, we find from Theorem 3.4.3 that for each $n \in \mathbb{N}$

$$\frac{1}{2}\left\|K\mathcal{R}_{\tau(n)}(y_n) - y\right\|^2 \le \|y - y_n\|^2 + \left\|K\mathcal{R}_{\tau(n)}(y_n) - y_n\right\|^2 \le \delta_n^2 + \frac{2c}{\tau(n)} + 2\delta_n^2$$

Since $\tau(n) \to \infty$ and $\delta_n \to 0$ as $n \to \infty$, this implies that

$$\lim_{n\to\infty} \left\|K\mathcal{R}_{\tau(n)}(y_n) - y\right\| = 0. \tag{3.71}$$

In particular, we have that

$$\sup_{n\in\mathbb{N}}\left(\frac{1}{2}\left\|K\mathcal{R}_{\tau(n)}(y_n)-y\right\|^2+J(\mathcal{R}_{\tau(n)}(y_n))\right)=: c_1<\infty$$

or in other words, $\mathcal{R}_{\tau(n)}(y_n)\in\Lambda(c_1)$ and thus according to assumption (R5) we can assume (possibly after dropping a subsequence) that $\mathcal{R}_{\tau(n)}(y_n)\to\hat{x}$ w.r.t. τ_X for an element $\hat{x}\in X$. Since $x\mapsto\|Kx-y\|$ is sequentially τ_X-lower semicontinuous (cf. Remark 2.1.2 (2)), we finally conclude from (3.71) that

$$\|K\hat{x}-y\|\le\liminf_{n\to\infty}\left\|K\mathcal{R}_{\tau(n)}(y_n)-y\right\|=0$$

and thus $K\hat{x}=y$. Together with the τ_X-lower semicontinuity of J and (3.70) we conclude that each τ_X-cluster point of $\left\{\mathcal{R}_{\tau(n)}(y_n)\right\}_{n\in\mathbb{N}}$ is a J-minimizing solution of (2.4) and the assertion is shown. □

3.6 Example: Quadratic Regularization Revisited

We revisit the special case where J is a quadratic functional, as it has been studied in Section 2.5. We will therefore assume the same setting; in particular, this means that X and Y are Hilbert spaces and

$$J(x)=\begin{cases}\frac{1}{2}\|L(x)\|^2 & \text{if } x\in D(L),\\ +\infty & \text{else,}\end{cases}$$

where $L: D(L)\subset X\to H$ is a linear, closed and densely defined operator, mapping X into a further Hilbert space H. In general, it is not required that L is bounded on $D(L)$. We furthermore recall that $D(\partial J)=D(L^*L)$ and

$$\partial J(x)=\begin{cases}L^*L(x) & \text{if } x\in D(L^*L),\\ \emptyset & \text{else,}\end{cases}$$

according to Lemma 2.5.1. In order to keep the presentation transparent, we shall assume that $p_0=0=L^*Lx_0$ for a $x_0\in\ker(L^*L)$.

Let $y\in Y$ be given. Then the functions $x(t):=\mathcal{R}_t(y)$ and $p(t):=\mathcal{R}_t^*(y)$ solve

$$dp(t)=y-Kx(t),\qquad K^*p(t)=L^*Lx(t),\tag{3.72a}$$
$$p(0)=0.\tag{3.72b}$$

We remark, that from the continuity of K^* it becomes evident that each x is also a solution of the equation

$$\frac{d}{dt}(L^*Lx(t))=K^*(y-Kx(t)),\quad L^*Lx(0)=0.$$

For a given sequence $\boldsymbol{\alpha}=\{\alpha_n\}_{n\in\mathbb{N}}\subset(0,\infty)$, the implicit scheme for (3.72) defined by the augmented Lagrangian method (cf. Algorithm 2.4.1) reads as

$$x_n\in\operatorname*{argmin}_{x\in X}\|Kx-y\|^2+\alpha_n\|L(x-x_{n-1})\|^2\tag{3.73a}$$
$$p_n=p_{n-1}+\alpha_n^{-1}(y-Kx_n)=L^*Lx_n.\tag{3.73b}$$

From Proposition 3.5.6 we gain an estimate for the approximation error of the implicit scheme (3.73). As in the previous section, assume that for $\nu \in \mathbb{N}$

$$\boldsymbol{\alpha}_\nu = \{\alpha_{\nu,1}, \alpha_{\nu;2}, \ldots\} \subset Y$$

is a sequence of partitions of $[0,\infty)$ such that $|\boldsymbol{\alpha}_\nu| \to \infty$ as $\nu \to \infty$. For the quadratic functional J we find for $x_1, x_2 \in X$ that

$$D_J(x_\nu(t), x(t)) = D_J(x(t), x_\nu(t)) = \|L(x_\nu(t) - x(t))\|^2 .$$

Hence the definition of Δ_J in (3.62) together with Proposition 3.5.6 gives

Corollary 3.6.1. Let $y \in Y$ and assume that for each $\nu \in \mathbb{N}$

$$\boldsymbol{x}_\nu = \{x_{\nu,1}, x_{\nu,2}, \ldots\} \subset X$$

is the primal sequence generated by the augmented Lagrangian method (Algorithm 2.4.1) w.r.t. the data y, the initial value p_0 and the parameters $\boldsymbol{\alpha}_\nu$. Moreover, assume that $x_\nu : [0,\infty) \to Y$ denotes that piecewise constant interpolation of the sequence $\boldsymbol{x}_\nu$ (cf. Table 3.1). Then,

$$\|L(x_\nu(t) - x(t))\|^2 \le \frac{\|Kx_0 - y\|^2}{\sqrt{2}\,|\boldsymbol{\alpha}_\nu|}.$$

In Section 2.5.2 we studied the special case, when $K \equiv \mathrm{Id}$. In this situation, the evolution equation (3.72) is equivalent to

$$\frac{d}{dt}(L^* L x(t)) = y - x(t), \quad L^* L x(0) = 0.$$

Moreover, we have that solutions are unique due to Proposition 3.4.2. We close this Section with the following example:

Example 3.6.2. Let $\Omega \subset \mathbb{R}^N$ be a open and bounded domain with smooth boundary $\partial\Omega$ and $X = \mathrm{L}^2(\Omega)$ as well as $H = \mathrm{L}^2(\Omega, \mathbb{R}^N)$ and set $L = \nabla$ with $\mathrm{H}^1(\Omega) = D(L) \subset \mathrm{L}^2(\Omega)$. Then L is linear, closed and densely defined (w.r.t. the L^2-topology). Moreover we find from [17, pp.63] that

$$D(L^* L) = D(\partial J) = \left\{x \in \mathrm{H}^2(\Omega) \ : \ \nabla x \cdot \nu = 0, \ \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega\right\}$$

and $\partial J(x) = -\Delta x$ (ν denotes the outer unit normal vector on $\partial\Omega$). The evolution equation (3.72) thus turns to the third order equation

$$\begin{aligned} \frac{d}{dt}\Delta x(t,s) &= x(t,s) - y(s) \quad \text{for all } s \in \Omega, && (3.74a)\\ \nabla x(s) \cdot \nu(s) &= 0 \quad \text{for } \mathcal{H}^{N-1}\text{-a.e. } s \in \partial\Omega, && (3.74b)\\ \Delta x(0,s) &= 0. && (3.74c) \end{aligned}$$

We show that the unique solution x of (3.74) is continuous (in time). To this end, we note that by Green's formula

$$\int_\Omega x(t,s) - y(s)\,\mathrm{d}s = \int_\Omega \frac{d}{dt}\Delta x(t,s)\,\mathrm{d}s = \frac{d}{dt}\int_{\partial\Omega} \nabla x(t,s) \cdot \nu\,\mathrm{d}s = 0. \tag{3.75}$$

Hence it follows that

$$\int_\Omega x(t_1,s) - x(t_2,s)\,\mathrm{d}s = 0$$

for all $t_1, t_2 \geq 0$. Consequently there exists a embedding constant $C = C(\Omega)$ such that (cf. [125, Thm.4.2.1])

$$\|x(t_1) - x(t_2)\|_{\mathrm{L}^2} \leq C\,\|\nabla(x(t_1) - x(t_2))\|_{\mathrm{L}^2(\Omega,\mathbb{R}^N)} = CD_J(x(t_1), x(t_2)).$$

Thus norm continuity follows from continuity w.r.t. the Bregman topology. In particular, the continuity of x and (3.75) imply

$$\int_\Omega x(0,s)\,\mathrm{d}s = \int_\Omega \lim_{t\to 0^+} x(t,s)\,\mathrm{d}s = \lim_{t\to 0^+} \int_\Omega x(t,s)\,\mathrm{d}s = \int_\Omega y(s)\,\mathrm{d}s.$$

Since $\Delta x(0,s) = 0$ and $\nabla x(0,s)\cdot \nu(s) = 0$ on $\partial\Omega$ this results in

$$x(0,s) = \frac{1}{\lambda^N(\Omega)} \int_\Omega y(s)\,\mathrm{d}s =: \bar{y}.$$

Next, we apply the error estimate in Corollary 3.6.1 to the present example. To do this, assume that $\{\alpha_n\}_{n\in\mathbb{N}}$ is a sequence of positive parameters and let $\{x_n(s)\}_{n\in\mathbb{N}} \subset D(L^*L)$ ($s \in \Omega$) be the sequence generated by the augmented Lagrangian method (3.73) (w.r.t. the data y and initial value $x_0 = \bar{y}$). This shows that for all $n \geq 1$

$$y(s) - x_n(s) = \alpha_n \Delta(x_{n-1}(s) - x_n(s)).$$

Again, by Green's formula one finds as in (3.75)

$$\int_\Omega y(s) - x_n(s)\,\mathrm{d}s = 0.$$

If for $\nu \in \mathbb{N}$ $\boldsymbol{\alpha}_\nu$, $\boldsymbol{x}_\nu$ and $x_\nu(t)$ are such as in Corollary 3.6.1, then it follows that

$$\int_\Omega x_\nu(t,s) - x(t,s)\,\mathrm{d}s = \int_\Omega x_\nu(t,s) - y(s)\,\mathrm{d}s + \int_\Omega y(s) - x(t,s)\,\mathrm{d}s = 0$$

for all $t \geq 0$. Consequently, we get with the same embedding constant C as above and Corollary 3.6.1

$$\|x_\nu(t) - x(t)\|_{\mathrm{L}^2}^2 \leq C\,\|\nabla(x_\nu(t) - x(t))\|_{\mathrm{L}^2(\Omega,\mathbb{R}^N)}^2 \leq \frac{C\int_\Omega |y(s) - \bar{y}|^2\,\mathrm{d}s}{\sqrt{2}\,|\boldsymbol{\alpha}_\nu|}.$$

3.7 Notes

In this Chapter we proved the existence of solutions of the evolution equation (3.4), where a solution consists of a pair of curves representing a primal and a dual variable (cf. Definition 3.1.1). Additionally regularizing properties of these solutions w.r.t. the constrained problem (2.5) were considered.

Equation of this type first appeared — at least in the context of image processing — in Burger et al. [33] (see also [31]), however, without any existence results for solutions. Existence analysis was given in Burger et al. [30], and in a more general setting by F. & Scherzer in [61]. Therein it was shown, that Equation (3.4) *decouples* in the sense that there always exist solution pairs whose dual component solves the differential inclusion (3.18), or in other words:

> Proving existence of solutions for (3.4) boils down to finding solutions of the dual equation (3.18) and a subsequent construction of a primal curve such that (3.4) holds.

Since the dual equation (3.18) comes in the shape of a *gradient flow* (or steepest descent) equation, solutions can be constructed via the implicit Euler scheme, which coincides in the present case, with the proximal point algorithm (Algorithm 2.2.16) studied in Chapter 2.

For the general (Banach space) case, we proved convergence in Theorem 3.3.3 making extensive use of the analysis in the recent book by Ambrosio et al. [9] (in particular, in Section A.3 and in the proof of Theorem 3.3.3). By requiring additional properties of the underlying spaces we derived improved convergence results for the dual sequence, which are listed in Table 3.2.

In fact, the analysis in [9] reaches far beyond the situation considered in this thesis: it provides a generalized approach for solving gradient flow equations in metric spaces by means of *curves of maximal slope* (originally introduced by de Giorgi et al. [46]), which turn out to be (strong) solutions of differential inclusions if consideration is restricted to (well-behaved) Banach spaces. Additionally, a generalization of the implicit Euler scheme for evolution equations as well as its convergence to curves of maximal slope is given within the framework of *(generalized) minimizing movements* which goes back to de Giorgi [45] (see also Ambrosio [8] or Gianazza & Savaré [64, 65]).

If the dual equation (3.18) is considered on a Hilbert space, the problem of finding solutions lies within the framework of *nonlinear semigroups* and considerably stronger results (concerning uniqueness, convergence and smoothness of solutions) are available (cf. Theorems 3.5.1 and 3.5.5 in Section 3.5). An exhaustive list of references on this vast topic lies beyond the scope of this section; thus we merely mention: Ambrosio et al [9], Barbu [17], Brézis [26], Crandall & Liggett [44], Miyadera [99] or Pavel [104].

$$\begin{tikzcd}
\text{Evolution Equation (3.4)} & \text{Dual Equation (3.18)} \arrow[l, "\text{Thm. 3.3.7}"'] \\
\text{Augmented Lagrangian Method (Algorithm 2.2.9)} \arrow[u, dotted] \arrow[r, "\text{Prop. 2.2.15}"'] & \text{Proximal Point Algorithm (Algorithm 2.2.16)} \arrow[u, "\text{Thm. 3.3.3 (Thm. 3.5.1)}"']
\end{tikzcd}$$

Figure 3.1: Methodology for constructing solutions of (3.4)

Finally, in Theorem 3.3.7 we have shown that each solution of (3.18) can be complemented to a solution of (3.4). To this end, a primal curve was constructed as as a pointwise limit of discrete (piecewise constant) functions generated by the augmented Lagrangian method (Algorithm 2.2.9 in Chapter 2). We considered special settings in order to come up with stronger convergence and smoothness results. Table 3.3 subsumes the derived results.

Figure 3.1 schematically summarizes the methodology for solving (3.4) used in this chapter.

4 Application: Image Denoising

In this chapter we return to the motivating example introduced in Chapter 1; the iterative image denoising method introduced by Osher et al. in [103]. We briefly recall:

Given a (noisy) image $f \in \mathrm{L}^2(\Omega)$ (defined on a image domain $\Omega \subset \mathbb{R}^2$), the authors propose the following algorithm to recover the underlying, noise-free image:

1. Set $v_0 := 0$.
2. For $n = 1, 2, \dots$ compute

$$\begin{aligned} u_{n+1} &= \operatorname*{argmin}_{u \in \mathrm{L}^2(\Omega)} \left\{ \frac{1}{2} \int_\Omega |f-u|^2 \, \mathrm{d}x + \alpha D_J^{v_n}(u, u_n) \right\}, \\ v_{n+1} &= v_n + \alpha^{-1}(f - u_{n+1}). \end{aligned}$$

Here, J denotes a convex and L^2-lower semicontinuous regularization functional (the BV-semi norm; cf. Section 4.1) and α a fixed positive parameter. $D_J^{v_n}(u, u_n)$ is the Bregman distance of u and u_n w.r.t. J and v_n (cf. Definition A.2.6)

$$D_J^{v_n}(u, u_n) = J(u) - J(u_n) - \int_\Omega v_n(u - u_n) \, \mathrm{d}x.$$

Later on, Burger et al. claimed in [31, 33] that the pair (u_n, v_n) can be considered as an approximation of solutions (u, v) of the equation system

$$\begin{aligned} dv(t) &= f - u(t), & v(t) &\in \partial J(u(t)), \\ v(0) &= 0. \end{aligned}$$

at time $t_n = \alpha^{-1} n$. However, neither existence of such solutions nor approximation estimates were given.

Recalling Algorithm 2.4.1, it becomes evident, that the above iteration falls within the scope of the augmented Lagrangian method and hence the analysis of the first part of this thesis applies. In this chapter we will therefore investigate (a slightly more general version of) this iterative image denoising method as well as the resulting evolution equation. We basically follow our work in Burger et al. [30].

This chapter is organized as follows: After introducing the basic functional analytic framework in Section 4.1, we proceed with a quick review of the analysis on pairings between vector valued, bounded functions and measures, as it was conducted by Anzellotti in [12]. With these preparations at hand, we will introduce a generalized definition of Meyer's *g-norm* in Section 4.3 and its relations to functions with finite total variation (cf. Section 4.4).

In Section 4.5 we eventually study the iterative image denoising technique and its corresponding evolution equation. Multiscale properties as well as characterization of solutions for a certain class of data images f are considered in Section 4.6. We close this chapter with a short summary and references for further reading (cf. Section 4.7).

4.1 Assumptions and Notation

Throughout this chapter we assume, that $N \geq 2$ and that $\Omega \subset \mathbb{R}^N$ is an open, simply connected and bounded subset. We moreover shall take for granted that the boundary $\partial\Omega$ can (locally) be represented as the graph of a Lipschitz map and we denote by $\nu : \partial\Omega \to S^{N-1}$ the unit outer normal on $\partial\Omega$. We assume that $1 < p < \infty$ and that p^* is such that $1/p + 1/p^* = 1$.

Synchronizing with the first part of the thesis (cf. Sections 2.1 and 3.1) we set

1. The spaces $X = Y = \mathrm{L}^p(\Omega)$, equipped with the topology $\tau_X = \tau_Y = \tau^{\omega}_{\mathrm{L}^p(\Omega)}$.

2. The weight function $\phi(s) = s^{p-1}$. According to Example A.1.3, one has that $\psi_\phi(s) = \frac{1}{p}s^p$ and for each $u \in \mathrm{L}^p(\Omega)$
$$\mathfrak{J}_\phi(u) = |u|^{p-1}\operatorname{sign}(u) \in \mathrm{L}^{p^*}(\Omega).$$

3. The linear Operator $K = \mathrm{Id} : \mathrm{L}^p(\Omega) \to \mathrm{L}^p(\Omega)$.

4. The convex functional $J_p : \mathrm{L}^p(\Omega) \to \overline{\mathbb{R}}$ defined by
$$J_p(u) = \begin{cases} |\mathrm{D}u|(\Omega) & \text{if } u \in \mathrm{BV}(\Omega) \cap \mathrm{L}^p(\Omega) \\ +\infty & \text{else.} \end{cases} \tag{4.1}$$
If the situation is clear, we shall simply write J instead of J_p.

We recall that the space $\mathrm{BV}(\Omega)$ consists of all functions $u \in \mathrm{L}^1(\Omega)$ whose distributional derivative is a vector-valued Radon measure $\mathrm{D}u$ on Ω with finite total variation $|\mathrm{D}u|$. For $u \in \mathrm{BV}(\Omega) \cap \mathrm{L}^p(\Omega)$ we have the following characterization (cf. e.g. [55, Chap. 5.1 Thm. 1])
$$|\mathrm{D}u|(\Omega) = \sup_{v \in K^{p^*}(\Omega)} \int_\Omega uv \,\mathrm{d}x, \tag{4.2}$$
where $K^{p*}(\Omega)$ is defined as
$$K^{p*}(\Omega) = \overline{\left\{\operatorname{div}(z) \;:\; z \in \mathcal{D}(\Omega)^N,\ \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \leq 1\right\}}^{\mathrm{L}^{p^*}}. \tag{4.3}$$
We recall a well known fact (see e.g. [1, Thm.2.3])

Lemma 4.1.1. The functional J is sequentially weakly lower semicontinuous on $\mathrm{L}^p(\Omega)$.

Proof. Assume that $\{u_n\}_{n\in\mathbb{N}} \subset \mathrm{L}^p(\Omega)$ is such that $u_n \rightharpoonup u$ for some $u \in \mathrm{L}^p(\Omega)$. For $v \in K^{p*}(\Omega)$ we find
$$\int_\Omega uv \,\mathrm{d}x = \liminf_{n\to\infty} \int_\Omega u_n v \,\mathrm{d}x \leq \liminf_{n\to\infty} J(u_n).$$
Taking the supremum over all such functions v proves the desired estimate. $\square$

Lemma 4.1.1 together with Remark 2.1.2 (1) shows that assumptions (R1) - (R3) are satisfied. Moreover, the spaces $\mathrm{L}^p(\Omega)$ are reflexive, which shows (R7) and thus (cf. Remark 3.1.5) for each $f \in \mathrm{L}^p(\Omega)$ and $\alpha > 0$ the sets

$$\Lambda(c) = \left\{ u \in \mathrm{L}^p(\Omega) \ : \ \frac{1}{p} \|f - u\|^p + \alpha J(u) \leq c \right\}$$

are sequentially weakly precompact. That is, (R5) holds. Finally, we note that $\emptyset \neq D(\partial J) \subset D(J)$. Indeed, since J is nonnegative and $0 = J(0) = \inf \{J(u) \ : \ u \in \mathrm{L}^p(\Omega)\}$ we have that

$$0 \in \partial J(0).$$

Therefore, there exists $u_0 \in \mathrm{L}^p(\Omega) \cap \mathrm{BV}(\Omega)$ and $v_0 \in \mathrm{L}^{p^*}(\Omega)$ such that $v_0 \in \partial J(u_0)$. In particular, u_0 is attainable ($K = \mathrm{Id}$). In other words, (R4) and (R6) are fulfilled. In order to safely apply the general results in the first part of the thesis, it remains to verify the closedness requirement (R8). This, however, requires some deeper analysis and will be proven in Section 4.4.

Throughout this chapter we make use of the spaces in the subsequent list

$$\begin{aligned}
X(\Omega)_\mu &:= \left\{ z \in \mathrm{L}^\infty(\Omega, \mathbb{R}^N) \ : \ \operatorname{div}(z) \text{ is a bounded measure on } \Omega \right\} \\
X(\Omega)_p &:= \left\{ z \in \mathrm{L}^\infty(\Omega, \mathbb{R}^N) \ : \ \operatorname{div}(z) \in \mathrm{L}^p(\Omega) \right\} \subset X(\Omega)_\mu \\
\mathrm{BV}(\Omega)_c &:= \mathrm{BV}(\Omega) \cap \mathrm{L}^\infty(\Omega) \cap C(\Omega) \\
\mathrm{BV}(\Omega)_p &:= \mathrm{BV}(\Omega) \cap \mathrm{L}^p(\Omega) \\
\mathrm{L}^p_\diamond(\Omega) &:= \left\{ u \in \mathrm{L}^p(\Omega) \ : \ \int_\Omega u \, \mathrm{d}x = 0 \right\}.
\end{aligned}$$

We additionally remark that $\mathrm{BV}(\Omega) = \mathrm{BV}(\Omega)_p$ for $1 \leq p \leq \frac{N}{N-1}$ (cf. e.g. [55, Chap. 5.6 Thm. 1]).

4.2 Pairings between Measures and Bounded Functions

In this section we are going to provide a brief overview over the work of Anzellotti as presented in [12]. Therein the author establishes a pairing concept between bounded, vector-valued functions $z \in \mathrm{L}^\infty(\Omega, \mathbb{R}^N)$ with $\operatorname{div}(z) \in \mathrm{L}^p(\Omega)$ and the vector measure $\mathrm{D}u$ ($u \in \mathrm{BV}(\Omega)$), that generalizes the inner product

$$\int_\Omega z \cdot \nabla u \, \mathrm{d}x$$

for functions $u \in \mathrm{W}^{1,1}(\Omega)$ as well as the corresponding Green's formula. We start with the following definition:

Let $u \in \mathrm{W}^{1,1}(\Omega) \cap \mathrm{BV}(\Omega)_c$ and $z \in X(\Omega)_\mu$. Then the following pairing between u and z is well defined

$$\langle z, u \rangle_{\partial\Omega} := \int_\Omega u \operatorname{div}(z) \, \mathrm{d}x + \int_\Omega z \cdot \nabla u \, \mathrm{d}x. \tag{4.4}$$

It is important to note, that the rightmost expression in (4.4) would not be properly defined if merely $u \in \mathrm{BV}(\Omega)_c$. However Anzellotti has shown the

Theorem 4.2.1. *There exists a bilinear extension* $\langle \cdot, \cdot \rangle_{\partial\Omega} : X(\Omega)_\mu \times BV_c \to \mathbb{R}$ *such that*

$$\langle z, u \rangle_{\partial\Omega} = \int_{\partial\Omega} uz \cdot \nu d\mathcal{H}^{N-1} \quad \textit{for all } z \in C^1(\overline{\Omega}, \mathbb{R}^N), \tag{4.5a}$$

$$|\langle z, u \rangle_{\partial\Omega}| \leq \|z\|_{L^\infty(\Omega,\mathbb{R}^N)} \int_{\partial\Omega} |u(x)| \, d\mathcal{H}^{N-1}. \tag{4.5b}$$

Proof. [12, Thm. 1.1] □

For a given $z \in X(\Omega)_\mu$ we define the mapping $F_z : \mathrm{L}^1(\partial\Omega) \to \mathbb{R}$ by

$$F_z(u) = \langle z, w \rangle_{\partial\Omega} \tag{4.6}$$

where $w \in \mathrm{BV}(\Omega)_c$ is such that $w_{|\partial\Omega} = u$. Then it follows from (4.5b) that

$$|F_z(u)| \leq \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \|u\|_{\mathrm{L}^1(\partial\Omega)} \, .$$

Thus F_z is linear and bounded and due to Riesz' representation theorem one can find a unique $T(z) \in \mathrm{L}^\infty(\partial\Omega)$ such that

$$F_z(u) = \int_{\partial\Omega} T(z) u \, \mathrm{d}\mathcal{H}^{N-1}. \tag{4.7}$$

We summarize this observation in the following (cf. [12, Thm.1.2])

Corollary 4.2.2. There exists a linear trace operator $T : X(\Omega)_\mu \to \mathrm{L}^\infty(\partial\Omega)$ such that

$$\|T(z)\|_{\mathrm{L}^\infty(\partial\Omega)} \leq \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \, , \tag{4.8a}$$

$$\langle z, u \rangle_{\partial\Omega} = \int_{\partial\Omega} T(z) u \, \mathrm{d}\mathcal{H}^{N-1} \quad \text{for all } u \in \mathrm{BV}(\Omega)_c, \tag{4.8b}$$

$$T(z)(x) = z(x) \cdot \nu(x) \quad \text{for all } x \in \partial\Omega \text{ and } z \in C^1(\Omega, \mathbb{R}^N). \tag{4.8c}$$

Remark 4.2.3. Since Ω is assumed to be bounded we find for $1 < p_1 \leq p_2 < \infty$ the ordering

$$X(\Omega)_{p_2} \subset X(\Omega)_{p_1} \subset X(\Omega)_\mu.$$

Consequently for each $p > 1$ there exists a trace operator $T_p : X(\Omega)_p \to \mathrm{L}^\infty(\partial\Omega)$ satisfying (4.8a) – (4.8c), simply given by $T_p = T_{|X(\Omega)_p}$ (where T is defined as in Corollary 4.2.2). If there is no chance for confusion, we write again T instead of T_p.

Theorem 4.2.4. *For all* $z \in X(\Omega)_p$ *and* $u \in W^{1,1}(\Omega) \cap L^{p^*}(\Omega)$ *we have that*

$$\int_\Omega u \mathrm{div}\,(z) \; dx + \int_\Omega \langle \nabla u, z \rangle \; dx = \int_{\partial\Omega} T(z) u d\mathcal{H}^{N-1}. \tag{4.9}$$

Proof. (cf. [12, Prop. 1.3]) From (4.4) and (4.8b) it is evident that (4.9) holds for all $u \in C^\infty(\overline{\Omega}) \subset \mathrm{W}^{1,1}(\Omega) \cap \mathrm{BV}(\Omega)_c$. By a standard mollifier argument (cf. [55, Chap. 4.2. Thm. 1]) we can choose a sequence $\{u_n\}_{n\in\mathbb{N}} \subset C^\infty(\Omega, \mathbb{R}^N)$ such that

$$u_n \to u \text{ in } \mathrm{W}^{1,1}(\Omega) \quad \text{and} \quad u_n \to u \text{ in } \mathrm{L}^{p^*}(\Omega).$$

From the strong convergence in $\mathrm{W}^{1,1}(\Omega)$ it follows from the trace theorem for Sobolev functions [55, Chap. 4.3 Thm. 1] that $u_n \to u$ in $\mathrm{L}^1(\partial\Omega)$ and we conclude

$$\begin{aligned}\int_\Omega u\operatorname{div}(z)\ \mathrm{d}x + \int_\Omega \langle \nabla u, z\rangle\ \mathrm{d}x &= \lim_{n\to\infty}\left(\int_\Omega u_n\operatorname{div}(z)\ \mathrm{d}x + \int_\Omega \langle \nabla u_n, z\rangle\ \mathrm{d}x\right)\\ &= \lim_{n\to\infty}\left(\int_{\partial\Omega} T(z)u_n \mathrm{d}\mathcal{H}^{N-1}\right) = \int_{\partial\Omega} T(z)u\mathrm{d}\mathcal{H}^{N-1}.\end{aligned}$$

□

With these preparations we are now in the position to establish the announced pairing. In order to do so let us assume that

$$u \in \mathrm{BV}(U)_{p^*} \quad \text{and} \quad z \in X(U)_p. \tag{4.10}$$

for all $U \subset\subset \Omega$ and define (cf. [12, Def. 1.4]) the linear functional $(z, \mathrm{D}u) : C_c^1(\Omega) \to \mathbb{R}$ by

$$(z, \mathrm{D}u)(\phi) = -\int_\Omega u\phi\operatorname{div}(z)\ \mathrm{d}x - \int_\Omega uz\cdot\nabla\phi\,\mathrm{d}x.$$

Then, for $u \in \mathrm{BV}(\Omega)\cap C^\infty(\Omega)$, $U \subset\subset \Omega$ and $\phi \in C_c^1(U)$ one finds by applying Theorem 4.2.4

$$\begin{aligned}(z, \mathrm{D}u)(\phi) &= \int_U u\phi\operatorname{div}(z)\ \mathrm{d}x + \int_U uz\cdot\nabla\phi\,\mathrm{d}x\\ &= -\int_U z\cdot\nabla(u\phi)\,\mathrm{d}x + \int_{\partial U} T(z)u\phi\mathrm{d}\mathcal{H}^{N-1} + \int_U uz\cdot\nabla\phi\,\mathrm{d}x\\ &= \int_U \nabla u\cdot z\phi\,\mathrm{d}x\end{aligned}\tag{4.11}$$

and therefore

$$|(z, \mathrm{D}u)(\phi)| \le \|\phi\|_{\mathrm{L}^\infty(U)}\|z\|_{\mathrm{L}^\infty(U,\mathbb{R}^N)}|\mathrm{D}u|(U). \tag{4.12}$$

For an arbitrary u satisfying (4.10) we can find (see for instance [55, Chap. 5.2 Thm. 2]) a sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathrm{BV}(\Omega)\cap C^\infty(\Omega)$ such that

$$\lim_{n\to\infty} u_n = u \text{ in } \mathrm{L}^1(\Omega) \quad \text{and} \quad \lim_{n\to\infty}|\mathrm{D}u_n|(\Omega) = |\mathrm{D}u|(\Omega).$$

This together with the definition of $(z, \mathrm{D}u)$ and (4.12) shows that

$$\begin{aligned}|(z, \mathrm{D}u)(\phi)| = \lim_{n\to\infty}|(z, \mathrm{D}u_n)(\phi)| &\le \lim_{n\to\infty}\|\phi\|_{\mathrm{L}^\infty(U)}\|z\|_{\mathrm{L}^\infty(U,\mathbb{R}^N)}|\mathrm{D}u_n|(U)\\ &\le \|\phi\|_{\mathrm{L}^\infty(U)}\|z\|_{\mathrm{L}^\infty(U,\mathbb{R}^N)}|\mathrm{D}u|(U).\end{aligned}$$

Finally, we remark that $C_c^1(\Omega) \subset C_c(\Omega)$ is dense w.r.t. $\|\cdot\|_{\mathrm{L}^\infty}$, as a consequence of which we can uniquely extend $(z, \mathrm{D}u)$ to $C_c(\Omega)$ by the Hahn-Banach Theorem (cf. e.g. [94, Thm. 1.9.1]).

Summarizing we note that for every compact $K \subset \Omega$ there exists a constant $c > 0$ such that for $\phi \in C_c(\Omega)$, $\operatorname{supp}(\phi) \subset K$ the following estimate holds

$$|(z, \mathrm{D}u)(\phi)| \le c\,\|\phi\|_{\mathrm{L}^\infty(K)}.$$

Hence Riesz' representation Theorem [55, Chap. 1.8 Thm. 1] is applicable and proves (cf. [12, Thm. 1.5])

Theorem 4.2.5. *If u and z satisfy* (4.10)*, then the functional* (z, Du) *is a Radon measure on* Ω*.*

Let $|(z, \mathrm{D}u)|$ denote the total variation measure of $(z, \mathrm{D}u)$, that is, for every open $U \subset \Omega$ define

$$|(z, \mathrm{D}u)|\,(U) = \sup\left\{(z, \mathrm{D}u)(\phi) \;:\; \phi \in C_c(\Omega),\, \|\phi\|_{\mathrm{L}^\infty} \leq 1,\, \mathrm{supp}(\phi) \subset U\right\}.$$

Then from (4.10) and the Radon-Nikodým Theorem (see e.g. [55, Chap. 1.6 Thm. 1]) it follows that

Corollary 4.2.6. Let u and z satisfy (4.10). The measures $(z, \mathrm{D}u)$ and $|(z, \mathrm{D}u)|$ are absolutely continuous w.r.t. the measure $|\mathrm{D}u|$ and

$$\left|\int_B (z, \mathrm{D}u)\right| \leq \int_B |(z, \mathrm{D}u)| \leq \|z\|_{\mathrm{L}^\infty(U,\mathbb{R}^N)}\,|\mathrm{D}u|\,(B)$$

for all Borel sets B and open sets U such that $B \subset U \subset \Omega$. Moreover, there exists a $|\mathrm{D}u|-$measurable function $\theta(z, \mathrm{D}u, \cdot) : \Omega \to \mathbb{R}$ such that

$$\int_B (z, \mathrm{D}u) = \int_B \theta(z, \mathrm{D}u, x)\,\mathrm{d}\,|\mathrm{D}u|\,(x), \qquad \text{for all Borel sets } B \subset \Omega$$
$$\text{and } \|\theta(z, \mathrm{D}u, \cdot)\|_{\mathrm{L}^\infty(\Omega,\mathbb{R},|\mathrm{D}u|)} \leq \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)}.$$

In addition to the above results on the measure $(z, \mathrm{D}u)$ the following two proposition (among others) were proven in [12]. The first one deals with invariance under compositions with nonincreasing functions, whereas the second provides a generalization of the Green's formula in Theorem 4.2.4.

Proposition 4.2.7. *Assume that u and z are as in* (4.10) *and let θ be the Radon-Nikodým derivative of (z, Du) w.r.t. $|Du|$ (cf. Corollary 4.2.6). If $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and increasing, one has*

$$\theta(z, Du, \cdot) = \theta(z, D(f \circ u), \cdot), \quad |Du|\text{-}a.e \text{ in } \Omega.$$

Proof. [12, Prop. 2.8] □

Remark 4.2.8. Note that due to the fact, that $f^{-1} : \mathrm{ran}(f) \to \mathbb{R}$ is again smooth and increasing, the above equality also holds $|\mathrm{D}(f \circ u)|$-a.e in Ω.

Proposition 4.2.9. *Let u and z be as in* (4.10)*. Then one has*

$$\int_\Omega u\mathrm{div}\,(z)\;dx + \int_\Omega (z, Du) = \int_{\partial\Omega} T_p(z) u d\mathcal{H}^{N-1}. \tag{4.13}$$

Proof. [12, Prop. 1.9] □

4.3 The g-Norm

With the analysis conducted in the previous chapter, we will now introduce a model for *oscillating patterns* in images where we follow the idea of Meyer in [95]. Oscillating patterns are measured by means of the *g-norm*. The concept of oscillating patterns will enable us (cf. Section 4.4) to formulate a convenient representation of the subgradient of the (extended) BV-seminorm J (as defined in (4.1)).

According to Meyer, an oscillating pattern (on $\mathbb{R}^2$) is a distribution v that admits the representation

$$v = \frac{\partial}{\partial x_1} z_1(x_1, x_2) + \frac{\partial}{\partial x_2} z_2(x_1, x_2) \tag{4.14}$$

where $z_1, z_2 \in \mathrm{L}^\infty(\mathbb{R}^2)$ and is measured by the *g-norm*

$$\inf \left\{ \sqrt{\|z_1\|_\infty^2 + \|z_2\|_\infty^2} \ : \ z_1, z_2 \in \mathrm{L}^\infty(\mathbb{R}^2) \text{ satisfy (4.14)} \right\}.$$

It turns out that the g-norm is less sensitive to oscillations than e.g. the L^p-norm. Moreover, it is a convenient tool to characterize the subdifferential of the total variation seminorm.

We generalize the definition of the g-norm to bounded domains. We follow the ideas of Aubert & Aujol in [13].

Definition 4.3.1. Let $T = T_p : X(\Omega)_p \to \mathrm{L}^\infty(\partial\Omega)$ be the trace operator in Corollary 4.2.2.

1. We call $v \in \mathrm{L}^p(\Omega)$ an *oscillating pattern on* Ω if there exists $z \in X(\Omega)_p$ such that $T_p(z) = 0$ and $v = \mathrm{div}\,(z)$. We define the set of all oscillating patterns as

$$\mathcal{L}^p_\diamondsuit(\Omega) := \mathrm{div}\,(\ker(T_p)) = \{\mathrm{div}\,(z) \ : \ z \in X(\Omega)_p,\ T_p(v) = 0\}\,.$$

2. Let $v \in \mathcal{L}^p_\diamondsuit(\Omega)$ be an oscillating pattern. The value

$$\|v\|_* := \inf_{z \in \ker(T_p)} \left\{ \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \ : \ \mathrm{div}\,(z) = v \right\}. \tag{4.15}$$

is called *g-norm* of v.

Theorem 4.3.2. *The function $\|\cdot\|_*$ renders $\mathcal{L}^p_\diamondsuit(\Omega)$ a normed space and for every $v \in \mathcal{L}^p_\diamondsuit(\Omega)$ there exists $z \in \ker(T_p)$ such that*

$$\mathrm{div}\,(z) = v \quad \text{and} \quad \|z\|_{L^\infty(\Omega,\mathbb{R}^N)} = \|v\|_*\,.$$

Proof. (cf. [13, Lem. 3.2]) Let $\{z_n\}_{n\in\mathbb{N}} \subset \ker(T_p)$ be such that $\mathrm{div}\,(z_n) = v$ and $\|z_n\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \to \|v\|_*$ as $n \to \infty$. Obviously

$$\sup_{n\in\mathbb{N}} \|z_n\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} < \infty,$$

as a consequence of which we can choose a selection $n \mapsto \rho(n)$ such that $z_{\rho(n)} \rightharpoonup^* z$ for a $z \in \mathrm{L}^\infty(\Omega,\mathbb{R}^N)$. Theorem 4.2.4 gives for all $\phi \in C_c^1(\Omega)$.

$$\int_\Omega \phi v \,\mathrm{d}x = \lim_{n\to\infty} -\int_\Omega \nabla\phi \cdot z_{\rho(n)} \,\mathrm{d}x = -\int_\Omega \nabla\phi \cdot z \,\mathrm{d}x = \int_\Omega \phi \mathrm{div}\,(z)\,\mathrm{d}x$$

where the last equation holds in the sense of distributions, but since $v \in \mathrm{L}^p(\Omega)$ this already shows that $z \in X(\Omega)_p$.

It remains to verify that $z \in \ker(T_p)$. To this end we again apply Green's Formula (Thm. 4.2.4) with a testfunction $\phi \in \mathrm{W}^{1,1}(\Omega) \cap \mathrm{L}^{p^*}(\Omega)$ which gives

$$0 = \int_{\partial\Omega} T_p(z)\phi \mathrm{d}\mathcal{H}^{N-1}.$$

For $w \in \mathrm{BV}(\Omega)_c$ it follows from [12, Lem. 5.2] that there exists $\{\phi_n\}_{n\in\mathbb{N}} \subset \mathrm{W}^{1,1}(\Omega) \cap \mathrm{L}^{p^*}(\Omega)$ such that $\phi_n \to w$ w.r.t. the L^{p^*}-topology and $w(x) = \phi_n(x)$ $\mathcal{H}^{n-1}$-a.e. on $\partial\Omega$. This shows that

$$\int_{\partial\Omega} T_p(z) w \mathrm{d}\mathcal{H}^{N-1} = \lim_{n\to\infty} \int_{\partial\Omega} T_p(z)\phi_n \mathrm{d}\mathcal{H}^{N-1} = 0.$$

Thus the functional $F_z : \mathrm{L}^1(\partial\Omega) \to \mathbb{R}$ as defined in (4.6) is identically zero and since $T_p(z) \in \mathrm{L}^\infty(\partial\Omega)$ is the unique element satisfying (cf. (4.7))

$$0 = F_z(\phi) = \int_{\partial\Omega} T_p(z)\phi \mathrm{d}\mathcal{H}^{N-1}$$

it follows that $T_p(z) = 0$ for $\mathcal{H}^{N-1}$ a.e. $x \in \partial\Omega$. Now we conclude from the weak* lower semicontinuity of $\|\cdot\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)}$ and the definition of $\|v\|_*$ that

$$\|v\|_* \le \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \le \liminf_{n\to\infty} \left\|z_{\rho(n)}\right\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} = \|v\|_*$$

The norm properties of $\|\cdot\|_*$ follow directly from the linearity of $\operatorname{div}(\cdot)$ and the norm properties of $\|\cdot\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)}$. □

Remark 4.3.3. From Green's formula it follows that $\mathcal{L}^p_\Diamond(\Omega) \subset \mathrm{L}^p_\Diamond(\Omega)$: let $v \in \mathcal{L}^p_\Diamond(\Omega)$ and choose $z \in \ker(T_p)$ such that $\operatorname{div}(z) = v$. Then Theorem 4.2.4 shows

$$\int_\Omega v \,\mathrm{d}x = \int_\Omega \operatorname{div}(z)\,\mathrm{d}x = \int_\Omega 1\operatorname{div}(z)\,\mathrm{d}x = -\int_\Omega \nabla(1)\cdot z\,\mathrm{d}x + \int_{\partial\Omega} T_p(z)\mathrm{d}\mathcal{H}^{N-1} = 0$$

In general the space $\mathcal{L}^p_\Diamond(\Omega)$ is *strictly* contained in $\mathrm{L}^p_\Diamond(\Omega)$. If $p > N$, however, one has $\mathrm{L}^p_\Diamond(\Omega) = \mathcal{L}^p_\Diamond(\Omega)$. In order to see this, note that

$$\Delta\phi = v \quad \text{in } \Omega, \tag{4.16}$$

$$\nu\cdot\nabla\phi = 0 \quad \text{on } \partial\Omega \tag{4.17}$$

attains a solution in $\mathrm{W}^{2,p}(\Omega)$ and thus $z := \nabla u \in \mathrm{W}^{1,p}(\Omega,\mathbb{R}^N)$. If $p > N$, it follows from [125, Thm. 2.4.4] that the embedding

$$\mathrm{W}^{1,p}(\Omega,\mathbb{R}^N) \hookrightarrow \mathrm{L}^\infty(\Omega,\mathbb{R}^N).$$

is continuous. This and (4.16) thus imply that $z \in \ker(T_p)$ and $\operatorname{div}(z) = v$. In other words $\mathrm{L}^p_\Diamond(\Omega) = \mathcal{L}^p_\Diamond(\Omega)$ if $p > N$.

In [25] Bourgain and Brézis showed the remarkable result, that for every $v \in \mathrm{L}^N_\Diamond(\Omega)$ there exists a $z \in C(\overline{\Omega},\mathbb{R}^N) \cap \mathrm{W}^{1,N}_0(\Omega,\mathbb{R}^N) \subset \mathrm{L}^\infty(\Omega,\mathbb{R}^N)$ such that $\operatorname{div}(z) = v$ (cf. [25, Thm.3]). This implies that even

$$\mathrm{L}^N_\Diamond(\Omega) = \mathcal{L}^N_\Diamond(\Omega).$$

This includes the case $N = 2$ and is therefore relevant for image processing tasks. In the context of image processing, Bourgain's and Brézis' result was first brought to attention by Aubert and Aujol in [13].

With the assertions in Theorem 4.3.2 we can easily prove the L^p-closedness of unit ball in $\mathcal{L}^p_\Diamond(\Omega)$ w.r.t. $\|\cdot\|_*$:

$$B^p_* := \left\{ v \in \mathcal{L}^p_\Diamond(\Omega) \ : \ \|v\|_* \leq 1 \right\}. \tag{4.18}$$

Corollary 4.3.4. The ball B^p_* is sequentially weakly closed in $\mathrm{L}^p(\Omega)$.

Proof. (cf. [13, Lem. 2.2]) Let $\{v_n\}_{n\in\mathbb{N}} \subset \mathcal{L}^p_\Diamond(\Omega)$ and $v \in \mathrm{L}^p_\Diamond(\Omega)$ such that

$$w\text{-}\lim_{n\to\infty} v_n = v.$$

From Theorem 4.3.2 it follows that there exists a sequence $\{z_n\}_{n\in\mathbb{N}} \in \ker(T_p)$ such that

$$\mathrm{div}\,(z_n) = v_n \quad \text{ and } 1 \geq \|v_n\|_* = \|z_n\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)}\,.$$

Hence we can choose $n \to \rho(n)$ such that $\left\{z_{\rho(n)}\right\}_{n\in\mathbb{N}}$ weakly* converges to an element $z \in \mathrm{L}^\infty(\Omega,\mathbb{R}^N)$ and in exactly the same way as in the proof of Theorem 4.3.2 one shows

$$z \in \ker(T_p), \quad \mathrm{div}\,(z) = v \quad \text{ and } \quad \|v\|_* = \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)}\,.$$

The assertion finally follows from the fact that

$$\|v\|_* = \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \leq \liminf_{n\to\infty} \left\|z_{\rho(n)}\right\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \leq 1.$$

□

Corollary 4.3.5. The mapping

$$\begin{aligned} G_p:\ \mathrm{L}^p(\Omega) &\longrightarrow \overline{\mathbb{R}} \\ v &\longmapsto \begin{cases} \|v\|_* & \text{if } v \in \mathcal{L}^p_\Diamond(\Omega) \\ +\infty & \text{else} \end{cases} \end{aligned} \tag{4.19}$$

is sequentially weakly lower semicontinuous w.r.t. the $\mathrm{L}^p(\Omega)$ topology. If there is no chance for confusion, we will write G instead of G_p.

In general, it is not possible to compare the topology on $\mathcal{L}^p_\Diamond(\Omega)$ induced by $\|\cdot\|_*$ to the induced strong or weak topology. This is indicated by the following

Example 4.3.6. Let $\Omega = [0,\pi]\times[0,\pi]$ and $p = 2$. Define for $n > 1$ an element $v_n \in \mathcal{L}^2_\Diamond(\Omega)$ by $v_n = \mathrm{div}\,(z_n)$, where

$$z_n(x,y) = \frac{1}{2n}\left(\sin((2n)^2 x), \sin((2n)^2 y)\right)^T.$$

Then clearly

$$\|v_n\|_* \leq \|z_n\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^2)} = \frac{1}{2n}$$

and thus $\|v_n\|_* \to 0$. However we have for $n > 0$

$$\|v_n\|^2_{\mathrm{L}^2} = 2n\pi,$$

that is, $\{v_n\}_{n\in\mathbb{N}}$ is unbounded in $\mathrm{L}^2(\Omega)$.

Example 4.3.6 is in the spirit of what has been said in the beginning in the section: The g-norm is less sensitive to oscillations than the L^p-norm. Restricted to L^p-bounded sets, however, the g-norm shows quite standard behavior

Theorem 4.3.7. *Let $\{v_n\}_{n\in\mathbb{N}} \subset \mathcal{L}^p_\Diamond(\Omega)$ and $v \in \mathcal{L}^p_\Diamond(\Omega)$ such that*

$$\lim_{n\to\infty} \|v_n - v\|_* = 0 \quad \text{and} \quad \sup_{n\in\mathbb{N}} \|v_n\|_{L^p} < \infty.$$

Then $v_n \rightharpoonup v$ in $L^p(\Omega)$.

Proof. Since $\{v_n\}_{n\in\mathbb{N}}$ is bounded in $\mathrm{L}^p(\Omega)$ we can find a selection $n \mapsto \rho(n)$ such that

$$v_{\rho(n)} \rightharpoonup \hat{v}, \quad \text{in } \mathrm{L}^p(\Omega) \tag{4.20}$$

for some $\hat{v} \in \mathrm{L}^p_\Diamond(\Omega)$. Moreover let $v \in \mathcal{L}^p_\Diamond(\Omega)$ such that $v_n \to v$ w.r.t. $\|\cdot\|_*$. Then it follows from Theorem 4.3.2 that there exists a sequence $\{z_{\rho(n)}\}_{n\in\mathbb{N}} \subseteq \mathrm{L}^\infty(\Omega, \mathbb{R}^N)$ such that

$$T_p(z_{\rho(n)}) = 0, \quad \operatorname{div}\left(z_{\rho(n)}\right) = v_{\rho(n)} - v \quad \text{and} \quad \left\|z_{\rho(n)}\right\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} = \left\|v_{\rho(n)} - v\right\|_*$$

for all $n \in \mathbb{N}$. This together with Theorem 4.2.4 shows that $\int_\Omega u(v_{\rho(n))} - v)\,\mathrm{d}x = -\int_\Omega \nabla u z_{\rho(n)}\,\mathrm{d}x$ for all $u \in \mathrm{W}^{1,1}(\Omega) \cap \mathrm{L}^{p_*}(\Omega)$ and therefore

$$\lim_{n\to\infty} \int_\Omega u v_{\rho(n)}\,\mathrm{d}x = \int_\Omega uv\,\mathrm{d}x$$

since $\left\|z_{\rho(n)}\right\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} = \left\|v_{\rho(n)} - v\right\|_* \to 0$. This together with (4.20) shows $\hat{v} = v$. Consequently, every subsequence of $\{v_n\}_{n\in\mathbb{N}}$ has a weakly convergent subsequence with limit v. This finally implies $v_n \rightharpoonup v$. □

For the remainder of this section we shed some light on an important duality relation between the total variation seminorm $|\mathrm{D}\cdot|(\Omega)$ and the g-norm $\|\cdot\|_*$ making use of the results in Section 4.2.

Corollary 4.3.8. Let $1 < p < \infty$ and p^* its conjugate exponent. For all $u \in \mathrm{BV}(\Omega)_{p^*}$ and $v \in \mathcal{L}^p_\Diamond(\Omega)$ there exists a $|\mathrm{D}u|$-measurable function $\vartheta(v, \mathrm{D}u, \cdot) : \Omega \to \mathbb{R}$ such that

$$\int_\Omega uv\,\mathrm{d}x = \int_\Omega \vartheta(v, \mathrm{D}u, x)\,\mathrm{d}\,|\mathrm{D}u| \quad \text{and} \quad \|\vartheta(v, \mathrm{D}u, \cdot)\|_{\mathrm{L}^\infty(\Omega,\mathbb{R},|\mathrm{D}u|)} \leq \|v\|_*\,.$$

In particular one has

$$\left|\int_\Omega uv\,\mathrm{d}x\right| \leq |\mathrm{D}u|\,(\Omega)\,\|v\|_*\,. \tag{4.21}$$

Moreover, if $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and increasing, one has

$$\vartheta(v, \mathrm{D}u, \cdot) = \vartheta(v, \mathrm{D}(f \circ u), \cdot), \quad |\mathrm{D}u|\text{-a.e and } |\mathrm{D}(f \circ u)|\text{-a.e in } \Omega.$$

Proof. Since $v \in \mathcal{L}^p_\Diamond(\Omega)$ it follows from Theorem 4.3.2 that there exists an element $z \in \ker(T_p)$ such that $\operatorname{div}(z) = v$ and $\|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} = \|v\|_*$. Then combination of Corollary 4.2.6 and Proposition 4.2.9 shows that there exists a $|\mathrm{D}u|$-measurable function $\theta(z, \mathrm{D}u, \cdot) : \Omega \to \mathbb{R}$ such that

$$\int_\Omega uv\,\mathrm{d}x = \int_\Omega u\operatorname{div}(z)\,\mathrm{d}x = -\int_\Omega \theta(z, \mathrm{D}u, x)\,\mathrm{d}\,|\mathrm{D}u|\,(x).$$

and
$$\|\theta(z, \mathrm{D}u, \cdot)\|_{\mathrm{L}^\infty(\Omega,\mathbb{R},|\mathrm{D}u|)} \leq \|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} = \|v\|_* \,.$$
The assertion follows with $\vartheta(v, \mathrm{D}u, \cdot) = \theta(z, \mathrm{D}u, \cdot)$ and Proposition 4.2.7. $\square$

In particular Corollary 4.3.8 shows that for all $u \in \mathrm{BV}(\Omega)_{p^*}$ such that $\mathfrak{J}_{p^*}(u) \in \mathcal{L}^p_\Diamond(\Omega)$ we have that
$$\|u\|^{p^*}_{\mathrm{L}^{p^*}} = \int_\Omega u\mathfrak{J}_{p^*}(u)\,\mathrm{d}x \leq |\mathrm{D}u|(\Omega)\,\|\mathfrak{J}_{p^*}(u)\|_* \,.$$
This gives rise to the following definition

Definition 4.3.9. A function $u \in \mathrm{BV}(\Omega)_p$ is called *calibrable* if $\mathfrak{J}_p(u) \in \mathcal{L}^{p^*}_\Diamond(\Omega)$ and
$$\|u\|^p_{\mathrm{L}^p} = |\mathrm{D}u|(\Omega)\,\|\mathfrak{J}_p(u)\|_* \,.$$

Remark 4.3.10. From Definition 4.3.9 and Corollary 4.3.8 it becomes clear, that in order to check whether or not a given function $u \in \mathrm{L}^p(\Omega)$ is calibrable or not, it suffices to find a vector field $z \in \mathrm{L}^\infty(\Omega, \mathbb{R}^N)$, such that $\operatorname{div}(z) = \mathfrak{J}_p(u)$ and
$$|\mathrm{D}u|(\Omega)\,\|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \leq \|u\|^p_{\mathrm{L}^p} \,.$$

4.4 Subdifferentiability and Duality

With the results in Sections 4.2 and 4.3 we have all the tools at hand necessary for a profound analysis of the subdifferential ∂J (recall that $J = J_p$ denotes the extension of the BV-seminorm on $\mathrm{L}^p(\Omega)$ as in (4.1)) and its duality relations with $G = G_{p^*}$ (as defined in (4.19)). Before we start recall the definition of the unit ball in $\mathcal{L}^{p^*}_\Diamond(\Omega)$ w.r.t. $\|\cdot\|_*$ in (4.18), that is
$$B^{p^*}_* = \left\{ v \in \mathcal{L}^{p^*}_\Diamond(\Omega) \;:\; \|v\|_* \leq 1 \right\}.$$

Lemma 4.4.1. For $u \in \mathrm{BV}(\Omega)_p$ one has
$$J(u) = \sup_{v \in B^{p^*}_*} \int_\Omega uv\,\mathrm{d}x.$$

Proof. Let $u \in \mathrm{BV}(\Omega)_p$ and recall from (4.2) that
$$J(u) = |\mathrm{D}u|(\Omega) = \sup_{v \in K^{p^*}(\Omega)} \int_\Omega uv\,\mathrm{d}x$$
where $K^{p^*}(\Omega)$ is defined as in (4.3). Obviously we have that $K^{p^*}(\Omega) \subset B^{p^*}_*$. Indeed, assume that $v \in K^{p^*}(\Omega)$. Then one can choose a sequence $\{z_n\}_{n\in\mathbb{N}} \subset \mathcal{D}(\Omega)^N$ such that
$$\sup_{n\in\mathbb{N}} \|z_n\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N)} \leq 1 \quad \text{and} \quad \lim_{n\to\infty} \operatorname{div}(z_n) \to v.$$
Since $v_n := \operatorname{div}(z_n) \in B^{p^*}_*$ for all $n \in \mathbb{N}$ it follows from Corollary 4.3.4 $v \in B^{p^*}_*$. Moreover, it follows from Corollary 4.3.8 that for all $v \in B^{p^*}_*$
$$\int_\Omega uv\,\mathrm{d}x \leq J(u)\,\|v\|_* \leq J(u).$$

Summarizing, we find

$$J(u) = \sup_{v \in K^{p^*}(\Omega)} \int_\Omega uv \,\mathrm{d}x \leq \sup_{v \in B_*^{p^*}} \int_\Omega uv \,\mathrm{d}x \leq J(u).$$

□

With this preparation, we can give a convenient representation of both, the subgradient and the Legendre – Fenchel conjugate of the (extended) total variation seminorm J by means of the unit ball $B_*^{p^*}$.

Theorem 4.4.2. *Let $u \in BV(\Omega)_p$ and $v \in L^{p^*}(\Omega)$.*

1. *The Legendre – Fenchel conjugate $J^* : L^{p^*}(\Omega) \to \overline{\mathbb{R}}$ of J is given by*

$$J^*(v) = \chi_{B_*^{p^*}}(v).$$

2. *One has $v \in \partial J(u)$ if and only if*

$$v \in B_*^{p^*} \qquad \textit{and} \qquad \int_\Omega uv \, dx = J(u).$$

Proof. (1). Observe that for $u \in \mathrm{BV}(\Omega)_p$ it follows from Lemma A.1.22 that

$$\left(\chi_{B_*^{p^*}}\right)^*(u) = \sup_{v \in B_*^{p^*}} \int_\Omega uv \,\mathrm{d}x = J(u).$$

Since $B_*^{p^*}$ is convex and (sequentially weakly) closed in $\mathrm{L}^{p^*}(\Omega)$ (cf. Corollary 4.3.4) it follows that $\chi_{B_*^{p^*}}$ is convex and (sequentially weakly) lower semicontinous. Thus we find from the Fenchel – Moreau Theorem (cf. [81, Chap. 3.3.3 Thm. 1]).

$$\chi_{B_*^{p^*}} = \left(\chi_{B_*^{p^*}}\right)^{**} = J^*.$$

(2). Note that according to Lemma A.2.12 $v \in \partial J(u)$ is equivalent to

$$J(u) + J^*(v) = \int_\Omega uv \,\mathrm{d}x.$$

If $u \in D(\partial J)$ one immediately finds $J^*(v) < \infty$ and thus $v \in B_*^{p^*}$. The assertion thus follows from (1). □

Corollary 4.3.8 and the properties of the density function $\vartheta(v, \mathrm{D}u, \cdot)$ (cf. Remark 4.2.7) provide the basic tools to prove the weak-weak closedness of the subgradient of J.

Theorem 4.4.3. *The L^p-subgradient $\partial J \subset L^p(\Omega) \times L^{p^*}(\Omega)$ is closed in the product topology of the weak topologies on $L^p(\Omega)$ and $L^{p^*}(\Omega)$.*

Proof. Let $\{u_n\}_{n\in\mathbb{N}} \subset \mathrm{L}^p(\Omega)$ and $\{v_n\}_{n\in\mathbb{N}} \subset \mathrm{L}^{p^*}(\Omega)$ as well as $u \in \mathrm{L}^p(\Omega)$ and $v \in \mathrm{L}^{p^*}(\Omega)$ such that

$$u_n \rightharpoonup u, \quad v_n \rightharpoonup v \quad \text{ and } \quad v_n \in \partial J(u_n).$$

Since $\{u_n\}_{n\in\mathbb{N}}$ and $\{v_n\}_{n\in\mathbb{N}}$ are weakly convergent, it follows that they are bounded. Thus Theorem 4.4.2 gives

$$\sup_{n\in\mathbb{N}} J(u_n) = \sup_{n\in\mathbb{N}} \int_\Omega u_n v_n \,\mathrm{d}x \le \|u_n\|_{\mathrm{L}^p} \|v_n\|_{\mathrm{L}^{p^*}} < \infty.$$

In other words, the sequence $\{u_n\}_{n\in\mathbb{N}}$ is bounded in $\mathrm{BV}(\Omega)$ and we therefore find from the compact embedding $\mathrm{BV}(\Omega) \hookrightarrow \mathrm{L}^1(\Omega)$ that every subsequence of $\{u_n\}_{n\in\mathbb{N}}$ has a strongly L^1-convergent subsequence with limit $\mathrm{L}^1(\Omega)$. Since weak convergence in $\mathrm{L}^p(\Omega)$ implies weak convergence in $\mathrm{L}^1(\Omega)$ and due to the fact that weak and strong limits coincide we conclude that all theses limits are equal to u. Therefore

$$\lim_{n\to\infty} u_n = u, \quad \text{strongly in } \mathrm{L}^1(\Omega)$$

and lower semicontinuity proves that $J(u) < \infty$.

Let $\gamma > 0$. We introduce the truncation operator

$$S_\gamma(r) := \begin{cases} \gamma & \text{if } r \ge \gamma, \\ -\gamma & \text{if } r \le -\gamma, \\ r & \text{else.} \end{cases} \tag{4.22}$$

and set $S_\gamma^\varepsilon = (S_\gamma * \eta_\varepsilon)$ where η_ε for $\varepsilon > 0$ denotes a standard mollifier[1]. Note that $S_\gamma^\varepsilon : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, increasing and Lipschitz continuous with Lipschitz constant $L = 1$. Corollary 4.3.8 implies that for every $n > 0$ there exists a $|\mathrm{D}u_n|$ - measurable function $\vartheta_n(v_n, \mathrm{D}u_n, \cdot) : \Omega \to \mathbb{R}$ such that $\vartheta_n(v_n, \mathrm{D}u_n, \cdot) \le 1$ $|\mathrm{D}u_n|$ - a.e. $x \in \Omega$,

$$\int_\Omega v_n u_n \,\mathrm{d}x = \int_\Omega \vartheta_k(v_n, \mathrm{D}u_n, x) \,\mathrm{d}|\mathrm{D}u_n|(x) \tag{4.23}$$

and

$$\vartheta_n(v_n, \mathrm{D}(S_\gamma^\varepsilon \circ u_n), x) = \vartheta_n(v_n, \mathrm{D}u_n, x), \quad |\mathrm{D}u_n| - \text{ a.e.} \tag{4.24}$$

Note that (4.24) also holds for $|\mathrm{D}S_\gamma^\varepsilon \circ u_n|$ - almost every $x \in \Omega$. From Theorem 4.4.2 and (4.23) it follows that

$$|\mathrm{D}u_n|(\Omega) = \int_\Omega u_n v_n \,\mathrm{d}x = \int_\Omega \vartheta_n(v_n, \mathrm{D}u_n, x) \,\mathrm{d}|\mathrm{D}u_n|(x).$$

Since $\vartheta_n(v_n, \mathrm{D}u_n, \cdot) \le 1$, $|\mathrm{D}u_n|$-almost surely, the above equation implies that $\vartheta_n(v_n, \mathrm{D}u_n, x) = 1$ for $|\mathrm{D}u_n|$ - almost every $x \in \Omega$ and consequently from (4.24) it follows that

$$\int_\Omega v_n \left(S_\gamma^\varepsilon \circ u_n\right) \,\mathrm{d}x = \int_\Omega \vartheta_n(v_n, \mathrm{D}(S_\gamma^\varepsilon \circ u_n), x) \,\mathrm{d}|\mathrm{D}S_\gamma^\varepsilon(u_n)|(x) = J\left(S_\gamma^\varepsilon \circ u_n\right). \tag{4.25}$$

[1] Let $\varepsilon > 0$ and $\eta \in C_c^\infty(\mathbb{R})$ such that $\eta \ge 0$ and

$$\int_{-\infty}^{\infty} \eta(\tau) \,\mathrm{d}\tau = 1.$$

Then a *standard mollifier* is defined as $\eta_\varepsilon(s) := \varepsilon^{-1}\eta(s\varepsilon^{-1})$ for $s \in \mathbb{R}$.

Since $\|u_n - u\|_{\mathrm{L}^1} \to 0$ it follows from the Lipschitz continuity of S_γ^ε that $\left\|S_\gamma^\varepsilon \circ u_n - S_\gamma^\varepsilon \circ u\right\|_{\mathrm{L}^1} \to 0$ as $n \to \infty$, and Lemma A.1.22 implies for every $\tilde{\varepsilon} > 0$

$$\lim_{n\to\infty} \lambda^N \left(\left\{x \in \Omega \;:\; \left|(S_\gamma^\varepsilon \circ u_n)(x) - (S_\gamma^\varepsilon \circ u)(x)\right| \geq \tilde{\varepsilon}\right\}\right) = 0.$$

Moreover, $S_\gamma^\varepsilon(u_k)$ is uniformly bounded in $\mathrm{L}^\infty(\Omega)$ (by γ) and therefore we can apply Lemma A.1.22 and observe

$$\lim_{n\to\infty} \left\|S_\gamma^\varepsilon \circ u_n - S_\gamma^\varepsilon \circ u\right\|_{\mathrm{L}^p} = 0.$$

The strong - weak continuity of the pairing on $\mathrm{L}^p(\Omega) \times \mathrm{L}^{p_*}(\Omega)$ and (4.25) then show

$$\lim_{n\to\infty} J\left(S_\gamma^\varepsilon \circ u_n\right) = \lim_{n\to\infty} \int_\Omega S_\gamma^\varepsilon \circ u_n v_n \,\mathrm{d}x = \int_\Omega S_\gamma^\varepsilon(u) v \,\mathrm{d}x \leq J\left(S_\gamma^\varepsilon(u)\right),$$

where the last inequality follows from the fact that $v \in B_*^{p^*}$ and Corollary 4.3.8. Together with the weak lower semicontinuity of J this implies that $J\left(S_\gamma^\varepsilon(u)\right) = \int_\Omega S_\gamma^\varepsilon(u) v \,\mathrm{d}x$ and thus $v \in \partial J(S_\gamma^\varepsilon \circ u)$ according to Theorem 4.4.2.

Moreover $S_\gamma^\varepsilon \to S_\gamma$ uniformly on $\mathbb{R}$ as $\varepsilon \to 0$: The mollifier function η is compactly supported in $(-1, 1)$ and $\int_{-1}^1 \eta(\sigma)\,\mathrm{d}\sigma = 1$. Therefore one finds for all $s \in \mathbb{R}$

$$\begin{aligned}
\left|S_\gamma(s) - S_\gamma^\varepsilon(s)\right| &= \left|S_\gamma(s) - \frac{1}{\varepsilon}\int_{-\infty}^\infty S_\gamma(s-\sigma)\eta(\varepsilon^{-1}\sigma)\,\mathrm{d}\sigma\right| \\
&= \left|\int_{-\infty}^\infty \left(S_\gamma(s) - S_\gamma(s-\varepsilon\sigma)\right)\eta(\sigma)\,\mathrm{d}\sigma\right| \\
&\leq \frac{1}{\varepsilon}\int_{-\varepsilon}^\varepsilon \left|S_\gamma(s) - S_\gamma(s-\sigma)\right|\,\mathrm{d}\sigma \\
&\leq \frac{1}{\varepsilon}\int_{-\varepsilon}^\varepsilon |\sigma|\,\mathrm{d}\sigma = \varepsilon.
\end{aligned}$$

We choose $\varepsilon(\gamma) = \gamma^{-1}$ and get

$$\begin{aligned}
2^{-p}\left\|S_\gamma^{\varepsilon(\gamma)}(u) - u\right\|_{\mathrm{L}^p}^p &\leq \left\|S_\gamma^{\varepsilon(\gamma)}(u) - S_\gamma(u)\right\|_{\mathrm{L}^p}^p + \|S_\gamma(u) - u\|_{\mathrm{L}^p}^p \\
&\leq \lambda^N(\Omega)\gamma^{-p} + \int_{\{s\in\Omega \;:\; |u(s)|\geq\gamma\}} |u|^p \,\mathrm{d}x.
\end{aligned}$$

Hence $S_\gamma^{\varepsilon(\gamma)}(u) \to u$ strongly in $\mathrm{L}^p(\Omega)$ as $\gamma \to 0^+$, and therefore $v \in \partial J(u)$ due to the strong closedness of ∂J. □

Remark 4.4.4. From Theorem 4.4.3 it follows that requirement (R8) is satisfied: (R8a) follows directly and (R8b) from the fact that $K^* = \mathrm{Id}$ and

$$v \in \partial J(u) \;\Leftrightarrow\; u \in \partial J^*(v)$$

according to Lemma A.2.12. Together with the argumentation in Section 4.1 this shows that Assumption 2.1.1 and Assumption 3.1.4 are satisfied and all general results deduced in the first part of the thesis are applicable.

4.5 Inverse Total Variation Flow

In this section we finally return to the augmented Lagrangian Algorithm 2.2.9 and the evolution equation (3.4) rewritten in the current framework (cf. Section 4.1). To this end, recall the definition of $J = J_p$ in (4.1) and assume that $v_0 \in \partial J(u_0)$ for an element $u_0 \in D(\partial J) \subset \mathrm{BV}(\Omega)_p$. In order to keep the presentation as transparent as possible, we will henceforth assume that $v_0 = 0$.

Further, let $\boldsymbol{\alpha} = \{\alpha_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ be a sequence of positive parameters such that

$$\lim_{n\to\infty} t_n(\boldsymbol{\alpha}) = \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{\alpha_j} = \infty.$$

The augmented Lagrangian algorithm reads as

Algorithm 4.5.1 (Augmented Lagrangian method for denoising)**.** Let $f \in \mathrm{L}^p(\Omega)$. For $n = 1, 2, \ldots$ compute

$$\mathcal{R}_n(f) := u_n \;=\; \operatorname*{argmin}_{u\in\mathrm{L}^p(\Omega)} \frac{1}{p}\int_\Omega |u-f|^p \,\mathrm{d}x + \alpha_n \left(J(u) - \int_\Omega v_{n-1} u \,\mathrm{d}x \right), \tag{4.26a}$$

$$\mathcal{R}_n^*(f) := v_n \;=\; v_{n-1} - \frac{1}{\alpha_n} |u_n - f|^{p-1} \operatorname{sign}(u_n - f) \in \partial J(u_n). \tag{4.26b}$$

Proposition 2.2.15 in Section 2.2 provides a dual characterization of $\{\mathcal{R}_n^*(f)\}_{n\in\mathbb{N}}$. We therefore remark that the functional $F^*(\cdot; f)$ in (2.13) in the current setting reads as

$$F^*(v;f) = \begin{cases} -\int_\Omega vf \,\mathrm{d}x & \text{if } v \in B_*^{p^*} \\ +\infty & \text{else.} \end{cases}$$

Moreover, it is of interest that for $f \in \mathrm{BV}(\Omega)_p$

$$\mu^*(f) := \inf_{v\in\mathrm{L}^{p^*}(\Omega)} F^*(v;f) = \inf_{v\in B_*^{p^*}} -\int_\Omega fv \,\mathrm{d}x = -\sup_{v\in B_*^{p^*}} \int_\Omega fv \,\mathrm{d}x = -J(f).$$

Thus we have that $J(f) + \mu^*(f) = 0$ for all $f \in \mathrm{BV}(\Omega)_p$ and consequently the infimum of $F^*(\cdot; f)$ is attained if and only if $f \in D(\partial J)$ according to Proposition 2.2.19. This is of course in accordance with Theorem 4.4.2. With this we can rewrite the proximal point method w.r.t. $F^*(\cdot; y)$ to

Algorithm 4.5.2. Let $f \in \mathrm{L}^p(\Omega)$. For $n = 1, 2, \ldots$ compute

$$v_n = \operatorname*{argmin}_{v\in B_*^{p^*}} \frac{\alpha^{p^*-1}}{p^*} \int_\Omega |v - v_{n-1}|^{p^*} \,\mathrm{d}x - \int_\Omega fv \,\mathrm{d}x.$$

Following Proposition 2.2.15 we find that $\{\mathcal{R}_n^*(f)\}_{n\in\mathbb{N}}$ is characterized by the proximal point method in Algorithm 4.5.2.

Recall that for the weight function $\phi(s) = s^{p-1}$ one has that $\psi_\phi(s) = \frac{1}{p}s^p$. Hence, Theorem 2.3.4 provides the following asymptotic estimate

Proposition 4.5.3. *Let $u \in BV(\Omega)_p$ and $f \in L^p(\Omega)$. For all $n \in \mathbb{N}$ one has*

$$\frac{1}{p}\int_\Omega |\mathcal{R}_n(f) - f|_{L^p}^p \; dx \le \frac{J(u)}{t_n(\boldsymbol{\alpha})} + \frac{1}{p}\int_\Omega |u-f|^p \; dx \tag{4.27}$$

and in particular

$$\lim_{n\to\infty} \mathcal{R}_n(f) = f. \tag{4.28}$$

Proof. Estimate (4.27) follows from Theorem 2.3.4 with $K = \mathrm{Id}$ and $\psi_\phi(s) = \frac{1}{p}s^p$. Moreover, since $\mathrm{BV}(\Omega)_p$ is a L^p-dense subset of $\mathrm{L}^p(\Omega)$, we can find for a given f and $\varepsilon > 0$ an element $u_\varepsilon \in \mathrm{BV}(\Omega)_p$ such that $\|f - u_\varepsilon\|_{\mathrm{L}^p} \le \varepsilon$. Thus (4.27) implies

$$\limsup_{n\to\infty} \int_\Omega |\mathcal{R}_n(f) - f|^p \, \mathrm{d}x \le \varepsilon^p.$$

Since $\varepsilon > 0$ is arbitrary small the assertion follows. □

Remark 4.5.4. Let $p = 2$ and $u \in \mathrm{BV}(\Omega)_2$. Moreover assume that for $\{u_n\}_{n\in\mathbb{N}} \subset \mathrm{L}^2(\Omega)$ one has that $\delta_n := \|u - u_n\|_{\mathrm{L}^2} \to 0$ as $n \to \infty$ and let $\Gamma : (0,\infty) \times \mathrm{L}^2(\Omega) \to \mathbb{N}$ be such that

$$\lim_{n\to\infty} \delta_n^2 t_{\Gamma(\delta_n,u_n)}(\boldsymbol{\alpha}) = 0 \quad \text{and} \quad \lim_{n\to\infty} t_{\Gamma(\delta_n,u_n)}(\boldsymbol{\alpha}) = \infty.$$

Then it follows from (4.27) and Theorem 2.4.4 that

$$\lim_{n\to\infty} \left\|\mathcal{R}_{\Gamma(\delta_n,u_n)}(u_n) - u\right\|_{\mathrm{L}^2} = 0 \quad \text{and} \quad \lim_{n\to\infty} J(\mathcal{R}_{\Gamma(\delta_n,u_n)}(u_n)) = J(u).$$

We will now prove a strong maximum principle for Algorithm 2.2.9. In short, this means that for bounded data $f \in \mathrm{L}^\infty(\Omega)$ the iterates are supposed to lie within the same bounds as f. Before we do so we prove the following

Lemma 4.5.5. Let $\gamma > 0$ and $u \in \mathrm{BV}(\Omega)$. Let $S_\gamma^+ : \mathbb{R} \to [0,\gamma]$ (resp. $S_\gamma^- : \mathbb{R} \to [-\gamma, 0]$) continuously differentiable and increasing (resp. decreasing). Moreover, assume that

$$\lim_{s\to\mp\infty} S_\gamma^\pm(s) = 0, \quad \lim_{s\to\pm\infty} S_\gamma^\pm(s) = \pm\gamma \quad \text{and} \quad \left|\frac{dS_\gamma^\pm(s)}{ds}\right| \le 1. \tag{4.29}$$

Then

$$\left|\mathrm{D}(S_\gamma^\pm \circ u)\right|(A) \le |\mathrm{D}u|(A)$$

for all Borel sets $A \subset \Omega$.

Proof. We prove the lemma for S_γ^+, for the arguments for second case being completely identical. Let $V \subset \Omega$ be open. According to [55, Chap.5.2. Thm.2] there exists a sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathrm{BV}(\Omega) \cap C^\infty(\Omega)$, such that

$$\lim_{n\to\infty} \|u_n - u\|_{\mathrm{L}^1} = 0 \quad \text{and} \quad \lim_{n\to\infty} |\mathrm{D}u_n|(V) = |\mathrm{D}u|(V).$$

From (4.29) it is evident that the mapping $S_\gamma^+ : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous (with a Lipschitz constant $L \le 1$) and thus we conclude that $\left\|S_\gamma^+ \circ u_n - S_\gamma^+ \circ u\right\|_{\mathrm{L}^1} \to 0$ as $n \to \infty$. Moreover, we find from the chain rule that for all $x \in \Omega$

$$\left|\nabla(S_\gamma^+ \circ u_n)(x)\right| = \left|\frac{dS_\gamma^+}{ds}(u_n(x))\nabla(u_n(x))\right| \le |\nabla u(x)|$$

and thus

$$\left|\mathrm{D}(S_\gamma^+ \circ u_n)\right|(V) = \int_V \left|\nabla(S_\gamma^+ \circ u_n)(x)\right| \,\mathrm{d}x \leq \int_V |\nabla u(x)| \,\mathrm{d}x = |\mathrm{D}u_n|(V).$$

From L^1-lower semicontinuity of $|\mathrm{D}\cdot|(V)$ we finally obtain

$$\left|\mathrm{D}(S_\gamma^+ \circ u)\right|(V) \leq \liminf_{n\to\infty} \left|\mathrm{D}(S_\gamma^+ \circ u_n)\right|(V) \leq \liminf_{n\to\infty} |\mathrm{D}u_n|(V) = |\mathrm{D}u|(V)$$

and the assertion is shown. □

Theorem 4.5.6. *Let $f \in L^\infty(\Omega)$. Then for all $n \in \mathbb{N}$ and a.e. $x_0 \in \Omega$ one has*

$$\operatorname*{ess\,inf}_{x\in\Omega} f(x) \leq [\mathcal{R}_n(f)](x_0) \leq \operatorname*{ess\,sup}_{x\in\Omega} f(x).$$

Proof. Let $f \in \mathrm{L}^\infty(\Omega)$ and note the abbreviations

$$u_n := \mathcal{R}_n(f) \quad \text{ and } \quad v_n := \mathcal{R}_n^*(f).$$

From Corollary 4.3.8 we conclude that for all $n \in \mathbb{N}$ there exists a $|\mathrm{D}u_n|$-measurable function $\vartheta : (v_{n-1}, \mathrm{D}u_n, \cdot) : \Omega \to \mathbb{R}$ such that $|\vartheta(v_{n-1}, \mathrm{D}u_n, x)| \leq 1$ for $|\mathrm{D}u_n|$-a.e. $x \in \Omega$ and

$$\int_\Omega v_{n-1} u_n \,\mathrm{d}x = \int_\Omega \vartheta(v_{n-1}, \mathrm{D}u_n, x) \,\mathrm{d}\,|\mathrm{D}u_n|(x).$$

Let $\varepsilon > 0$, $\gamma := \operatorname{ess\,sup}_{x\in\Omega} f(x)$ and define $S_\gamma^{+,\varepsilon} = S_\gamma^+ * \eta_\varepsilon$, where S_γ^+ denotes the upper truncation operator

$$S_\gamma^+(s) := \begin{cases} \gamma & \text{if } s \geq \gamma \\ s & \text{else.} \end{cases}$$

and η_ε denotes a standard mollifier. Therefore, the function $S_\gamma^{+,\varepsilon}$ satisfies the requirements in Lemma 4.5.5. This and the estimate

$$1 - |\vartheta(v_{n-1}, \mathrm{D}u_n, x)| \geq 0 \quad \text{ for } |\mathrm{D}u_n| \text{-a.e. } x \in \Omega$$

result in

$$\begin{aligned} J(u_n) - \int_\Omega v_{n-1} u_n \,\mathrm{d}x &= \int_\Omega (1 - \vartheta(v_{n-1}, \mathrm{D}u_{n-1}, x)) \,\mathrm{d}\,|\mathrm{D}u_n|(x) \\ &\geq \int_\Omega (1 - \vartheta(v_{n-1}, \mathrm{D}u_{n-1}, x)) \,\mathrm{d}\left|\mathrm{D}(S_\gamma^{+,\varepsilon} \circ u_n)\right|(x). \end{aligned} \tag{4.30}$$

Now observe that according to Corollary 4.3.8

$$\vartheta(v_{n-1}, \mathrm{D}u_{n-1}, x) = \vartheta(v_{n-1}, \mathrm{D}(S_\gamma^{+,\varepsilon} \circ u_{n-1}), x)$$

$|\mathrm{D}u_n|$-a.e. and $\left|\mathrm{D}(S_\gamma^{+,\varepsilon} \circ u_n)\right|$-a.e. in Ω. This and (4.30) hence imply

$$\begin{aligned} J(u_n) - \int_\Omega v_{n-1} u_n \,\mathrm{d}x &\geq \int_\Omega (1 - \vartheta(v_{n-1}, \mathrm{D}(S_\gamma^{\varepsilon,+} \circ u_{n-1}), x)) \,\mathrm{d}\left|\mathrm{D}(S_\gamma^{+,\varepsilon} \circ u_n)\right|(x) \\ &= J(S_\gamma^{+,\varepsilon} \circ u_n) - \int_\Omega \vartheta(v_{n-1}, \mathrm{D}(S_\gamma^{+,\varepsilon} \circ u_{n-1}), x)) \,\mathrm{d}\left|\mathrm{D}(S_\gamma^{\varepsilon,+} \circ u_n)\right|(x) \\ &= J(S_\gamma^{+,\varepsilon} \circ u_n) - \int_\Omega v_{n-1}(S_\gamma^{+,\varepsilon} \circ u_n) \,\mathrm{d}x. \end{aligned}$$

Now define $\bar{u}_n := S_\gamma^+ \circ u_n$. Then as in the proof of Theorem 4.4.3 we conclude from the fact that $S_\gamma^{+,\varepsilon} \to S_\gamma^+$ uniformly as $\varepsilon \to 0^+$ that

$$\lim_{\varepsilon \to 0^+} \int_\Omega \left| S_\gamma^{+,\varepsilon} \circ u_n - \bar{u}_n \right|^p \, \mathrm{d}x = 0.$$

Lower semicontinuity thus gives

$$\begin{aligned} J(\bar{u}_n) - \int_\Omega v_{n-1} \bar{u}_n \, \mathrm{d}x \leq \liminf_{\varepsilon \to 0^+} \left(J(S_\gamma^{+,\varepsilon}(u_n)) - \int_\Omega v_{n-1}(S_\gamma^{+,\varepsilon} \circ u_n) \, \mathrm{d}x \right) \\ \leq J(u_n) - \int_\Omega v_{n-1} u_n \, \mathrm{d}x. \end{aligned}$$

Moreover it is easy to see that

$$\frac{1}{p} \int_\Omega |\bar{u}_n - f|^p \, \mathrm{d}x \leq \frac{1}{p} \int_\Omega |u_n - f|^p \, \mathrm{d}x.$$

Combination of the previous two estimates with optimality in (4.26a) shows $\bar{u}_n = u_n$ and therefore

$$\operatorname*{ess\,sup}_{x \in \Omega} u_n(x) \leq \gamma.$$

Clearly we can proceed analogously for $\gamma = \operatorname{ess\,inf}_{x \in \Omega} f(x)$ and the lower truncation operator S_γ^- and thus nothing remains to be proven. □

We will now turn out attention to evolution equation (3.4) in the current setting. Let $f \in \mathrm{L}^p(\Omega)$ be the given (noisy) data. From Theorem 4.4.2 it follows that (3.4) can equivalently be written as

$$dv(t) = |f - u(t)|^{p-1} \operatorname{sign}(f - u(t)), \qquad \int_\Omega u(t) v(t) \, \mathrm{d}x = J(u(t)), \tag{4.31a}$$

$$v(0) = 0, \qquad \|v(t)\|_* \leq 1. \tag{4.31b}$$

Assume that $\{\boldsymbol{\alpha}_\nu\}_{\nu \in \mathbb{N}}$ is a sequence of partitions of $[0, \infty)$ (cf. Definition 3.2.1), satisfying

$$\lim_{\nu \to \infty} |\boldsymbol{\alpha}_\nu| = \infty.$$

Moreover let $\{\boldsymbol{u}_\nu\}_{\nu \in \mathbb{N}}$ and $\{\boldsymbol{v}_\nu\}_{\nu \in \mathbb{N}}$ be the corresponding sequences defined by Algorithm 4.5.1 w.r.t. to the data f, initial value $v_0 = 0$ and the sequence of parameters $\boldsymbol{\alpha}_\nu$. As in Section 3.3 (see Table 3.1) we use the abbreviations

$$u_\nu(t) := c(\boldsymbol{\alpha}_\nu, \boldsymbol{u}_\nu)(t) \quad \text{and} \quad v_\nu(t) := l(\boldsymbol{\alpha}_\nu, \boldsymbol{v}_\nu)(t).$$

for the piecewise constant and piecewise affine interpolants respectively (cf. Definition 3.2.2).

From the considerations in Section 4.1 and from Remark 4.4.4 it becomes evident that all assumptions of Chapter 2 ((R1) - (R6)) as well as Chapter 3 ((R7) and (R8)) are satisfied.

Moreover, since the spaces $\mathrm{L}^p(\Omega)$ are uniformly convex for $1 < p < \infty$ (see Clarkson [43] for the pioneering work) they are in particular strictly convex and thus E-spaces (cf. [94, Prop. 5.3.16]). Therefore the requirements of Corollary 3.3.6 and Theorem 3.3.10 are satisfied and one gets

Proposition 4.5.7. *There exist solutions* $(u, v) : [0, \infty) \to L^p(\Omega) \times L^{p^*}(\Omega)$ *for Equation* (4.31) *such that*

$$\lim_{\nu \to \infty} \|v_\nu(t) - v(t)\|_{L^{p^*}} = 0 \ \text{ for all } \ t > 0$$

and $v(0) = 0$. *Furthermore one has for all* $T > 0$ *that*

$$\lim_{\nu \to \infty} u_\nu = u, \ \text{ in } \ L^p((0,T) \times \Omega) \quad \text{ and } \quad \lim_{\nu \to \infty} v_\nu = v \ \text{ in } \ W^{1,p}((0,T) \times \Omega).$$

and $u(t) \in BV(\Omega)_p$ *for all* $t > 0$.

As in the discrete situation above we introduce for $t \geq 0$

$$\mathcal{R}_t(f) = u(t) \quad \text{and} \quad \mathcal{R}_t^*(f) = v(t)$$

where $(u(t), v(t))$ denote solutions of (4.31) as in Proposition 4.5.7 w.r.t. the data f. Then Theorem 3.4.3 generalizes the asymptotic estimate in Proposition 4.5.3 by simply replacing $\mathcal{R}_n(f)$ by $\mathcal{R}_t(f)$ and $t_n(\boldsymbol{\alpha}))$ by t. In particular, on has

$$\lim_{t \to \infty} \mathcal{R}_t(f) = f. \tag{4.32}$$

Moreover, from Theorem 4.5.6 we immediately find

Corollary 4.5.8. Let $f \in \mathrm{L}^\infty(\Omega)$. Then for all $t \geq 0$ and a.e. $x \in \Omega$ one has

$$\operatorname*{ess\,inf}_{x \in \Omega} f(x) \leq [\mathcal{R}_t(f)](x) \leq \operatorname*{ess\,sup}_{x \in \Omega} f(x).$$

From Proposition 4.5.7 it follows that for a given $f \in \mathrm{L}^p(\Omega)$ the mapping $t \mapsto \mathcal{R}_t(f)$ is measurable and locally p-summable; properties that are not sufficient in order to predict the behavior of solutions for $t \to 0^+$. The next theorem will shed some more light on this problem

Theorem 4.5.9. *Let* $f \in L^p(\Omega)$. *Then,*

$$\lim_{t \to \infty} \mathcal{R}_t(f) = f \quad \text{and} \quad \lim_{t \to 0^+} \mathcal{R}_t(f) = c(f), \tag{4.33}$$

where both limits hold w.r.t the strong L^p*-topology. Here,* $c(f) \in \mathbb{R}$ *is (the unique constant) such that*

$$\int_\Omega |c(f) - f|^{p-1} \, \mathit{sign}(c(f) - f) \, dx = 0.$$

Proof. For the sake of simplicity we write

$$u(t) = \mathcal{R}_t(f) \quad \text{and} \quad v(t) = \mathcal{R}_t^*(f).$$

As augmented above (cf. (4.32)), the first limit in (4.33) already follows from Theorem 3.4.3 (with $K = \mathrm{Id}$ and $\psi_\phi(s) = \frac{1}{p} s^p$).

Therefore let $\{t_n\}_{n \in \mathbb{N}} \subset (0, 1)$ be such that $t_n \to 0^+$ as $n \to \infty$. Then we have

$$\sup_{n \in \mathbb{N}} J(u(t_n)) < \infty \quad \text{and} \quad \sup_{n \in \mathbb{N}} \|u(t_n) - f\|_{\mathrm{L}^p} < \infty,$$

as a result of which we conclude that there exists a selection $n \mapsto \rho(n)$ and an element $u^* \in \mathrm{BV}(\Omega)_p$ such that $u(t_{\rho(n)}) \to u^*$ weakly in $\mathrm{L}^p(\Omega)$ and weakly* in $\mathrm{BV}(\Omega)$. Moreover, we have that

$$v(t_{\rho(n)}) \in \partial J(u(t_{\rho(n)}))$$

and hence $0 = v_0 \in \partial J(u^*)$, due to the strong-weak closedness of ∂J (cf. Lemma A.2.2 (1)) and the fact that $v(t_{\rho(n)}) \to 0$. Thus $J(u^*) = 0$ and $u^* \equiv c$ for some $c \in \mathbb{R}$. We shall prove that $c = c(f)$.

For $\gamma > 0$ let $S_\gamma(r)$ be the truncation operator in (4.22). Since $u(t_{\rho(n)}) \rightharpoonup^* u^*$ in $\mathrm{BV}(\Omega)$ we particularly have that $u(t_{\rho(n)}) \to u^*$ in $\mathrm{L}^1(\Omega)$ and hence we conclude as in the proof of Theorem 4.4.3 that

$$\begin{aligned} \lim_{n\to\infty} (S_\gamma \circ (u(t_{\rho(n)}) - f)) &= S_\gamma \circ (u^* - f) \quad \text{in } \mathrm{L}^p(\Omega) \\ \lim_{\gamma\to\infty} S_\gamma \circ (u^* - f) &= u^* - f \quad \text{in } \mathrm{L}^p(\Omega). \end{aligned}$$

Choose a sequence $\{\gamma(n)\}_{n\in\mathbb{N}}$ such that $\gamma(n) \to \infty$ as $n \to \infty$ and

$$\lim_{n\to\infty} (S_{\gamma(n)} \circ (u(t_{\rho(n)}) - f)) = u^* - f.$$

According to [42, Prop. 4.8] we have that the duality mapping $\mathfrak{J}_p = |\cdot|^{p-1} \operatorname{sign}(\cdot)$ is norm-norm continuous. Additionally, from the definition of $\mathfrak{J}_p$ it is evident that for an arbitrary $w \in \mathrm{L}^p(\Omega)$

$$\mathfrak{J}_p(S_\gamma \circ w) = S_{\gamma^{p-1}} \circ \mathfrak{J}_p(w).$$

Combination of these observations together with (4.31) results in

$$\begin{aligned} \lim_{n\to\infty} (S_{\gamma(n)^{p-1}} \circ v'(t_{\rho(n)})) &= \lim_{n\to\infty} (S_{\gamma(n)^{p-1}} \circ \mathfrak{J}_p(f - u(t_{\rho(n)})) \\ &= \lim_{n\to\infty} \mathfrak{J}_p(S_\gamma \circ (f - u(t_{\rho(n)}))) = \mathfrak{J}_p(u^* - f). \end{aligned} \tag{4.34}$$

For $n \in \mathbb{N}$ we introduce the sets

$$\Omega_n^+ := \left\{ x \in \Omega \ : \ v'(t_{\rho(n)}) > \gamma(n)^{p-1} \right\} \quad \text{and} \quad \Omega_n^- := \left\{ x \in \Omega \ : \ v'(t_{\rho(n)}) < -\gamma(n)^{p-1} \right\}. \tag{4.35}$$

Since $v(t) \in \mathrm{L}^{p^*}_\Diamond(\Omega)$ for all $t > 0$ it follows that $v'(t_{\rho(n)}) \in \mathrm{L}^{p^*}_\Diamond(\Omega)$ for all $n \in \mathbb{N}$ and consequently one finds with Hölder's inequality that

$$\begin{aligned} \left| \int_\Omega S_{\gamma(n)^{p-1}}(v'(t_{\rho(n)}))\,\mathrm{d}x \right| &= \left| \int_\Omega S_{\gamma(n)^{p-1}}(v'(t_{\rho(n)}))\,\mathrm{d}x - \int_\Omega v'(t_{\rho(n)})\,\mathrm{d}x \right| \\ &= \int_{\Omega_n^+} v(t_{\rho(n)}) - \gamma^{p-1}\,\mathrm{d}x - \int_{\Omega_n^-} v(t_{\rho(n)}) + \gamma^{p-1}\,\mathrm{d}x \\ &\leq \int_{\Omega_n^+} v(t_{\rho(n)})\,\mathrm{d}x - \int_{\Omega_n^-} v(t_{\rho(n)})\,\mathrm{d}x \\ &\leq (\lambda^N(\Omega_n^+) + \lambda^N(\Omega_n^-)) \left\| v'(t_{\rho(n)}) \right\|_{\mathrm{L}^{p^*}}. \end{aligned} \tag{4.36}$$

Since $\left\| v'(t_{\rho(n)}) \right\|_{\mathrm{L}^{p^*}}$ is uniformly bounded in $n \in \mathbb{N}$ and since $\gamma(n) \to \infty$ as $n \to \infty$ it follows (4.35) that

$$\lim_{n\to\infty} \lambda^N(\Omega_n^\pm) = 0.$$

Thus we can apply the dominated convergence theorem and find from (4.34) and (4.36) that

$$\int_\Omega \mathfrak{J}_p(u^* - f) = \int_\Omega \lim_{n\to\infty} S_{\gamma(n)^{p-1}}(v'(t_{\rho(n)}))\,\mathrm{d}x = \lim_{n\to\infty} \int_\Omega S_{\gamma(n)^{p-1}}(v'(t_{\rho(n)}))\,\mathrm{d}x = 0$$

as desired.

From [4, Lem. 1.5.5] we learn that $\mathfrak{J}_p$ is a strict monotone operator, that is, for all $c_1 < 0 < c_2$ and all $w \in \mathrm{L}^p(\Omega)$ one has that

$$\int_\Omega \mathfrak{J}_p(w + c_1)\,\mathrm{d}x < \int_\Omega \mathfrak{J}_p(w)\,\mathrm{d}x < \int_\Omega \mathfrak{J}_p(w + c_2)\,\mathrm{d}x.$$

Therefore $c(f)$ is the only (constant) element satisfying $\int_\Omega \mathfrak{J}_p(c(f) - f) = 0$, wherefore we conclude with a standard sub-subsequence argument that $u(t_n) \to c(f)$ weakly in $\mathrm{L}^p(\Omega)$ and weakly* in $\mathrm{BV}(\Omega)$. Lower semicontinuity and monotonicity of the residual $\|u(t) - f\|$ then imply that

$$\|c(f) - f\| \le \liminf_{n\to\infty} \|u(t_n) - f\| \le \limsup_{n\to\infty} \|u(t_n) - f\| \le \|c(f) - f\|\,,$$

which shows — by the Radon-Riesz property of $\mathrm{L}^p(\Omega)$ — that $\|u(t_n) - c(f)\|_{\mathrm{L}^p} \to 0$ and the assertion is shown. □

Remark 4.5.10. In connection with image denoising the residual $\|f - u\|$, where f denotes the given noisy data and u some approximation, is often referred to as *data fidelity term*, since it measures the quality of the reconstruction.

If $\{u(t)\}_{t\ge 0} \subset X^{\mathbb{R}}$ is a scale of images in an image space X (e.g. $X = \mathrm{L}^p(\Omega)$) generated by a (noisy) image f, it is said, that the scale space has the *fidelity property* if the finest scale (usually $u(0)$) contains all information of f, that is $u(0) = f$. This notion suggests to call property (4.33) *inverse fidelity* and Equation (4.31) *inverse total variation flow equation.* A solution $\mathcal{R}_t(f)$ of the inverse total variation equation is coherently called an *inverse scale space.*

4.6 Exact Solutions and Multiscale Properties

In the following we discuss the multiscale properties of the inverse total variation equation. From the interpretation of (3.4) as an inverse scale space method, we expect that large scales are reconstructed for small times, while finer scales take a longer time to be included in the reconstruction. We provide an example giving a experimental justification. Throughout the whole section we make again the assumption that $v_0 = 0$. For a given function $f \in \mathrm{L}^p(\Omega)$ we will assume that $c(f) = 0$, where $c(f)$ is as in Theorem 4.5.9. In other words, $\mathfrak{J}_p(f) \in \mathrm{L}^{p^*}_\diamondsuit(\Omega)$.

Theorem 4.6.1. *Let $f \in L^p(\Omega)$ such that $\mathfrak{J}_p(f) \in \mathcal{L}^{p_*}_\diamondsuit(\Omega)$. Then, for $t\,\|\mathfrak{J}_p(f)\|_* \le 1$, a solution pair of the inverse total variation flow equation* (3.4) *is given by*

$$u(t) = 0 \quad \textit{and} \quad v(t) = t\;\mathfrak{J}_p(f). \tag{4.37}$$

Proof. In the following let $t \le \frac{1}{\|\mathfrak{J}_p(f)\|_*}$. We verify that (u, v) defined by (4.37) is indeed a solution pair of (3.4). Since $v'(t) = \mathfrak{J}_p(f) = \mathfrak{J}_p(f - u(t))$ it follows that

$$\mathfrak{J}_{p_*}(v') = f - u.$$

According to Theorem 4.3.2 there exists $z \in \ker(T_{p_*})$ such that $\operatorname{div}(z) = \mathfrak{J}_p(f)$ and

$$\|z\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^N} = \|\mathfrak{J}_p(f)\|_* .$$

Then, for all $\phi \in C_c^1(\Omega)$ we have

$$\begin{aligned} J(\phi) - J(u(t)) - \int_\Omega v(t)\ (\phi - u(t))\,\mathrm{d}x &= J(\phi) - \int_\Omega t\ \mathfrak{J}_p(f)\ \phi\,\mathrm{d}x \\ &= \int_\Omega (|\nabla\phi| + t\ z\ \nabla\phi)\ \mathrm{d}x \\ &\geq (1 - t\,\|\mathfrak{J}_p(f)\|_*) \int_\Omega |\nabla\phi|\ \mathrm{d}x \geq 0. \end{aligned}$$

By standard continuity and density arguments we can now extend the inequality

$$J(\phi) - J(u(t)) - \int_\Omega v(t)\ (\phi - u(t))\ \mathrm{d}x \geq 0$$

to all $\phi \in \mathrm{L}^p(\Omega)$ and hence, $v(t) \in \partial J(0) = \partial J(u(t))$. Thus, (v, u) is a solution pair of (3.4). □

Theorem 4.6.1 shows that the variable u does not change in an initial stage of the evolution up to time

$$t_*(f) := \frac{1}{\|\mathfrak{J}_p(f)\|_*}, \tag{4.38}$$

while the dual variable changes at linear rate in time. It is important to note that this result only holds for data satisfying $\mathfrak{J}_p(f) \in \mathcal{L}_\Diamond^{p^*}(\Omega)$. If, however, $\|\mathfrak{J}_p(f)\|_* = \infty$, then Theorem 4.6.1 gives no assertion about the initial behavior of solutions and one has to be satisfied with the asymptotic result in Theorem 4.5.9.

In order to illustrate the behavior for $t > t_*$, we consider calibrable functions (recall Definition 4.3.9). We continue with a characterization of the subgradient of J for this functionclass.

Lemma 4.6.2. Let $u \in \mathrm{L}^p(\Omega)$. Then

$$\frac{\mathfrak{J}_p(u)}{\|\mathfrak{J}_p(u)\|_*} \in \partial J(u) \quad \Leftrightarrow \quad u \text{ is calibrable.}$$

Proof. $\Rightarrow$: From Theorem 4.4.2 we conclude that

$$\int_\Omega \frac{\mathfrak{J}_p(u)}{\|\mathfrak{J}_p(u)\|_*} u\,\mathrm{d}x = J(u)$$

and hence

$$\|u\|_{\mathrm{L}^p}^p = \int_\Omega \mathfrak{J}_p(u) u\,\mathrm{d}x = J(u)\,\|\mathfrak{J}_p(u)\|_* .$$

Thus, u is calibrable.

$\Leftarrow$: For an arbitrary $w \in \mathrm{L}^p(\Omega)$ we have

$$\begin{aligned} J(u) + \int_\Omega \frac{\mathfrak{J}_p(u)}{\|\mathfrak{J}_p(u)\|_*}(w-u)\,\mathrm{d}x &= J(u) + \frac{1}{\|\mathfrak{J}_p(u)\|_*}\left(\int_\Omega (\mathfrak{J}_p(u) w\,\mathrm{d}x - \int_\Omega \mathfrak{J}_p(u) u\,\mathrm{d}x\right) \\ &= J(u) + \frac{1}{\|\mathfrak{J}_p(u)\|_*}\left(\int_\Omega \mathfrak{J}_p(u) w\,\mathrm{d}x - \|u\|_{\mathrm{L}^p}^p\right) \\ &= J(u) + \frac{1}{\|\mathfrak{J}_p(u)\|_*}\int_\Omega \mathfrak{J}_p(u) w\,\mathrm{d}x - J(u) \\ &\le \frac{1}{\|\mathfrak{J}_p(u)\|_*}\|\mathfrak{J}_p(u)\|_* \, J(w) = J(w). \end{aligned}$$

□

The following theorem clarifies the importance of calibrable functions in connection with (3.4): They are *exactly* those functions which can be recovered in finite time.

Theorem 4.6.3. *Let $f \in L^p(\Omega)$ such that $\mathfrak{J}_p(f) \in L_\diamondsuit^{p^*}(\Omega)$ and let $t_* > 0$. Then the following two statements are equivalent*

1. *The pair*
$$u(t) = f \qquad \text{and} \qquad v(t) = t_*\,\mathfrak{J}_p(f)$$
is a solution of (3.4) *for $t \ge t_*$.*
2. *The function f is calibrable and $t_* = t_*(f)$.*

Proof. (1) $\Rightarrow$ (2): If (u,v) solves (3.4) then $v(t) \in \partial J(u(t))$ for $t \ge t_*$ and therefore $\int_\Omega u(t)v(t)\,\mathrm{d}x = J(u(t))$. From Corollary 4.3.8 we hence find that

$$\frac{J(f)}{t_*} = \int_\Omega \mathfrak{J}_p(f) f\,\mathrm{d}x = \|f\|_{\mathrm{L}^p}^p \le J(f)\,\|\mathfrak{J}_p(f)\|_*$$

and thus $1 \le t_*\,\|\mathfrak{J}_p(f)\|_*$. However from Theorem 4.4.2 it follows that $\|t_*\,\mathfrak{J}_p(f)\|_* \le 1$. Consequently $t_*\,\|\mathfrak{J}_p(f)\|_* = 1$, i.e. $t_* = t_*(f)$ and f is calibrable.

(2) $\Rightarrow$ (1): We have noticed in Lemma 4.6.2 that for calibrable functions the inclusion

$$v(t) = t_*(f)\;\mathfrak{J}_p(f) = \frac{\mathfrak{J}_p(f)}{\|\mathfrak{J}_p(f)\|_*} \in \partial J(f) = \partial J(u(t))$$

holds. Moreover, $\mathfrak{J}_{p_*}(v'(t)) = 0 = f - u(t)$ is obviously satisfied. Hence (u,v) is a solution of (3.4). □

Remark 4.6.4. Combining Theorems 4.6.1 and 4.6.3 shows that for calibrable initial data f a solution pair (u,v) of (3.4) is given by

$$u : t \to \begin{cases} 0 & \text{if } 0 \le t \le t_*(f), \\ f & \text{else} \end{cases} \qquad \text{and} \qquad v : t \to \begin{cases} t\;\mathfrak{J}_p(f) & \text{if } 0 \le t \le t_*(f), \\ t_*(f)\;\mathfrak{J}_p(f) & \text{else.} \end{cases}$$

We illustrate this result by means of a simple example.

Example 4.6.5. Assume that $p = N = 2$ and $0 < r < R$. We set $\Omega = B_R(0)$ and define for $(x, y) \in \Omega$

$$u_{R,r}(x,y) = \frac{R^2 - r^2}{R^2}\chi_{B_r(0)} - \frac{r^2}{R^2}\chi_{B_R(0)\setminus B_r(0)}.$$

Moreover we introduce the vector field $z_{R,r}(x,y) = g_{R,r}(\sqrt{x^2+y^2})(x,y)^T$, where

$$g_{R,r}(s) = \begin{cases} \frac{R^2-r^2}{2R^2} & \text{for } 0 \le s \le r, \\ \frac{r^2}{2s^2} - \frac{r^2}{2R^2} & \text{for } r < s \le R. \end{cases}$$

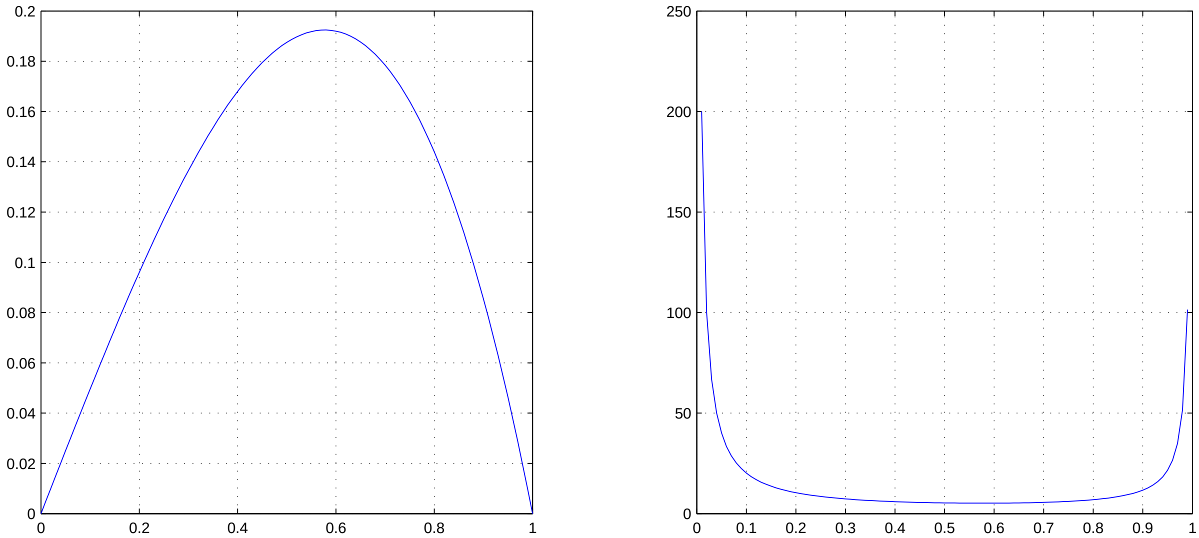

Figure 4.1: Left: g-norm $\|u_{1,r}\|_*$ for $0 \le r \le 1$. Right: Reconstruction time $t_*(u_{1,r})$ for $0 \le r \le 1$.

We obviously have that $\operatorname{div}(z_{R,r}) = u_{R,r}$ and moreover $z_{R,r}(x,y) = (0,0)$, whenever $\sqrt{x^2+y^2} = R$, which implies, that $u_{R,r} \in \mathcal{L}^2_\Diamond(\Omega)$. Direct computation gives

$$|\mathrm{D}u_{R,r}|(\Omega)\,\|z_{R,r}\|_{\mathrm{L}^\infty(\Omega,\mathbb{R}^2)} = \frac{(R^2-r^2)r^2\pi}{R^2} = \|u_{R,r}\|^2_{\mathrm{L}^2}.$$

Thus, $u_{R,r}$ is calibrable and one has

$$\|u_{R,r}\|_* = \frac{(R^2-r^2)r}{2R^2} \qquad \text{and} \qquad t_*(u_{R,r}) = \frac{2R^2}{(R^2-r^2)r}.$$

We note that the smaller the spatial features are (that is the smaller $\|u_{R,r}\|_*$ is) the longer it takes to recover the signal. By increasing the width r of the peak the reconstruction time decreases as long as $r \le R/\sqrt{3}$. Beyond this point the negative part of $u_{R,r}$, that is, $\{(x,y) \in \Omega \,:\, |(x,y)| \ge r\}$, behaves like a peak and therefore again a larger reconstruction time is required.

Figure 4.1 depicts the g-norm $\|u_{1,r}\|_*$ (left image) and the reconstruction time $t_*(u_{1,r})$ (right image) for $0 \le r \le 1$.

The results collected in Remark 4.6.4 lead to the suggestion that also for general (not necessarily calibrable) images small features take longer to be reconstructed than large ones. This is confirmed by the following (numerical) example.

Example 4.6.6. As in Example 4.6.5 we set $p = N = 2$ and assume that the image f (as depicted in Figure 4.2) is composed of geometric objects of different intensities and sizes.

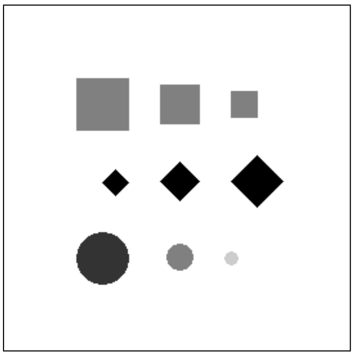

Figure 4.2: Image f composed of geometric objects.

We compute the solution $u(t)$ of (4.31) for $t = 0 \dots 1.5$. The left image in Figure 4.3 shows the residuals $\|f - u(t)\|_{\mathrm{L}^2}$ (solid line) and the a-priori bound given by Proposition 4.5.3 (dashed line) given by

$$\|f - u(t)\|_{\mathrm{L}^2} \leq \sqrt{\frac{2J(f)}{t}}.$$

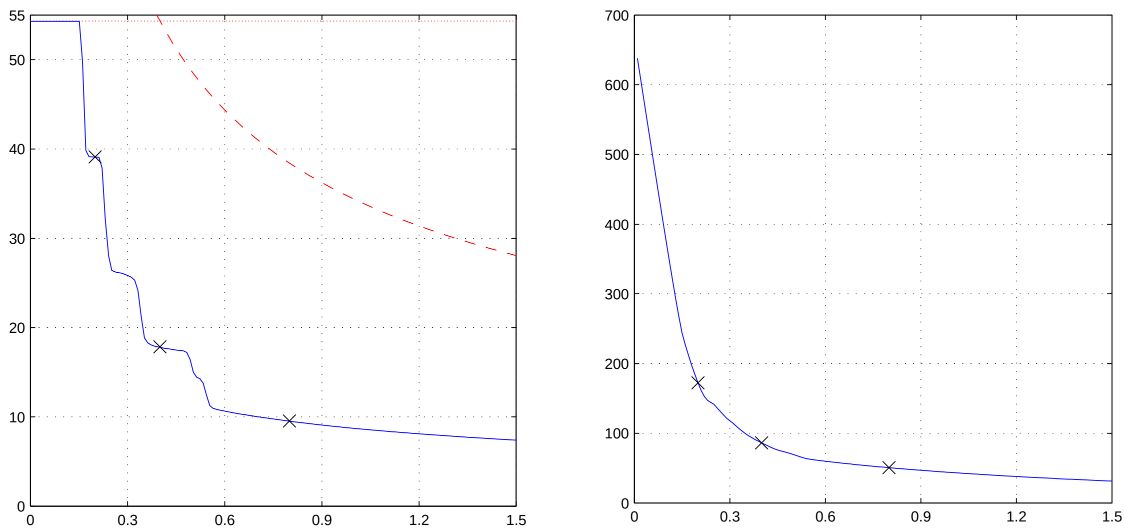

Figure 4.3: Left: Residual $\|f - u(t)\|_{\mathrm{L}^2}$ (solid line) and a-priori bound (dashed line). Right: Bregman distance $D_J^{v(t)}(f, u(t))$

The cascaded structure of the graph of $\|f - u(t)\|_{\mathrm{L}^2}$ indicates that the solution $u(t)$ behaves as expected, that is, starting from a constant image with value (cf. Theorem 4.6.1)

$$u(t) = \int_\Omega f \, \mathrm{d}x, \qquad \text{for } 0 \le t \le t_*(f)$$

one after the other object is reconstructed (almost) *instantaneously* where the order of reconstruction is determined by the feature size (i.e. geometrical size *and* intensity). The right image in Figure 4.3 depicts the Bregman distance $D_J^{v(t)}(f, u(t))$ of f and $u(t)$ w.r.t. J and $v(t)$ (right image). Recall that $v(t) \in \partial J(u(t))$ is the dual solution of (4.31). From Theorem 4.4.2 it follows that that

$$D_J^{v(t)}(f, u(t)) = J(f) - J(u(t)) - \int_\Omega v(t)(f - u(t)) \, \mathrm{d}x = J(f) - \int_\Omega v(t) f \, \mathrm{d}x. \tag{4.39}$$

Figure 4.4 displays the solutions $u(t)$ at time $t = 0.2, 0.4$ and 0.8. We note, that $t_*(f) \approx 0.18$ according to Figure 4.3.

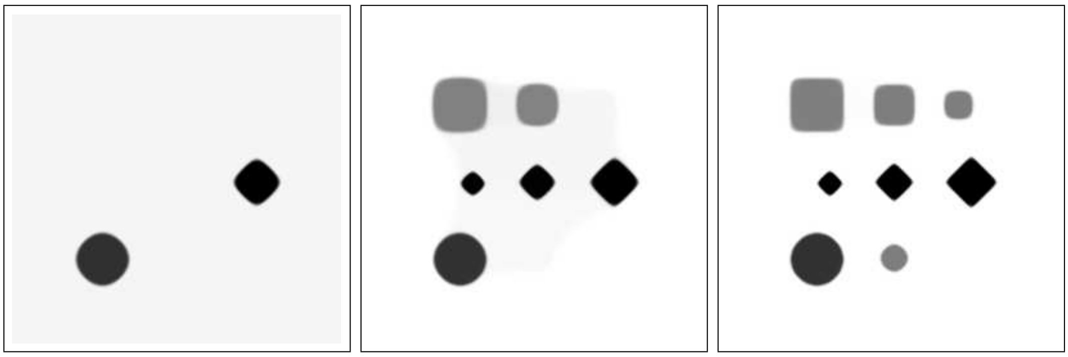

Figure 4.4: Solutions of (4.31) at times $t = 0.2, 0.4$ and 0.8.

For solving (4.31) numerically, the implicit time scheme given by Algorithm 4.5.1 with $\alpha_n = \alpha = 100$ (that is $\Delta t = 0.01$) was used. Each time step hence requires the computation of the minimizing element u_n of the functional

$$u \mapsto \frac{1}{2} \int_\Omega |u - (f + \alpha v_{n-1})|^2 \, \mathrm{d}x + \alpha J(u). \tag{4.40}$$

In order to tackle this problem we used the method proposed by Dobson & Vogel in [48]: For $\varepsilon > 0$ and u sufficiently smooth (e.g. $u \in \mathrm{W}^{1,\infty}(\Omega)$) we introduce

$$J_\varepsilon(u) := \int_\Omega \sqrt{|\nabla u|^2 + \varepsilon^2} \, \mathrm{d}x$$

as an approximation of the TV semi-norm J. Indeed, as it was shown e.g. in [1, Thm. 2.2], one has that

$$\lim_{\varepsilon \to 0^+} J_\varepsilon(u) = J(u).$$

Moreover, $J_\varepsilon : \mathrm{L}^2(\Omega) \to \overline{\mathbb{R}}$ is differentiable for smooth functions u and

$$\langle \partial J_\varepsilon(u), v \rangle = \int_\Omega \frac{\nabla u \cdot \nabla v}{\sqrt{|\nabla u|^2 + \varepsilon^2}}$$

for all test functions $v \in \mathcal{D}(\Omega)$. Thus the weak form of the Euler-Lagrange equations for (4.40) can be written as

$$\int_\Omega (u - f_{n-1})v + \alpha \frac{\nabla u \cdot \nabla v}{\sqrt{|\nabla u|^2 + \varepsilon^2}} \, dx = 0, \qquad \text{for all } v \in \mathcal{D}(\Omega) \tag{4.41}$$

where we set $f_{n-1} = f + \alpha v_{n-1}$. By setting $L(u) = \left(|\nabla u|^2 + \varepsilon^2\right)^{-1/2}$ we can rewrite (4.41) to

$$\int_\Omega uv + \alpha L(u) \nabla u \nabla v \, dx = \int_\Omega f_{n-1} v \, dx, \qquad \text{for all } v \in \mathcal{D}(\Omega).$$

We compute a numerical approximation of this equation by a finite element approach. To this end, choose $N \in \mathbb{N}$ and for $1 \leq i \leq N$ linear independent functions $\phi_i \in \mathrm{W}^{1,\infty}(\Omega)$ and set

$$V_N = \operatorname{span}\{\phi_1, \ldots, \phi_N\} \subset \mathrm{W}^{1,\infty}(\Omega) \subset \mathrm{L}^2(\Omega).$$

Solving (4.41) on the space V_N amounts to finding an element $u_N = \sum_{i=1}^N U_i \phi_i \in V_N$ such that

$$\sum_{i=1}^N U_i \int_\Omega \phi_i \phi_j + \alpha L(u_N) \nabla \phi_i \nabla \phi_j \, dx = \sum_{i=1}^N F_{n-1} \int \phi_i \phi_j \, dx \qquad \text{for } 1 \leq j \leq N. \tag{4.42}$$

Here $F_{n-1} \in \mathbb{R}^N$ denotes the coordinate vector of the orthogonal L^2-projection of f_{n-1} onto V_N. Note that due to the expression $L(u_N)$ this equation is *non-linear*. We overcome this problem by applying a fixed point iteration following [48]. To this end define for $u \in \mathrm{W}^{1,\infty}(\Omega)$ the matrix $S(u) \in \mathbb{R}^{N \times N}$ as

$$[S(u)]_{ij} = M_{ij} + \alpha \int_\Omega L(u) \nabla \phi_i \nabla \phi_j \, dx$$

where M is the mass matrix defined by $M_{ij} = \int_\Omega \phi_i \phi_j \, dx$. The fixed point iteration then reads as

- Choose an arbitrary initial guess $u^0 \in V_N$.
- For $k = 1, 2, \ldots$ compute the solution $U^k \in \mathbb{R}^N$ of the linear equation

  $$S(u^{k-1})U = MF_{n-1}$$

  and set

  $$u^k = \sum_{i=1}^N U_i^k \phi_i.$$

From [48, Thm. 4.1] it follows that this fixed point iteration converges for every initial guess $u^0 \in V_N$ to the solution of (4.42).

In the computations for Example 4.6.6 we used $\Omega = [0, 256]^2$ and the standard bilinear basis functions ϕ_i on a uniform grid with unit step size. The image f (and hence as well its projection onto V_N) has values in $[0, 1]$. In all our computations we used $\varepsilon = 10^{-8}$ for regularizing the TV semi-norm. The integral in the matrix $S(u)$ is computed numerically by means of the midpoint rule.

For the computations we used the C++ library *Imaging*, developed by the Infmath Imaging group at the University of Innsbruck. We refer to [62] for a documentation of the source code.

4.7 Notes

In this chapter the iterative image denoising method (4.5.1) and the resulting evolution equation (4.31) as they were introduced by Osher et al. in [103] and Burger et al. in [33, 31] respectively, have been studied.

We pursued the following strategy: In Section 4.2 we reviewed a pairing technique for bounded, vector valued functions and vector valued Radon-measures due to the work of Anzellotti in [12]. We used this pairing concept in Section 4.3 in order to define the *g-norm*, originally introduced in the inspiring work by Meyer in [95] (for the case $\Omega = \mathbb{R}^2$).

Aujol & Aubert modified Meyer's definition and came up with a similar model for bounded domains $\Omega \subset \mathbb{R}^2$ in [13]. Our definition (Definition 4.3.1) holds in arbitrary space dimensions $N \geq 2$ and each $\mathrm{L}^p(\Omega)$ for $1 < p < \infty$ and reduces to the version introduced in [13] for $N = 2$ and $p = 2$. We remark, that the g-norm as studied in this chapter already appeared in the textbook [11] on the total variation flow equation by Andreu-Vaillo et al., although it is not explicitly used as a norm.

With these preparations we investigated duality relations of the g-norm with the total variation seminorm in Section 4.4, the Cauchy – Schwartz type inequality (4.21) being the fundamental estimate. It is the basis for a convenient characterization of the subgradient and the Legendre – Fenchel conjugate for the total variation seminorm (Theorem (4.4.2)). Moreover, it gives rise to the definition of *calibrable functions* (cf. Definition 4.3.9), established by Meyer in [95], then called *simple functions*. We prefer the first notion according to the work of Bellettini et al. [20, 21] (see also Alter et al. [6, 7]) who termed a set $C \subset \mathbb{R}^2$ calibrable if its characteristic function is calibrable according to our definition.

The analysis in Section 4.4 culminates in the assertion that the subgradient of the total variation seminorm (when viewed on $\mathrm{L}^p(\Omega)$) is weakly-weakly closed in the sense of Definition 3.1.2 (cf. Theorem 4.4.3). In view of Assumption 3.1.4, this paves the way to the analysis of the inverse total variation flow equation (4.31) in Section 4.5. Additionally to the general results proved in the first part of the thesis, we showed that Algorithm 4.5.1 (as well as the evolution equation (4.31)) satisfies a maximum principle (cf. Theorem 4.5.6).

Finally, we studied multiscale properties of the inverse total variation flow equation in Section 4.6. We found out, that for all (normalized) functions with finite g-norm, the solutions of Equation (4.31) do not change for small times (cf. Theorem 4.6.1) and that a signal can be reconstructed in finite time, if and only if it is calibrable (cf. Theorem 4.6.3).

Summary

In this thesis we studied iterative methods of augmented Lagrangian type for solving linear and ill-posed operator equations on Banach spaces. Furthermore, we have shown that this regularization technique — when viewed as an implicit time scheme — approximates solutions of a class of evolution equations. We applied the derived results to the image denoising problem. It turned out that the augmented Lagrangian algorithm in the context of image denoising constitutes an inverse scale space.

In Chapter 2 we introduced the augmented Lagrangian Algorithm 2.2.9 in a general Banach space setting allowing for nonsmooth regularization functionals. We proved that this algorithm constitutes a regularization method for the ill-posed operator equation (2.4). Improved results (including convergence rates) were derived additionally assuming Hilbert space data. In order to prove the regularization results we showed that the dual sequence generated by Algorithm 2.2.9 is characterized by the proximal point algorithm (Algorithm 2.2.16).

Chapter 3 was devoted to a class of evolution equations and their relation to the augmented Lagrangian algorithm. The equations under consideration (cf. (3.4)) consist of a system of coupled abstract differential inclusions depending on two time depending variables. Existence of solutions was shown by construction: Interpolations of the sequences generated by the augmented Lagrangian algorithm were considered, where the sampling points are determined by the sequence of regularization parameters in Algorithm 2.2.9.

Using the dual representation derived in Chapter 2 we first showed convergence of the (piecewise affine) interpolations of the dual sequences to a continuous time dependent (dual) function as the density of the sampling points increases. It turned out that this function already solves a gradient flow equation. This results was established by making extensive use of the analysis in Ambrosio et al. [9]. With this function at hand, we also proved convergence of the primal sequences to a Bregman-continuous function that together with the previously constructed dual function constitutes a solution of the evolution equation (3.4).

We considered special situations (e.g. additional smoothness for the Banach spaces in use, strict (total) convex regularization functionals etc.) in order to come up with stronger convergence results and improved smoothness assertions for solutions. Moreover, we studied regularizing properties of solutions of (3.4) w.r.t. the operator equation (2.4) where we again placed special emphasis on data in Hilbert spaces.

We applied the general results derived in Chapter 2 and Chapter 3 to the image denoising problem in Chapter 4. In this special case the augmented Lagrangian method (and the related evolution equation) reduce to a image denoising technique proposed by Osher et al. in [103] (and by Burger et al. in in [31, 33]), which uses the total variation seminorm for denoising. This was the actual starting point of this work.

We proved that the general assumptions in Chapter 2 and Chapter 3 are met and concluded that the iteration process is well defined and that solutions of the related evolution equation (*inverse total variation flow equation*) exist. Additional to the results in the first part of the thesis we proved a maximum principle for the inverse total variation flow and studied multiscale properties of solutions.

We finally mention that aside to the pure theoretical challenge the consideration of contin-

uous regularization techniques is justified from an application point of view as well. Whereas the augmented Lagrangian method introduced in Chapter 2 constituted a first order and thus a rather slow method for solving operator equations, the continuous formulation (3.4) in principle converges arbitrarily fast. Proving existence of solutions of (3.4) therefore paves the way for novel regularization techniques by finding alternative numerical schemes for (3.4).

Moreover, we showed in Chapter 4 that for a certain class of input images exact solutions of (3.4) can be computed. It turned out for the inverse total variation flow that – in contrast to classical scale spaces – the solution trajectories in general are *discontinuous* in time. From a scale space point of view this means that the different scales in an image are not selected smoothly by the scale parameter but rather in a sudden and abrupt way. Although inverse scale space methods have recently become very popular, this fact has not been discussed by the image processing community so far.

Nevertheless the inverse scale space method is a promising approach for further image processing tasks as for example *image decomposition*. Current research copes with the application of the augmented Lagrangian method and the associated evolution equations on nonsmooth decomposition methods such as *Meyer's problem* (cf. [95]).

A Mathematical Preliminaries

A.1 Functional Analytic Facts

In this section basic analytic notions and theorems are collected. In what follows we shall assume that X is a Banach space and we use the notation

$$\langle x^*, x\rangle := \langle x^*, x\rangle_{X^*,X} := x^*(x)$$

for the pairing of X^* and X. We write X^{**} for its *bidual* space, i.e. $X^{**} = (X^*)^*$. For $x \in X$ the *natural mapping* $i_X : X \to X^{**}$ is defined by

$$i_X(x)(x^*) = \langle x^*, x\rangle \, .$$

The canonical mapping i is a linear isometry, that is $\|i_X(x)\|_{X^{**}} = \|x\|_X$ for all $x \in X$ and surjective if and only if X is reflexive.

A.1.1 Weight Functions and Duality Mappings

Definition A.1.1. 1. A continuous and increasing function $\phi : [0,\infty) \to [0,\infty)$ is called a *weight function*, if $\phi(0) = 0$ and

$$\lim_{t\to\infty} \phi(t) = \infty.$$

2. Let ϕ be a weight function. The multivalued function $\mathfrak{J}_\phi : X \to \mathfrak{P}(X^*)$ defined by

$$\mathfrak{J}_\phi(x) = \{x^* \in X^* \;:\; \langle x^*, x\rangle = \|x^*\|_{X^*} \|x\|_X \, , \, \|x^*\|_{X^*} = \phi(\|x\|_X)\}$$

is called *duality mapping of weight* ϕ.

Remark A.1.2. Let ϕ be a weight function and $x \in X$. For $y = x\phi(\|x\|_X)$ there exists, according to the Hahn–Banach Theorem, an element $y^* \in X^*$ such that

$$\|y^*\|_{X^*} = 1 \quad \text{and} \quad y^*(y) = \|y\|_X \, .$$

Then $x^* := y^*\phi(\|x\|_X) \in \mathfrak{J}_\phi(x)$. Therefore, $\mathfrak{J}_\phi(x) \neq \emptyset$ for all $x \in X$. In particular, one has $\mathfrak{J}_\phi(0) = \{0\}$ and $\mathfrak{J}_\phi(x) = -\mathfrak{J}_\phi(-x)$.

For a given weight function ϕ we define *the primitive function w.r.t. to* ϕ, $\psi_\phi : [0,\infty) \to [0,\infty)$ as

$$\psi_\phi(s) = \int_0^s \phi(\sigma)\,\mathrm{d}\sigma$$

Obviously ψ_ϕ is convex and continuously differentiable. Moreover, ϕ^{-1} is well defined and also a weight function. Figure A.1 depicts a possible weight function ϕ and the corresponding primitives ψ_ϕ and $\psi_{\phi^{-1}}$.

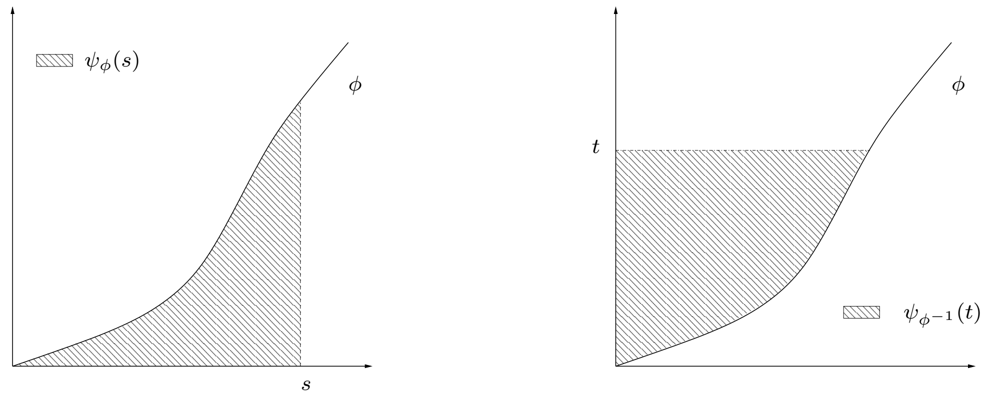

Figure A.1: Weight functions and their primitives. Left: weight function ϕ with ψ_ϕ. Right: weight function ϕ^{-1} (view from the ordinate axis) with $\psi_{\phi^{-1}}$.

Example A.1.3. For $p > 0$ and let p^* be its conjugate exponent, i.e. $1/p + 1/p^* = 1$. We define $\phi(s) = s^{p-1}$. Then ϕ is a weight function and

$$\psi_\phi(s) = \int_0^s \sigma^{p-1}\,\mathrm{d}\sigma = \frac{1}{p}s^p, \quad \phi^{-1}(s) = s^{p^*-1} \quad \text{and} \quad \psi_{\phi^{-1}}(s) = \frac{1}{p^*}s^{p^*}.$$

Now, let $N \geq 2$ and $\Omega \subset \mathbb{R}^N$ be open, bounded and suppose that $\partial\Omega$ is local Lipschitz and set $X = \mathrm{L}^p(\Omega)$. Then, according to [42, Thm. 4.8], one has

$$\mathfrak{J}_\phi(u) = |u|^{p-1}\,\mathrm{sign}(u).$$

The follow-up lemma collects some assertions on monotonicity and asymptotic behavior of weight functions and their primitives.

Lemma A.1.4. Let ϕ be a weight function.

1. The mapping $\varphi_1(s) = s^{-1}\psi_\phi(s)$ is nondecreasing and
$$\lim_{s\to\infty} \varphi_1(s) = \infty.$$
2. The mapping $\varphi_2(s) = s^{-1}\psi_\phi^{-1}(s)$ is nonincreasing and
$$\lim_{s\to\infty} \varphi_2(s) = 0.$$
3. The mapping $\varphi_3(s) = s\psi_\phi^{-1}(s^{-1})$ is nondecreasing.

Proof. (1). First note that

$$\varphi_1(s) = s^{-1}\int_0^s \phi(\sigma)\,\mathrm{d}\sigma = \int_0^1 \phi(s\sigma)\,\mathrm{d}\sigma.$$

Hence $s \mapsto \varphi_1(s)$ is nondecreasing due to the monotonicity of ϕ.

For given $s > 0$ let $\xi_s \in [0, \infty)$ be such that

$$\psi_\phi(s) = \xi_s s.$$

Assume that there exists a constant $M > 0$ such that $\xi_s \leq M$ for all $s \in [0, \infty)$. This implies that

$$\int_0^s \phi(\tau)\, \mathrm{d}\tau \leq sM \tag{A.1}$$

Set $\tilde{\phi}(s) = \phi(s) - M$ and let $s_0 \in [0, \infty)$ be the unique element such that $\tilde{\phi}(s_0) = 0$. Then it follows from (A.1) that for $s \geq s_0$

$$0 \geq \int_0^s \phi(\tau)\, \mathrm{d}\tau - sM = \int_0^s \tilde{\phi}(\tau)\, \mathrm{d}\tau = \int_0^{s_0} \tilde{\phi}(\tau)\, \mathrm{d}\tau + \int_{s_0}^s \tilde{\phi}(\tau)\, \mathrm{d}\tau.$$

This amounts to

$$\int_{s_0}^s \tilde{\phi}(\tau)\, \mathrm{d}\tau \leq -\int_0^{s_0} \tilde{\phi}(\tau)\, \mathrm{d}\tau < \infty$$

for all $s \geq s_0$ which is clearly a contradiction, since $\tilde{\phi}(s) \to \infty$ for $s \to \infty$. Therefore $\lim_{s\to\infty} \xi_s = \infty$ and (1) follows.

(2). Let $\varepsilon > 0$. From (1) it becomes clear that for all $M > 0$ there exists $s_M > 0$ such that

$$\frac{\psi_\phi(s)}{s} \geq M$$

for all $s \geq s_M$. Choose M such that $M\varepsilon > 1$ and set $s_0 = M s_M$. Then for all $s > s_0$ we have that $sM^{-1} > s_M$ and we observe

$$\frac{\psi_\phi(sM^{-1})}{s} \geq 1 \;\Leftrightarrow\; \psi_\phi^{-1}(s) \leq sM^{-1} < s\varepsilon.$$

Thus $\varphi_2(s) \to 0$ as $s \to \infty$.

We now show $\varphi_2'(s) \leq 0$. From definition it is clear that $s = \psi_\phi(\varphi_2(s)s)$ for all $s > 0$ and hence

$$s = \int_0^{\varphi_2(s)s} \phi(\sigma)\, \mathrm{d}\sigma \leq \varphi_2(s) s \phi(\varphi_2(s)s)$$

or in other words

$$\frac{1}{\phi(\varphi_2(s)s)} \leq \varphi_2(s).$$

With this we get the estimate

$$\varphi_2'(s) = \frac{1}{s}\left(\frac{1}{\phi(\varphi_2(s)s)} - \varphi_2(s)\right) \leq 0.$$

(3). Define

$$t = t(s) = \psi_\phi^{-1}(s^{-1}).$$

Consequently we have $s(t) = (\psi_\phi(t))^{-1}$ and we further introduce

$$\tilde{\varphi}_3(t) := \varphi_3(s(t)) = \frac{t}{\psi_\phi(t)}.$$

Since

$$\tilde{\varphi}_3'(t) = \frac{\psi_\phi(t) - t\phi(t)}{\psi_\phi^2(t)} \leq 0, \text{ and } s'(t) = \frac{-\phi(t)}{\psi_\phi^2(t)} \leq 0$$

we conclude from

$$\tilde{\varphi}_3'(t) = \varphi_3'(s(t)) \cdot s'(t)$$

and from the surjectivity of $s(t) = (\psi_\phi(t))^{-1}$ on $[0, \infty)$ that $\varphi_3'(s) \geq 0$ for all $s \in [0, \infty)$. □

Lemma A.1.5. Let $\phi : [0, \infty) \to [0, \infty)$ be a weight function and let $\mathfrak{J}_\phi : X \to X^*$ be the duality mapping on X w.r.t to ϕ and $\mathfrak{J}_{\phi^{-1}} : X^* \to X^{**}$ the duality mapping on X^* w.r.t. ϕ^{-1}. Then we have that

$$\mathfrak{J}_{\phi^{-1}}\left(\mathfrak{J}_\phi(x)\right) = \{i_X(x)\}.$$

Proof. Let x^{**} be an element of $\mathfrak{J}_{\phi^{-1}}(x^*)$, where in turn $x^* \in \mathfrak{J}_\phi(x)$ for $x \in X$. From the definition of the duality mapping it follows that $\|x^{**}\| = \phi^{-1}(\|x^*\|) = \phi^{-1}(\phi(\|x\|)) = \|x\|$. Thus we find that

$$\langle x^{**}, x^* \rangle_{X^{**}, X^*} = \|x^{**}\| \, \|x^*\| = \|x\| \, \|x^*\| = x^*(x)$$

or in other words that $x^{**} = i_X(x)$. □

A.1.2 Integration

In this paragraph we review the notion of Bochner-integrable functions and some basic properties. In particular, the connection between the Bochner-integral and derivatives of absolutely continuous functions is established. For the presented definitions and further details we refer the reader to the books by Diestel & Uhl [47, Chap. II] or Dunford & Schwartz [49, Chap. III].

For the sake of simplicity we restrict our considerations on functions defined on an open subset Ω of the euclidean space $\mathbb{R}^N$ with values in the Banach space X. Let $\mathcal{B}(\mathbb{R}^N)$ denote the Borel σ-algebra generated by the open sets of $\mathbb{R}^N$ and λ^N the N-dimensional Lebesgue measure.

Definition A.1.6.

1. Consider $f : \Omega \to X$ such that $\operatorname{ran}(f) = \{x_1, \ldots, x_n\} \subset X$ and

   $$f^{-1}(x_i) = \{x \in X \; : \; f(x) = x_i\} \in \mathcal{B}(\mathbb{R}^N).$$

   A function $g : \Omega \to X$ is called *simple*, whenever it equals such an f λ^N-a.e. in Ω.

2. A function $f : \Omega \to X$ is called *measurable*, if there exists a sequence of simple functions $\{f_n\}_{n \in \mathbb{N}}$ such that

   $$\lim_{n \to \infty} \lambda^n \left(\{s \in \Omega \; : \; \|f_n(s) - f(s)\| \geq \varepsilon\}\right) = 0 \tag{A.2}$$

   for all $\varepsilon > 0$. The function f is called *weakly-measurable*, if for all $x^* \in X^*$ the numerical function $s \mapsto \langle x^*, f(s) \rangle$ is measurable.

3. A simple function g is called *integrable*, if it is λ^N-a.e. equal to a function f of the form

   $$f = \sum_{j=1}^{n} x_j \chi_{E_j}$$

where $\{E_j\}_{1\leq j\leq n}$ is a partition of $\mathbb{R}^N$ and $x_j = 0$ if $\lambda^n(E_j) = \infty$. We define *the integral of g over Ω* as

$$\int_\Omega f(s)\,\mathrm{d}\lambda^N(s) := \int_\Omega f\,\mathrm{d}\lambda^N := \sum_{j=1}^n x_j \lambda^N(\Omega\cap E_j)$$

where $0\cdot\infty := 0$.

4. A measurable function $g : \Omega \to X$ is called *integrable*, if there exists a sequence of simple functions $\{f_n\}_{n\in\mathbb{N}}$ converging to f in the sense of (A.2) such that

$$\lim_{m,n\to\infty} \int_\Omega \|f_n(s) - f_m(s)\|\,\mathrm{d}\lambda^N(s) = 0. \tag{A.3}$$

Proposition A.1.7. *1. Let $f : \Omega \to X$ be integrable and $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of simple functions satisfying* (A.2) *and* (A.3). *Then the limit*

$$\int_\Omega f(s)\,d\lambda^N(s) := \int_\Omega f\,d\lambda^N := \lim_{n\to\infty}\int_\Omega f_n\,d\lambda^N$$

exists and does not depend on the choice of $\{f_n\}_{n\in\mathbb{N}}$. We call $\int_\Omega f(s)\,d\lambda^N(s)$ the integral of f over Ω.

2. A function $f : \Omega \to X$ is integrable if and only if it is measurable and

$$\int_\Omega \|f(s)\|\,d\lambda^n(s) < \infty$$

holds.

Proof. [49, Chap.III. Lem.16] and [49, Chap.III. Lem.18] □

Definition A.1.8. For $1 \leq p < \infty$ we define $\mathrm{L}^p(\Omega, X)$ as the space of all (equivalence classes of) measurable functions $f : \Omega \to X$ such that

$$\|f\|_{\mathrm{L}^p(\Omega,X)} := \left(\int_\Omega \|f\|^p\,\mathrm{d}\lambda^N\right)^{\frac{1}{p}} < \infty.$$

For $p = \infty$ the symbol $\mathrm{L}^\infty(\Omega, X)$ stands for the space of all (equivalence classes of) measurable functions $f : \Omega \to X$ that are essentially bounded on Ω, that is f satisfies

$$\|f\|_{\mathrm{L}^\infty(\Omega,X)} := \operatorname*{ess\,sup}_{s\in\Omega}\|f(s)\| := \inf_{\lambda^N(A)=0}\ \sup_{s\in\Omega\setminus A}\|f(s)\| < \infty.$$

The spaces $\mathrm{L}^p(0, T; X)$ are called *Lebesgue – Bochner spaces*.

It can be shown that for $1 \leq p \leq \infty$ the functions $\|\cdot\|_{\mathrm{L}^p(\Omega,X)}$ render the spaces $\mathrm{L}^p(\Omega, X)$ normed vector spaces, which are complete whenever X is complete (see [49, Chap.III Thm.5] and [47, Chap. IV]). Moreover we have

Proposition A.1.9. *If $1 \leq p < \infty$, the set of integrable and simple functions is dense in $L^p(\Omega, X)$ w.r.t. to the topology induced by $\|\cdot\|_{L^p(\Omega,X)}$.*

Proof. [49, Chap.III Cor.8] □

From this result we draw the following conclusion

Corollary A.1.10. Let $1 \leq p < \infty$. The space $\mathrm{L}^p(\Omega, X)$ is separable, whenever X is separable.

Proof. Let $\Sigma \subset \mathcal{B}(\mathbb{R}^N)$ be the collection of all sets Q satisfying

$$Q = \prod_{j=1}^{N} (a_j, b_j), \quad \text{for} \quad a_j, b_j \in \mathbb{Q} \quad \text{and} \quad Q \subset \Omega.$$

Moreover, assume that $D \subset X$ is a countable and dense subset. In view of Proposition A.1.9 it is sufficient to prove that the set

$$S := \left\{ f : \Omega \to D \ : \ f \text{ is simple, integrable and } f^{-1}(X) \subset \Sigma \right\}.$$

lies dense (w.r.t. $\|\cdot\|_{\mathrm{L}^p(\Omega,X)}$) in the set of all simple and integrable functions on Ω (note, that S is countable).

To this end, let $f : \Omega \to X$ be simple. Without loss of generality we can assume that

$$f(s) = \begin{cases} x_0 & \text{if } s \in A, \\ 0 & \text{else.} \end{cases}$$

for a set $A \in \mathcal{B}(\mathbb{R}^N)$ and $x_0 \in X$. We first show, that for each $\varepsilon > 0$ there exists $M(\varepsilon) \in \mathbb{N}$, a finite family of disjoint sets $\{Q_m^\varepsilon\}_{1 \leq m \leq M(\varepsilon)} \subset \Sigma$ and a λ^N-measurable set N^ε such that

$$\bigcup_{m=1}^{M(\varepsilon)} Q_m^\varepsilon = A + N^\varepsilon \quad \text{and} \quad \lambda^N(N^\varepsilon) \leq \varepsilon. \tag{A.4}$$

Indeed, since λ^N is a Radon-measure one has (cf. [55, Chap. 1.1 Thm. 4])

$$\lambda^N(A) = \inf \left\{ \lambda^N(O) \ : \ A \subset O, \ O \text{ is open} \right\}.$$

For a given $\varepsilon > 0$, there hence exists an open $O^\varepsilon \supset A$ such that $\lambda^N(A) \geq \lambda(O^\varepsilon) - \varepsilon/2$. This shows

$$\lambda^N(O^\varepsilon \backslash A) = \lambda^N(O^\varepsilon) - \lambda^N(A) \leq \frac{\varepsilon}{2}.$$

Moreover, since O^ε is open, it can be written as countable union of disjoint sets $Q_m^\varepsilon \in \Sigma$ $(m \in \mathbb{N})$ and thus $M(\varepsilon) \in \mathbb{N}$ can be chosen such that

$$\sum_{m=M(\varepsilon)}^{\infty} \lambda^N(Q_m) \leq \frac{\varepsilon}{2}.$$

Therefore (A.4) follows with

$$N^\varepsilon = \bigcup_{m=M(\varepsilon)}^{\infty} Q_m^\varepsilon \cup (O^\varepsilon \backslash A).$$

Now let $x^\varepsilon \in D$ be such that $\|x_0 - x^\varepsilon\| \leq \varepsilon$ and set

$$f^\varepsilon(s) = \begin{cases} x^\varepsilon & \text{if } s \in A^\varepsilon, \\ 0 & \text{else,} \end{cases}$$

where $A^\varepsilon := \bigcup_{m=1}^{M(\varepsilon)} Q_m^\varepsilon$. We have that $f^\varepsilon \in S$ and furthermore, it follows that

$$\int_\Omega \|f_\varepsilon - f\|^p \, \mathrm{d}\lambda^N = \int_{N^\varepsilon} \|x_\varepsilon\|^p \, \mathrm{d}\lambda^N + \int_{A^\varepsilon} \|x^\varepsilon - x_0\|^p \, \mathrm{d}\lambda^N \leq \varepsilon \|x_\varepsilon\|^p + \lambda^N(A^\varepsilon)\varepsilon^p$$

In other words, this means

$$\lim_{\varepsilon \to 0^+} \|f^\varepsilon - f\|_{\mathrm{L}^p(\Omega,X)} = 0$$

and the assertion follows. □

Theorem A.1.11. *Let $1 \leq p < \infty$ and q be such that $p^{-1} + q^{-1} = 1$ (agreeing upon $q = \infty$ for $p = 1$). If X is reflexive we have that*

$$(L^p(\Omega, X))^* = L^q(\Omega, X^*). \tag{A.5}$$

Proof. [47, Chap.IV Thm.1]. □

Corollary A.1.12. If X is separable and reflexive, every norm-bounded subset of $\mathrm{L}^\infty(\Omega, X^*)$ is sequentially weakly* compact.

Proof. Since X is separable it follows from Corollary A.1.10 that $\mathrm{L}^1(\Omega; X)$ is also separable. Moreover, Theorem A.1.11 shows that

$$\mathrm{L}^\infty(\Omega; X^*) = \left(\mathrm{L}^1(\Omega; X)\right)^*.$$

In other words, $\mathrm{L}^\infty(\Omega, X^*)$ is the dual of a separable Banach space. Thus it follows from [94, Thm. 2.6.23] that for each bounded subset $C \subset \mathrm{L}^\infty(\Omega, X^*)$ the relative weak* topology on C is induced by a metric. Since by the Banach – Alaoglu Theorem [94, Thm. 2.6.18] C is already weakly* compact, the assertion follows from the fact that sequential compactness and compactness are equivalent on metric spaces. □

Corollary A.1.10 and Theorem A.1.11 are two example of a whole class of mathematical problems arising integration theory of vector valued functions:

> If X satisfies a certain property, when does this property lifts to the corresponding Lebesgue-Bochner space $\mathrm{L}^p(\Omega; X)$?

For separability and reflexivity positive answers are given by Corollary A.1.10 and Theorem A.1.11 respectively (see also [47, Chap. IV.1 Cor. 2] for the latter). Moreover, it was shown by Smith & Turett in [118] that this also stays true for various notions of convexity of the Banach space X. By the same authors it was pointed out though, that in general the Radon – Riesz property of a Banach space X is not transferred to $\mathrm{L}^p(0, T; X)$.

We recall that a Banach space X has the *Radon – Riesz property* if for all sequences $\{x_n\}_{n\in\mathbb{N}} \subset X$ the following implication holds

$$\left.\begin{aligned} \lim_{n\to\infty} \|x_n\| = \|x\| \\ w\text{-}\lim_{n\to\infty} x_n = x \end{aligned}\right\} \Rightarrow \lim_{n\to\infty} x_n = x. \tag{A.6}$$

Some authors prefer the notion *Kadeč – Klee property* or *property (H)* (cf. [42]).

In [92] Lin & Lin proved that $\mathrm{L}^p(0,T;X)$ inherits the Radon – Riesz property from X, if X is reflexive and has no subspace isomorphic to ℓ^1, which is for example satisfied if X is reflexive (see [94, Thm 1.13.8]). We summarize

Theorem A.1.13. *Let $T > 0$ and $1 < p < \infty$. If a reflexive and strictly convex Banach space X has the Radon – Riesz property, so does $L^p(0,T;X)$.*

This gives rise to the following

Definition A.1.14. A reflexive and strictly convex Banach space that has the Radon – Riesz property is called an *E-space.*

For the sake of completeness, we mention that E-spaces are sometimes also called *strongly convex* (see e.g. [94]) and were originally introduced by Fan and Glicksberg in [56] (with a different definition: cf. [94, Def. 5.3.15]). They turn out to be exactly those spaces, on which the norm attains a unique minimizer on every convex and closed subset, such that all minimizing sequences are strongly convergent. Therefore they play an important role in convex minimization. For further reading we refer to Holmes [79, Part V].

Lemma A.1.15. For every integrable function $f \in \mathrm{L}^1(\Omega, X)$ and $x^* \in X^*$ one has

$$\left\langle x^*, \int_\Omega f \,\mathrm{d}\lambda^N \right\rangle = \int_\Omega \langle x^*, f(s)\rangle \,\mathrm{d}\lambda^N(s).$$

Proof. [47, Chap. II Thm. 6] □

We close this paragraph with the fundamental theorem of calculus for Bochner-integrable functions. For further reading we refer e.g. to the textbook by Miyadera [99, Chap. 1.3].

Theorem A.1.16. *Let X be reflexive and $-\infty < a \leq b < \infty$. For a function $f : [a,b] \to X$ the following two assertions are equivalent:*

1. *The function f is absolutely continuous on $[a,b]$.*
2. *The (strong) derivative df of f exists λ^1-a.e. in $[a,b]$ and is Bochner-integrable. Moreover, one has*
$$f(s) = f(a) + \int_a^s df(\tau)\, d\tau, \qquad \text{for all } a \leq s \leq b.$$

Proof. [99, Chap. 1.3 Thm. 1.17] □

Theorem A.1.16 gives rise to the following

Definition A.1.17. Let X be reflexive and $-\infty < a \leq b < \infty$. For $1 \leq p \leq \infty$ we define

$$\mathrm{W}^{1,p}(a,b;X) = \{f : [a,b] \to X \;:\; f, df \in \mathrm{L}^p(0,T;X)\}\,.$$

The spaces $\mathrm{W}^{1,p}(0,T;X)$ are Banach spaces when equipped with the norm

$$\|f\|_{\mathrm{W}^{1,p}(0,T;X)} := \|f\|_{\mathrm{L}^p(0,T;X)} + \|df\|_{\mathrm{L}^p(0,T;X)}$$

and are called *Sobolev – Bochner spaces.*

Remark A.1.18. Let X be reflexive and $-\infty < a \leq b < \infty$. From Theorem A.1.16 and Definition A.1.17 it is evident that

$$\mathrm{W}^{1,1}(a,b;X) = \{f : [a,b] \to X \;:\; f \text{ is absolutely continuous on } [a,b]\}\,.$$

A.1.3 Miscellaneous

In this section we collect some fundamental theorems situated in the orbit of functional analysis, which may lie beyond the scope of a standard calculus lecture.

Theorem A.1.19 (Arzelà – Ascoli). *Let X be reflexive and $-\infty < a \le b < \infty$. Assume that $K \subset X$ is bounded and that $\{x_n : [a,b] \to X\}_{n\in\mathbb{N}}$ is a family of functions satisfying*

1. *the values $x_n(t) \in K$ for all $t \in [a,b]$ and $n \in \mathbb{N}$ and*
2. *for all $s,t \in [a,b]$ that*
$$\limsup_{n\to\infty} \|x_n(s) - x_n(t)\| \le \omega(s,t),$$

for $\omega : [a,b]^2 \to [0,\infty)$ satisfying

$$\lim_{(s,t)\to(r,r)} \omega(s,t) = 0, \qquad \text{in } [a,b]\backslash N,$$

where $\lambda^1(N) = 0$. Then there exists a function $x : [a,b] \to X$ that is absolutely continuous in $[a,b]\backslash N$ such that

$$x_n(t) \rightharpoonup x(t), \qquad \text{for all } t \in [a,b]\backslash N.$$

Proof. [9, Prop.3.3.1] □

Theorem A.1.20 (Helly). *Let $-\infty < a \le b < \infty$ and $f_n : [a,b] \to [-\infty,\infty]$ be a sequence of nonincreasing functions. Then one can find a nonincreasing function $f : [a,b] \to [-\infty,\infty]$ and a selection $n \mapsto \rho(n)$ such that for all $t \in [a,b]$*

$$\lim_{n\to\infty} f_{\rho(n)}(t) = f(t).$$

Proof. [9, Lem. 3.3.3] □

Theorem A.1.21 (Vitali). *Assume that $(\Omega, \mathcal{S}, \mu)$ is a measure space and let $0 < p < \infty$. For $f, f_n \in L^p(\Omega, \mathbb{R}, \mu)$ ($n \in \mathbb{N}$) the following two statements are equivalent*

1. *The sequence $\{f_n\}_{n\in\mathbb{N}}$ converges strongly to f in L^p.*
2. a) *For all $\varepsilon > 0$ and $A \in \mathcal{S}$ satisfying $\mu(A) < \infty$ one has*
$$\lim_{n\to\infty} \mu(\{x \in \Omega \ : \ |f(x) - f_n(x)| \ge \varepsilon\}) = 0.$$
   b) *For all $\varepsilon > 0$ there exists $E \in \mathcal{S}$, such that $\mu(E) < \infty$ and*
$$\int_{\Omega\backslash E} |f_n|^p \; d\mu < \varepsilon, \qquad \text{for all } n \in \mathbb{N}.$$
   c) *For all $\varepsilon > 0$ there exists $\delta > 0$, such that for all $A \in \mathcal{S}$ satisfying $\mu(A) < \delta$ and all $n \in \mathbb{N}$ one has*
$$\int_A |f_n|^p \; d\mu < \varepsilon.$$

Proof. [52, Chap. VI Thm.5.6] □

From Vitali's theorem A.1.21 we derive the following

Lemma A.1.22. Let $(\Omega, \mathcal{S}, \mu)$ be a measure space, such that $\mu(\Omega) < \infty$. Let $0 < r < p \leq \infty$, $f : \Omega \to \mathbb{R}$ a μ-measurable function and $\{f_n\}_{n\in\mathbb{N}} \subset \mathrm{L}^p(\Omega, \mathbb{R}, \mu)$ a L^p-bounded sequence. Then the following two statements are equivalent

1. For all $\varepsilon > 0$
$$\lim_{n\to\infty} \mu\left(\{x \in \Omega \ : \ |f_n(x) - f(x)| > \varepsilon\}\right) = 0.$$
2. $f \in \mathrm{L}^p(\Omega, \mathbb{R}, \mu)$ and
$$\lim_{n\to\infty} \|f_n - f\|_{\mathrm{L}^r(\Omega,\mathbb{R},\mu)} = 0.$$

Proof. We define $M := \sup_{n\in\mathbb{N}} \|f_n\|_{\mathrm{L}^p(\Omega,\mathbb{R},\mu)}$. (1) $\Rightarrow$ (2). Set $k_0 = 1$. For every $n \in \mathbb{N}$ choose $k_n > k_{n-1}$ such that

$$\mu\left(\left\{x \in \Omega \ : \ |f(x) - f_m(x)| > \frac{1}{n}\right\}\right) < \frac{1}{n^2}.$$

for all $m > k_n$. Define $A_n := \left\{x \in \Omega \ : \ |f_{k_n}(x) - f(x)| > \frac{1}{n}\right\}$ and set for $k > 0$ $B_k := \bigcup_{l=k}^{\infty} A_l$. Now let $\delta > 0$ and $n_0 \in \mathbb{N}$ such that $\sum_{j=n}^{\infty} \frac{1}{l^2} \leq \delta$ for all $n \geq n_0$. This implies that

$$\mu(B_n) = \mu\left(\bigcup_{l=n}^{\infty} A_l\right) \leq \sum_{l=n}^{\infty} \mu(A_l) \leq \sum_{l=n}^{\infty} \frac{1}{l^2} \leq \delta.$$

and moreover, for all $x \in \Omega \backslash B_n$ it follows from construction that

$$|f(x) - f_{k_n}(x)| < \frac{1}{n}.$$

This means, that with the subsequence $n \mapsto \rho(n) = k_n$ one has, $f_{\rho(n)} \to f$ (uniformly) a.e. in Ω. Fatou's Lemma now yields

$$\int_\Omega |f|^p \, \mathrm{d}\mu = \int_\Omega \liminf_{n\to\infty} \left|f_{\rho(n)}\right|^p \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega \left|f_{\rho(n)}\right|^p \, \mathrm{d}\mu \leq M^p.$$

This shows $f \in \mathrm{L}^p(\Omega, \mathbb{R}, \mu)$. Moreover, observe from Hölder's inequality that for all $A \in \mathcal{S}$

$$\int_A |f_n|^r \, \mathrm{d}\mu \leq \mu(A)^{\frac{p-r}{p}} M^p.$$

Thus the requirements of Vitali's convergence theorem (cf. Theorem A.1.21) are satisfied and (2) follows.

(2) $\Rightarrow$ (1). Observe that for all $\varepsilon > 0$

$$\mu\left(\{x \in \Omega \ : \ |f_n(x) - f(x)| > \varepsilon\}\right) \leq \int_\Omega \left|(f_n - f)\varepsilon^{-1}\right|^r \, \mathrm{d}\mu = \varepsilon^{-r} \|f_n - f\|^r_{\mathrm{L}^r(\Omega,\mathbb{R},\mu)}$$

and (2) clearly implies (1). $\square$

A.2 Convex Analysis

In this section we review techniques from convex analysis used in the thesis, basically concerning subdifferential calculus, the Bregman distance and the Legendre – Fenchel duality concept. Standard textbooks dealing with these topics are Ekeland & Temam [51], Aubin [14], Ioffe & Tichomirov [81] or Barbu & Precupanu [18]. Throughout this section we assume that X is Banach space, unless stated differently.

A.2.1 Subdifferentiability and Slopes

Definition A.2.1. Let $J : X \to \overline{\mathbb{R}}$ be convex.

1. J is called *subdifferentiable at* $x_0 \in X$ if there exists an element $\xi^* \in X^*$ such that
$$J(x) \geq J(x_0) + \langle \xi^*, x - x_0 \rangle \, .$$
We call ξ^* a *subgradient* of J at x_0 and the collection of all subgradients $\partial J(x_0) \subset X^*$ the *subdifferential* of J at x_0. The set
$$D(\partial J) = \{x \in X \ : \ \partial J(x) \neq \emptyset\}$$
is called the *effective domain* of the subgradient ∂J.

2. Each function $\partial^0 J : D(\partial J) \subset X \to X^*$ satisfying
$$\partial^0 J(x) \in \operatorname{argmin} \{\|p^*\|_* \ : \ p^* \in \partial J(x)\}$$
is called a *minimal section* of ∂J.

3. Let $\partial^0 J$ be a minimal section of J. The function $|\partial J| : X \to \overline{\mathbb{R}}$ defined by
$$|\partial J|\,(x) = \begin{cases} \left\|\partial^0 J(x)\right\|_* & \text{if } x \in D(\partial J) \\ +\infty & \text{else} \end{cases}$$
is called the *slope* of J.

Lemma A.2.2. Let $J : X \to \overline{\mathbb{R}}$ be convex and lower semicontinuous.

1. The set $\partial J \subset X \times X^*$ is strongly-weakly* closed and consequently there exists at least one minimal section of ∂J.

2. For all $x \in X$ we have that
$$|\partial J|\,(x) = \sup_{y \neq x} \frac{[J(x) - J(y)]^+}{\|x - y\|}.$$

3. Let $T > 0$. For all absolutely continuous and a.e. differentiable functions $x : [0, T] \to X$ we have that
$$|J(x(s)) - J(x(t))| \leq \int_s^t |\partial J|\,(x(\tau))\,\left|x'(\tau)\right|\,\mathrm{d}\tau, \qquad \text{for all } 0 \leq s \leq t \leq T.$$

Proof. [9, Prop.1.4.4] and additionally [9, Thm.1.2.5] for (3) □

We proceed with a sufficient condition for sequentially weak lower semicontinuity of the slope of a convex functional.

Lemma A.2.3. Let $J : X \to \overline{\mathbb{R}}$ be convex and lower semicontinuous and assume that ∂J is weakly-weakly* closed in the sense of Definition 3.1.2. Then the slope $|\partial J|$ is sequentially weakly lower semicontinuous.

Proof. [9, Chap. 2 Lem. 2.3.6] □

Proposition A.2.4 (Asplund)**.** *Let ϕ be a weight function. Then*

$$\mathfrak{J}_\phi = \partial\left(\psi_\phi(\|\cdot\|)\right)$$

Proof. [42, Chap. 1 Thm. 4.4]. □

Lemma A.2.5. Let $J : Y \to \overline{\mathbb{R}}$ be convex and $K : X \to Y$ linear. Then for all $x \in X$

$$K^*(\partial J(K(x)) \subset \partial(J \circ K)(x).$$

If additionally J is continuous in at least on point of $\mathrm{ran}(K)$, then equality holds.

Proof. [81, Chap. 4.2 Thm. 2] □

A.2.2 Bregman Distance and Total Convexity

Definition A.2.6. Let $J : X \to \overline{\mathbb{R}}$ be convex and $\xi^* \in X^*$ and define $D_J^{\xi^*} : D(J) \times D(\partial J) \to \mathbb{R}$ by

$$D_J^{\xi^*}(u, v) = J(u) - J(v) - \langle \xi^*, u - v\rangle .$$

If $u \in D(J)$, $v \in D(\partial J)$ and $\xi^* \in \partial J(v)$, then the quantity $D_J^{\xi^*}(u, v)$ is called *Bregman Distance of u and v with respect to J and ξ^*.*

It is clear that the Bregman distance is not a metric, since it is in general neither symmetric nor satisfies the triangle inequality. However, from the definition it immediately becomes clear, that for given $v \in D(\partial J)$ and $\xi^* \in \partial J(v)$ we have that $D_J^{\xi^*}(u, v) \geq 0$, where $u = v$ implies equality. In the literature often a variant of the Bregman distance is considered: Define $D_J : D(J) \times D(J) \to \overline{\mathbb{R}}$ as

$$D_J(u, v) = J(u) - J(v) - J'(v)(u - v),$$

where $J'(u)(w)$ denotes the directional derivative of J at u in direction w, i.e.

$$J'(u)(w) = \lim_{h \to 0^+} \left(\frac{1}{h}\left(J(u + hw) - J(u)\right)\right).$$

The directional derivative of a convex function is everywhere defined in $D(J)$ (with values in $[-\infty, \infty]$) and we have that

$$0 \leq D_J(u, v) \leq D_J^{\xi^*}(u, v)$$

for all $(u, v) \in D(J) \times D(\partial J)$ and $\xi^* \in \partial J(v)$.

The Bregman distance has proven to be a valuable tool to study convergence (and convergence rates) of various regularization problems (Hofmann et al. [78], Resmerita [110], Resmerita et al. [111], Burger et al. [32]). The success of this technique is due to the fact, that the Bregman distance automatically provides the suitable topology subordinate to the given problem.

To be more precise: Let $J : X \to \overline{\mathbb{R}}$ be a convex functional and $x_0 \in X$. For every $\varepsilon > 0$ we define

$$B_J(x_0, \varepsilon) := \{x \in X \ : \ D_J(x, x_0) < \varepsilon\} \tag{A.7}$$

where we agree upon $D_J(x, x_0) = \infty$ for $x_0 \notin D(J)$ and set

$$\tau_X^J = \{U \subset X \ : \ \forall\, x \in U\, \exists\, \varepsilon > 0 \text{ such that } B_J(x, \varepsilon) \subset U\}\,. \tag{A.8}$$

It is quite obvious that the system τ_X^J forms a topology on X. This gives rise to the following

Definition A.2.7. The topology τ_X^J on X is called *Bregman topology with respect to J.*

Remark A.2.8.

1. For an arbitrary convex functional $J : X \to \overline{\mathbb{R}}$ the topology τ_X^J is in general not Hausdorff, that is, the set of points in X that can not be separated by ε-balls of type (A.7) is not empty. Indeed, any to points $x, y \in X$ satisfying

$$J'(x)(y - x) = -J'(y)(x - y)$$

can not be separated by τ_X^J-neighborhoods.

2. From Definition A.2.7 it becomes clear that for $x \notin D(J)$, the sets $\{x\}$ are open in τ_X^J, that is, the induced topology of τ_X^J on $X \backslash D(J)$ is discrete. In the sequel we will therefore consider the induced topology of τ_X^J on $D(J)$ rather than on the whole space X, but will use the same notation.

3. Let $\Omega \subset \mathbb{R}^N$. A function $x : \Omega \to X$ is τ_X^J-sequential continuous in $s_0 \in \Omega$ if

$$\lim_{s \to s_0} D_J(x(s)x(s_0)) = 0.$$

We write $C_J(\Omega, X)$ for the collection of all these functions.

It is of natural interest to characterize the collection of functionals J for which the Bregman topology on X is finer than the norm topology.

Definition A.2.9. Let $J : X \to \overline{\mathbb{R}}$ be convex and $y \in D(J)$. The function $\nu_J(y, \cdot) : [0, \infty) \to [0, \infty]$ defined by

$$\nu_J(y, t) = \inf\{D_J(x, y) \ : \ x \in D(J),\ \|x - y\| = t\}$$

is called *modulus of total convexity of J at y*. The functional J is called *totally convex at y* if $\nu_J(y, \cdot) > 0$ and *totally convex* if it is totally convex at every $y \in D(J)$.

The modulus of total convexity and totally convex functions were introduced by Butnariu et al. in [34] (under the name *modulus of local convexity* and *very convex functions*) and received remarkable attention since. From the definition it becomes clear, that every total convex function is already strict convex. Moreover we can prove the

Lemma A.2.10. Let $J : X \to \overline{\mathbb{R}}$ be convex. Then the following statements are equivalent

1. J is totally convex
2. The topology τ_X^J is finer than the norm topology on $D(J)$.

Proof. (cf. Resmerita [109]) (1) $\Rightarrow$ (2). Let $x \in D(J)$ and J be totally convex at x. Moreover, assume that $r > 0$ and let $B_r(x)$ be the ball with radius r centered at x in $D(J)$ w.r.t the norm on X. Then for $\varepsilon := \nu_J(x, \frac{r}{2})$ and all $z \in D(J)$ such that $\|z - x\| = \frac{r}{2}$ we have that $D_J(z, x) \geq \varepsilon$. Thus the implication

$$\forall\, z \in D(J) : D_J(z, x) < \varepsilon \;\Rightarrow\; \|z - x\| \leq \frac{r}{2} < r$$

holds and thus $B_\varepsilon^J(x) \subset B_r(x)$, which shows that τ_X^J is finer than the topology induced by the norm.

(2) $\Rightarrow$ (1). Assume that J is not totally convex at $x \in D(J)$. Then there exists a $t_0 > 0$ such that $\nu_J(x, t_0) = 0$. In other words, this means that for all $\varepsilon > 0$ there exists an element $z_\varepsilon \in D(J)$ such that $\|z_\varepsilon - x\| = t_0$ and $z_\varepsilon \in B_\varepsilon^J(x)$. Since by assumption τ_X^J has more open balls than the norm topology, we can find for $r = \frac{t_0}{2}$ a number $\varepsilon_0 > 0$ such that

$$z_{\varepsilon_0} \in B_{\varepsilon_0}^J(x) \subset B_r(x)$$

which is clearly a contradiction to $\|x - z_{\varepsilon_0}\| = 2r$. □

A.2.3 Legendre – Fenchel Calculus

Definition A.2.11. Let $J : X \to \overline{\mathbb{R}}$. The *Legendre – Fenchel conjugate* $J^* : X^* \to \overline{\mathbb{R}}$ of J is defined as

$$J^*(\xi^*) = \sup_{x \in X} \left(\langle \xi^*, x \rangle - J(x) \right).$$

From the definition of J^* it becomes clear, that for arbitrary $x \in X$ and $\xi^* \in X^*$

$$J(x) + J^*(\xi^*) \geq \langle \xi^*, x \rangle\,, \tag{A.9}$$

which is often referred to as *Fenchel's inequality.* Moreover one has

Lemma A.2.12. Let X be a normed space, $J : X \to \overline{\mathbb{R}}$ be a proper functional and $x \in X$, $\xi^* \in X^*$. Consider the statements

(a) $\xi^* \in \partial J(x)$,

(b) $J(x) + J^*(\xi^*) = \langle \xi^*, x \rangle$,

(c) $i_X(x) \in \partial J^*(\xi^*)$.

If J is convex, then $(a) \Leftrightarrow (b) \Rightarrow (c)$ and if additionally J is lower semicontinuous then also $(b) \Leftarrow (c)$.

Proof. [18, Chap.2 Prop.2.1] □

Example A.2.13. Let ϕ be a weight function and define $\omega : \mathbb{R} \to \overline{\mathbb{R}}$ as $\omega(s) = \psi_\phi(|s|)$. Then it is easy to see that $\omega^*(s) = \psi_{\phi^{-1}}(|s|)$ (note that $\phi^{-1} : [0,\infty) \to [0,\infty)$ is also a weight function). Moreover, for $s, t > 0$ one has the following relation

$$\psi_{\phi^{-1}}(t) = \sup_{s > \phi^{-1}(t)} \left(st - \psi_\phi(s) \right).$$

Note, that Lemma A.2.12 implies for $s, t \geq 0$ that

$$\psi_{\phi^{-1}}(t) + \psi_\phi(s) = st \;\Leftrightarrow\; t = \phi(s) \tag{A.10}$$

for $\partial\psi_\phi(\cdot) = \phi(\cdot)$. Note that in the case $t \neq \phi(s)$ one has

$$\psi_{\phi^{-1}}(t) + \psi_\phi(s) \geq st. \tag{A.11}$$

Figure A.2 depicts the situation.

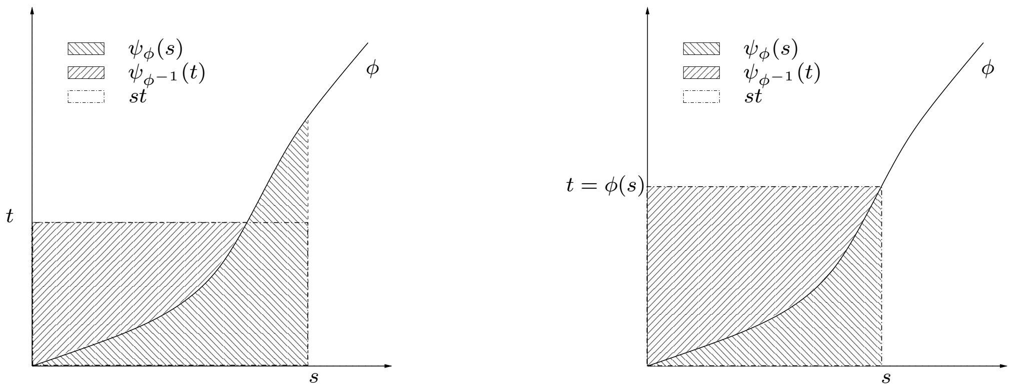

Figure A.2: Geometrical interpretation of Fenchel's inequality (A.11) (left) and equality (A.10) (right).

Lemma A.2.14. The Legendre – Fenchel conjugate $J^* : X^* \to \overline{\mathbb{R}}$ of a functional $J : X \to \overline{\mathbb{R}}$ is convex and sequentially lower semicontinuous in the weak* topology of X^*

Proof. Convexity of J^* follows directly from the definition. Assume that $\{\xi_n^*\}_{n\in\mathbb{N}}$ is a weakly* convergent sequence in X^*, i.e. there exists an element ξ^*, such that $\lim_{n\to\infty} \langle \xi_n^*, x\rangle = \langle x^*, x\rangle$ for all $x \in X$. From Fenchel's inequality (A.9) it follows that $J(x) + J^*(\xi^*) \geq \langle \xi^*, x\rangle$ for arbitrary $x \in X$ and $\xi^* \in \xi^*$. In particular we have for all $x \in X$ and $n \in \mathbb{N}$ that

$$J^*(\xi_n^*) \geq \langle \xi_n^*, x\rangle - J(x).$$

This implies that

$$\liminf_{n\to\infty} J^*(\xi_n^*) \geq \langle \xi^*, x\rangle - J(x)$$

for all $x \in X$. Hence taking the supreme over all $x \in X$ on the right hand side of above inequality gives the desired assertion. $\square$

Lemma A.2.15. Let $J : X \to \mathbb{R}$ and $\lambda \in \mathbb{R}\backslash\{0\}$. For $J_1(x) = \lambda J(x)$ and $J_2(x) = J(\lambda x)$ we have that

$$J_1^*(\xi^*) = \lambda J^*(\lambda^{-1}\xi^*), \quad J_2^*(\xi^*) = J^*(\lambda^{-1}\xi^*).$$

Proof. First we observe that

$$J_1^*(\xi^*) = \sup_{x\in X}\{\langle\xi^*, x\rangle - \lambda J(x)\} = \lambda \sup_{x\in X}\left\{\lambda^{-1}\langle\xi^*, x\rangle - J(x)\right\} = \lambda J^*(\lambda^{-1}\xi^*)$$

and analogously

$$J_2^*(\xi^*) = \sup_{x\in X}\{\langle\xi^*, x\rangle - J(\lambda x)\} = \lambda \sup_{x\in X}\left\{\lambda^{-1}\langle\xi^*, \lambda x\rangle - J(\lambda x)\right\} = J^*(\lambda^{-1}\xi^*)$$

□

In what follows we summarize some facts dealing with the composition of convex functionals $J : X \to \overline{\mathbb{R}}$ with linear operators $K : X \to Y$, where we assume that X and Y are at least normed vector spaces. We start with a

Lemma A.2.16. Assume that X and Y are normed spaces. If $K : X \to Y$ is a bounded and linear operator, then its adjoint $K^* : Y^* \to X^*$ is weak*-to-weak* continuous.

Proof. [94, Thm. 3.1.11] □

Corollary A.2.17. Assume that X and Y are normed spaces and that $K : X \to Y$ is a bounded and linear operator. If $J : X \to \overline{\mathbb{R}}$ is an arbitrary functional, then the composition $(J^* \circ K^*) : Y^* \to \overline{\mathbb{R}}$ is weakly* sequentially lower semicontinuous on Y^*.

Proof. Follows from Lemmata A.2.14 and A.2.16. □

Lemma A.2.18. Let ϕ be a weight function and $y \in Y$. Then

$$\partial\left(\psi_\phi\left(\|K(\cdot) - y\|\right)\right)(x) = K^*(\mathfrak{J}_\phi(K(x) - y))$$

for all $x \in X$.

Proof. Follows from Asplund's Theorem A.2.4 □

A.3 Analysis of the Proximal Point Method

In this section we collect some technical details of one single step in the proximal point method (cf. Algorithm 2.2.16). The upcoming results are used in Chapter 3. We essentially follow the analysis in Ambrosio et al. [9, Chap. 3.1].

In what follows, we assume that Assumptions 2.1.1 and 3.1.4 are satisfied. In particular, this means that X and Y are Banach spaces, where Y is assumed to be reflexive and that $K : X \to Y$ is linear and continuous. Moreover, we assume that $J : X \to \overline{\mathbb{R}}$ is convex, proper and lower semicontinuous.

For given data $y \in Y$, we recall the definition of the functional $F^*(\cdot; y) : Y^* \to \overline{\mathbb{R}}$ (cf. (2.13)) and of $\mu^*(y)$:

$$F^*(q^*; y) = J^*(K^*q^*) - \langle q^*, y\rangle \quad \text{and} \quad \mu^*(y) := \inf_{q^*\in Y^*} F^*(q^*; y).$$

From Lemma 2.2.14 it follows that $q^* \mapsto F^*(q^*; y)$ is convex, proper and sequentially weakly lower semicontinuous and that $\mu^*(y)$ is finite, when y is attainable, that is, when there exists a $x \in D(J)$ such that $Kx = y$.

Each step in Algorithm 2.2.16 consists in evaluating the resolvent operator (cf. Remark 2.2.18) defined by

$$R^\alpha_{F^*(\cdot;y)}(p^*) = \operatorname*{argmin}_{q^* \in Y^*} \left\{ \alpha^{-1} \psi_{\phi^{-1}}(\alpha \|q^* - p^*\|) + F^*(q^*; y) \right\},$$

where $p^* \in Y^*$ and $\alpha > 0$. We shall first show that $R^\alpha_{F^*(\cdot;y)}(p^*) \neq \emptyset$ for each $p^* \in Y^*$. Before we do so, we prove

Lemma A.3.1. Assume that $y \in Y$ and $\{\alpha_n\}_{n\in\mathbb{N}} \subset [\varepsilon, \infty)$ for $\varepsilon > 0$ and that $\{p_n^*\}_{n\in\mathbb{N}} \subset Y^*$ is bounded. Moreover, suppose that $\{q_n^*\}_{n\in\mathbb{N}} \subset Y^*$ is such that

$$\sup_{n\in\mathbb{N}} \left(\alpha_n^{-1} \psi_{\phi^{-1}}(\alpha_n \|q_n^* - p_n^*\|) + F^*(p_n^*; y) \right) < \infty.$$

Then $\{q_n^*\}_{n\in\mathbb{N}}$ is bounded.

Proof. According to assumption (R4) there exists at least one attainable element. Let $y_0 \in Y$ be such an element and observe from Lemma 2.2.14 that $\mu^*(y_0)$ is finite. This and the definition of $F^*(\cdot; y)$ show that

$$\langle q_n^*, y_0 - y \rangle = F^*(q_n^*; y) - F^*(q_n^*; y_0) \leq F^*(q_n^*; y) - \mu^*(y_0).$$

This, together with the assumptions of the Lemma gives

$$\begin{aligned} &\sup_{n\in\mathbb{N}} \left(\alpha_n^{-1} \psi_{\phi^{-1}}(\alpha_n \|q_n^* - p_n^*\|) + \langle q_n^*, y_0 - y \rangle \right) \\ &\qquad \leq \sup_{n\in\mathbb{N}} \left(\alpha_n^{-1} \psi_{\phi^{-1}}(\alpha_n \|q_n^* - p_n^*\|) + F^*(q_n^*; y) \right) - \mu^*(y_0) =: c < \infty. \end{aligned} \tag{A.12}$$

for an appropriately chosen constant $c \in \mathbb{R}$. Since $\alpha_n \geq \varepsilon > 0$ for all $n \in \mathbb{N}$ it follows from Lemma A.1.4 (1) that $\varepsilon^{-1} \psi_{\phi^{-1}}(\varepsilon \|q_n^* - p_n^*\|) \leq \alpha_n^{-1} \psi_{\phi^{-1}}(\alpha_n \|q_n^* - p_n^*\|)$ and consequently we find from (A.12)

$$\varepsilon^{-1} \psi_{\phi^{-1}}(\varepsilon \|q_n^* - p_n^*\|) \leq c + \langle q_n^*, y - y_0 \rangle .$$

With the substitution $r_n^* = q_n^* - p_n^*$ it follows for each $n \geq 1$

$$\varepsilon^{-1} \psi_{\phi^{-1}}(\varepsilon \|r_n^*\|) \leq c + \langle r_n^* + p_n^*, y - y_0 \rangle \leq c + \|y - y_0\| (\|p_n^*\| + \|r_n^*\|). \tag{A.13}$$

Assume, by contradiction, that $\sup_{n\in\mathbb{N}} \|q_n^*\| = \infty$. Since $\{p_n^*\}_{n\in\mathbb{N}}$ is assumed to be bounded it follows (possibly after dropping a subsequence) that

$$\lim_{n\to\infty} \|r_n^*\| = \infty.$$

Thus in particular $r_n^* \neq 0$ for all n large enough and estimate (A.13) eventually implies

$$\frac{\varepsilon^{-1} \psi_{\phi^{-1}}(\varepsilon \|r_n^*\|)}{\|r_n^*\|} \leq \|y_0 - y\| + \frac{c + \|p_n^*\| \, \|y - y_0\|}{\|r_n^*\|}.$$

Obviously the right hand side of the previous inequality stays bounded as $n \to \infty$ which, however, is a contradiction to Lemma A.1.4 (1) stating

$$\lim_{s\to\infty} s^{-1} \psi_{\phi^{-1}}(s) = \infty.$$

Hence the Lemma is shown. □

Lemma A.3.2. Let $y \in Y$, $\alpha > 0$ and $p^* \in Y^*$. Then there exists an element $p^*_\alpha \in Y^*$ such that

$$p^*_\alpha \in R^\alpha_{F^*(\cdot;y)}(p^*).$$

In particular, there exist $\xi \in \mathfrak{J}_{\phi^{-1}}(\alpha(p^*_\alpha - p^*))$ and $\theta \in \partial F^*(p^*_\alpha; y)$ such that $\xi + \theta = 0$.

Proof. Let $y \in Y$ and $\alpha > 0$. Moreover, assume that $p^* \in Y^*$ is arbitrary and assume that $\{q^*_n\}_{n\in\mathbb{N}} \subset Y^*$ satisfies

$$\begin{aligned}\lim_{n\to\infty} \alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^*_n - p^*\|) + F^*(q^*_n; y)\\ = \inf_{q^*\in Y^*} \alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^* - p^*\|) + F^*(q^*; y) =: \mu_0 \in [-\infty, \infty).\end{aligned}$$

Thus there exists a constant $c > 0$ such that for all $n \in \mathbb{N}$

$$\alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^*_n - p^*\|) + F^*(p^*_n; y) \le c, \tag{A.14}$$

as a consequence of which $\{q^*_n\}_{n\in\mathbb{N}}$ is uniformly bounded according to Lemma A.3.1.

According to requirement (R7), Y is reflexive and thus norm-bounded sets in Y^* are sequentially weakly compact. In other words, there exists an element $p^*_\alpha \in Y^*$ and a selection $n \mapsto \rho(n)$ such that

$$w\text{-}\lim_{n\to\infty} q^*_{\rho(n)} = p^*_\alpha.$$

The sequential weak lower semicontinuity of F^* and of the norm on Y^* (see e.g. [94, Thm. 2.5.21]) imply that

$$\alpha^{-1}\psi_{\phi^{-1}}(\alpha \|p^*_\alpha - p^*\|) + F^*(p^*_\alpha) \le \liminf_{n\to\infty} \alpha^{-1}\psi_{\phi^{-1}}\left(\alpha\left\|q^*_{\rho(n)} - p^*\right\|\right) + F^*(q^*_{\rho(n)}) = \mu_0.$$

Hence $p^*_\alpha \in R^\alpha_{F^*(\cdot;y)}(p^*)$.

The map $f : q^* \mapsto \alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^* - p^*\|)$ is continuous and thus it follows from optimality of p^*_α and the Moreau – Rockafellar Theorem (cf. [51, Chap. 1 Prop. 5.6]) that

$$0 \in \partial(f + F^*(\cdot; y))(p^*_\alpha) = \partial f(p^*_\alpha) + \partial F^*(p^*_\alpha; y).$$

From Asplund's theorem A.2.4 it eventually follows that $\partial f(p^*_\alpha) = \mathfrak{J}_{\phi^{-1}}(\alpha(p^*_\alpha - p^*))$. Thus there exist $\xi \in \mathfrak{J}_{\phi^{-1}}(\alpha(p^*_\alpha - p^*))$ and $\theta \in \partial F^*(p^*_\alpha; y)$ such that $\xi + \theta = 0$ and the Lemma is shown. □

For the remainder of this section we shall assume that $y \in Y$ is fixed and we shall write $F^*(p^*)$ instead of $F^*(p^*; y)$. Moreover, for each $\alpha > 0$ and $p^* \in Y^*$ we define

$$g(\alpha, p^*) := \inf_{q^*\in Y^*} \left\{\alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^* - p^*\|) + F^*(q^*)\right\}, \tag{A.15}$$

$$\mathrm{d}^+_\alpha(p^*) := \sup_{q^*\in R^\alpha_{F^*}(p^*)} \|q^* - p^*\| \quad \text{and} \quad \mathrm{d}^-_\alpha(p^*) := \inf_{q^*\in R^\alpha_{F^*}(p^*)} \|q^* - p^*\|. \tag{A.16}$$

The assertions of Lemmata A.3.3 - A.3.6 correspond to [9, Lem. 3.1.2], where the results are proven for the particular weight function $\phi^{-1}(s) = s^{p-1}$ for $p > 1$. Since the proofs in [9] are rather condensed, the passage to general ϕ^{-1} may not always be obvious. If necessary we therefore give the proofs, however, we shall always follow the ideas in [9].

Lemma A.3.3. Let $0 < \alpha_1 \leq \alpha_2$ and $p^* \in Y^*$. Then we have for all $p_{\alpha_i} \in R^{\alpha_i}_{F^*}(p^*)$ $(i = 1, 2)$ that

$$F^*(p^*) \geq g(\alpha_2, p^*) \geq g(\alpha_1, p^*) \quad \text{and} \quad \left\|p^*_{\alpha_2} - p^*\right\| \leq \left\|p^*_{\alpha_1} - p^*\right\|.$$

Proof. Let $q^* \in Y^*$ be chosen arbitrarily and define $s = \|q^* - p^*\|$. Then it follows from Lemma A.1.4 (1) that

$$\alpha_1^{-1}\psi_\phi(\alpha_1 s) = s(\alpha_1 s)^{-1}\psi_\phi(\alpha_1 s) \leq s(\alpha_2 s)^{-1}\psi_\phi(\alpha_2 s) = \alpha_2^{-1}\psi_\phi(\alpha_2 s).$$

Consequently we find

$$\alpha_1^{-1}\psi_\alpha(\alpha_1 \|q^* - p^*\|) + F^*(q^*) \leq \alpha_2^{-1}\psi_\alpha(\alpha_2 \|q^* - p^*\|) + F^*(q^*)$$

for all $q^* \in Y^*$. This implies that $g(\alpha_1, p^*) \leq g(\alpha_2, p^*)$.

We show that $F^*(p^*) \geq g(\alpha, p^*)$ for all $\alpha > 0$. To this end, we choose an arbitrary $p^*_\alpha \in R^\alpha_{F^*}(p^*)$ and note that according to Lemma A.3.2 there exist elements $\xi, \theta \in Y (= Y^{**})$ such that

$$\xi \in \mathfrak{J}_{\phi^{-1}}(\alpha(p^*_\alpha - p^*)), \quad \theta \in \partial F^*(p^*_\alpha; y) \quad \text{and} \quad \theta + \xi = 0.$$

From the definition of the subgradient it hence follows that

$$F^*(p^*; y) \geq F^*(p^*_\alpha; y) + \alpha^{-1}\left\langle \theta, \alpha(p^* - p^*_\alpha)\right\rangle = F^*(p^*_\alpha; y) + \alpha^{-1}\left\langle \xi, \alpha(p^*_\alpha - p^*)\right\rangle \tag{A.17}$$

The properties of the duality mapping $\mathfrak{J}_{\phi^{-1}}$ (cf. Definition A.1.1) imply

$$\left\langle \xi, \alpha(p^*_\alpha - p^*)\right\rangle = \|\xi\| \, \|\alpha(p^*_\alpha - p^*)\| = \phi^{-1}(\|\alpha(p^*_\alpha - p^*)\|) \, \|\alpha(p^*_\alpha - p^*)\|. \tag{A.18}$$

By observing that $\psi_{\phi^{-1}}(s) = \int_0^s \phi^{-1}(\sigma)\,\mathrm{d}\sigma \leq s\phi^{-1}(s)$, estimates (A.17) and (A.18) result in

$$F^*(p^*; y) \geq F^*(p^*_\alpha; y) + \alpha^{-1}\psi_{\phi^{-1}}(\alpha \|p^*_\alpha - p^*\|) = g(\alpha, p^*)$$

and the first part of the Lemma is shown.

It remains to prove that $\left\|p^*_{\alpha_2} - p^*\right\| \leq \left\|p^*_{\alpha_1} - p^*\right\|$. Since $p^*_{\alpha_i} \in R^{\alpha_i}_{F^*}(p^*)$ for $i = 1, 2$ we obtain

$$\alpha_i^{-1}\psi_{\phi^{-1}}(\alpha_i \left\|p^*_{\alpha_i} - p^*\right\|) + F^*(p^*_{\alpha_i}; y) \leq \alpha_i^{-1}\psi_{\phi^{-1}}(\alpha_i \left\|p^*_{\alpha_j} - p^*\right\|) + F^*(p^*_{\alpha_j}; y)$$

for $i \neq j$. Adding up the inequalities for $i = 1$ and $j = 2$ and vice versa yields

$$\begin{aligned} &\alpha_2^{-1}\psi_{\phi^{-1}}(\alpha_2 \left\|p^*_{\alpha_2} - p^*\right\|) - \alpha_1^{-1}\psi_{\phi^{-1}}(\alpha_1 \left\|p^*_{\alpha_2} - p^*\right\|) \\ &\qquad \leq \alpha_2^{-1}\psi_{\phi^{-1}}(\alpha_2 \left\|p^*_{\alpha_1} - p^*\right\|) - \alpha_1^{-1}\psi_{\phi^{-1}}(\alpha_1 \left\|p^*_{\alpha_1} - p^*\right\|). \end{aligned} \tag{A.19}$$

We note that for each $\alpha > 0$ the substitution $\sigma \to \alpha\sigma$ yields for $s > 0$

$$\alpha^{-1}\psi_{\phi^{-1}}(\alpha s) = \alpha^{-1}\int_0^{\alpha s} \phi^{-1}(\sigma)\,\mathrm{d}\sigma = \int_0^s \phi^{-1}(\alpha\sigma)\,\mathrm{d}\sigma.$$

Thus (A.19) says

$$\int_0^{\left\|p^*_{\alpha_2} - p^*\right\|} \phi^{-1}(\alpha_2\sigma) - \phi^{-1}(\alpha_1\sigma)\,\mathrm{d}\sigma \leq \int_0^{\left\|p^*_{\alpha_1} - p^*\right\|} \phi^{-1}(\alpha_2\sigma) - \phi^{-1}(\alpha_1\sigma)\,\mathrm{d}\sigma.$$

Since $\alpha_2 \geq \alpha_1$ and ϕ^{-1} is nondecreasing we conclude that the integrand $\phi^{-1}(\alpha_2\sigma) - \phi^{-1}(\alpha_1\sigma)$ is nonnegative and therefore $\left\|p^*_{\alpha_2} - p^*\right\| \leq \left\|p^*_{\alpha_1} - p^*\right\|$. $\square$

Lemma A.3.4. Let $0 < \alpha_1 \le \alpha_2$ and $p^* \in Y^*$. Then we have for all $p_{\alpha_i} \in R_{F^*}^{\alpha_i}(p^*)$ $(i = 1, 2)$ that

$$F^*(p^*_{\alpha_1}) \le F^*(p^*_{\alpha_2}) \le F^*(p^*) \quad \text{and} \quad \mathrm{d}^+_{\alpha_2}(p^*) \le \mathrm{d}^-_{\alpha_1}(p^*) \le \mathrm{d}^+_{\alpha_1}(p^*).$$

Moreover for λ^1- almost every $\alpha \in [0, \infty)$ we find

$$\mathrm{d}^+_{\alpha}(p^*) = \mathrm{d}^-_{\alpha}(p^*) =: \mathrm{d}^{\pm}_{\alpha}(p^*). \tag{A.20}$$

Proof. First, note that Lemma A.3.3 yields

$$F^*(p^*) \ge g(\alpha_2, p^*) = \alpha_2^{-1}\psi_\phi(\alpha_2 \left\|p^* - p^*_{\alpha_2}\right\|) + F^*(p^*_{\alpha_2}) \ge F^*(p^*_{\alpha_2}).$$

Moreover, from the optimality of $p^*_{\alpha_1}$ and again from Lemma A.3.3 it follows that

$$\begin{aligned}\alpha_1^{-1}\psi_{\phi^{-1}}(\alpha_1 \left\|q^* - p^*_{\alpha_1}\right\|) + F^*(p^*_{\alpha_1}) &\le \alpha_1^{-1}\psi_{\phi^{-1}}(\alpha_1 \left\|q^* - p^*_{\alpha_2}\right\|) + F^*(p^*_{\alpha_2}) \\ &\le \alpha_1^{-1}\psi_{\phi^{-1}}(\alpha_1 \left\|q^* - p^*_{\alpha_1}\right\|) + F^*(p^*_{\alpha_2}).\end{aligned}$$

Hence $F^*(p^*_{\alpha_1}) \le F^*(p^*_{\alpha_2})$.

The inequality $\mathrm{d}^-_{\alpha_1}(p^*) \le \mathrm{d}^+_{\alpha_1}(p^*)$ is obvious. Moreover, from Lemma A.3.3 it is evident that

$$\mathrm{d}^+_{\alpha_2}(p^*) = \sup_{q^* \in R^{\alpha_2}_{F^*}(p^*)} \|q^* - p^*\| \le \left\|p^*_{\alpha_1} - p^*\right\|, \qquad \text{for all } p^*_{\alpha_1} \in R^{\alpha_1}_{F^*}(p^*)$$

Taking the infimum over all $p^*_{\alpha_1} \in R^{\alpha_1}_{F^*}(p^*)$ results in

$$\mathrm{d}^+_{\alpha_2}(p^*) \le \mathrm{d}^-_{\alpha_1}(p^*). \tag{A.21}$$

In particular, this shows that the mapping $\alpha \mapsto \mathrm{d}^+_{\alpha}(p^*)$ is nonincreasing and therefore continuous λ^1-a.e. in $(0, \infty)$. Let $\alpha_0 > 0$ be a point of continuity of $\mathrm{d}^+_{\alpha}(q^*)$. Then (A.21) yields

$$\mathrm{d}^+_{\alpha_0}(q^*) = \lim_{\alpha \to \alpha_0^+} \mathrm{d}^+_{\alpha}(q^*) \le \mathrm{d}^-_{\alpha_0}(q^*) \le \mathrm{d}^+_{\alpha_0}(q^*)$$

and thus $\mathrm{d}^+_{\alpha_0}(q^*) = \mathrm{d}^-_{\alpha_0}(q^*)$. □

Lemma A.3.5. Let $p^* \in Y^*$ and $\alpha > 0$. Then there exists $p^*_\alpha \in R^{\alpha}_{F^*}(p^*)$ such that

$$\inf_{q^* \in R^{\alpha}_{F^*}(p^*)} F^*(q^*) = F^*(p^*_\alpha).$$

Proof. From Lemma A.3.1 we find that every minimizing sequence $\{q^*_n\}_{n\in\mathbb{N}} \subset R^{\alpha}_{F^*}(p^*)$ for F^* is bounded due to the fact that

$$\alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^*_n - p^*\|) + F^*(q^*_n; y) = g(\alpha, p^*) < \infty.$$

Thus there exists a selection $n \mapsto \rho(n)$ and $p^*_\alpha \in Y^*$ such that $p^*_{\rho(n)} \rightharpoonup p^*_\alpha$ and sequential weak lower semicontinuity of $\|\cdot\|_{Y^*}$ and F^* eventually shows that

$$\alpha^{-1}\psi_{\phi^{-1}}(\alpha \|p^*_\alpha - p^*\|) + F^*(p^*_\alpha) \le g(\alpha, p^*) \quad \text{and} \quad F^*(p^*_\alpha) \le \inf_{q^* \in R^{\alpha}_{F^*}(p^*)} F^*(q^*).$$

This, on the one hand, implies that $p^*_\alpha \in R^{\alpha}_{F^*}(p^*)$ and, on the other hand, proves

$$\inf_{q^* \in R^{\alpha}_{F^*}(p^*)} F^*(q^*) = F^*(p^*_\alpha).$$

□

Lemma A.3.6. For every $p^* \in \overline{D(F^*(\cdot;y))}$ we have that

$$\lim_{\alpha\to\infty} g(\alpha,p^*) = \lim_{\alpha\to\infty} \inf_{q^*\in R^{\alpha}_{F^*}(p^*)} F^*(q^*) = F^*(p^*).$$

Proof. Let $\{\alpha_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ be a increasing sequence such that

$$\lim_{n\to\infty} \alpha_n = \infty.$$

From Lemma A.3.5 it follows that for each $n \in \mathbb{N}$ there exists an element $p^*_{\alpha_n} \in R^{\alpha_n}_{F^*}(p^*)$ such that

$$\inf_{q^*\in R^{\alpha_n}_{F^*}(p^*)} F^*(q^*) = F^*(p^*_{\alpha_n}).$$

Then we find, according to Lemma A.3.3, that

$$\left\|p^*_{\alpha_n}\right\| \le \|p^*\| + \left\|p^* - p^*_{\alpha_n}\right\| \le \|p^*\| + \left\|p^* - p^*_{\alpha_1}\right\| < +\infty,$$

that is, $\left\{\left\|p^*_{\alpha_n}\right\|\right\}_{n\in\mathbb{N}}$ is bounded and thus we can find an element $p^*_\infty \in Y^*$ with

$$w\text{-}\lim_{n\to\infty} p^*_{\alpha_{\rho(n)}} = p^*_\infty$$

for a suitable selection $n \mapsto \rho(n)$. We show that $p^*_\infty = p^*$ and will then conclude (by a standard sub-subsequence argument), that the whole sequence $\left\{p^*_{\alpha_n}\right\}_{n\in\mathbb{N}}$ weakly converges to p^*.

In order to keep the notation transparent, we will assume in the following paragraph, that $p_{\alpha_n} \rightharpoonup p^*_\infty$. Since $p^*_{\alpha_n} \in R^{\alpha_n}_{F^*}(p^*)$, it follows from Lemma A.3.2 that for all $n \in \mathbb{N}$ there exist elements $\xi_n, \theta_n \in Y$ such that

$$\xi_n \in \mathfrak{J}_{\phi^{-1}}(\alpha_n(p^*_{\alpha_n} - p^*)), \quad \theta_n \in \partial F^*(p^*_{\alpha_n}) \quad \text{and} \quad \xi_n + \theta_n = 0.$$

Now choose an arbitrary pair $(\zeta^*, \zeta) \in Y^* \times Y$ such that $\zeta \in \partial F^*(\zeta^*)$. Then the monotonicity of the subgradient ∂F^* yields

$$\left\langle \zeta - \theta_n, \zeta^* - p^*_{\alpha_n}\right\rangle \ge 0, \qquad \text{for all } n \ge 1. \tag{A.22}$$

Moreover, from the fact that $\theta_n = -\xi_n \in \mathfrak{J}_{\phi^{-1}}(\alpha_n(p^* - p^*_{\alpha_n}))$ we find

$$\left\langle \theta_n, \alpha_n(p^* - p^*_{\alpha_n})\right\rangle = \|\theta_n\| \left\|\alpha_n(p^*_{\alpha_n} - p^*)\right\| = \phi^{-1}\left(\left\|\alpha_n(p^*_{\alpha_n} - p^*)\right\|\right)\left\|\alpha_n(p^*_{\alpha_n} - p^*)\right\|. \tag{A.23}$$

Using the abbreviation $\eta(s) = s\phi^{-1}(s)$ and combining (A.22) and (A.23) show

$$\alpha_n^{-1}\eta\big(\left\|\alpha_n(p^*_{\alpha_n} - p^*)\right\|\big) \le \langle\theta_n, p^* - \zeta^*\rangle + \left\langle \zeta, \zeta^* - p^*_{\alpha_n}\right\rangle. \tag{A.24}$$

Assume now that $\inf_{n\in\mathbb{N}}\left\|p^*_{\alpha_n} - p^*\right\| \ge \gamma$ for a constant $\gamma > 0$ (Note that if such a γ can not be chosen, we can drop a further subsequence that strongly converges to p^* and nothing remains to be shown). Then the monotonicity of ϕ^{-1} implies that

$$\beta_n := \phi^{-1}\left(\alpha_n \left\|p^*_{\alpha_n} - p^*\right\|\right) \ge \phi^{-1}(\alpha_n\gamma)$$

and consequently one has that $\beta_n \to \infty$ as $n \to \infty$. Dividing (A.24) by $\beta_n \neq 0$ eventually gives (note that $\left\{\left\|p^*_{\alpha_n}\right\|\right\}_{n\in\mathbb{N}}$ is bounded)

$$\left\|p^*_{\alpha_n} - p^*\right\| \le \beta_n^{-1}\left(\left\langle\theta_n, p^* - \zeta^*\right\rangle + \left\langle \zeta, \zeta^* - p^*_{\alpha_n}\right\rangle\right) \le \beta_n^{-1}\|\theta_n\|\,\|p^* - \zeta^*\| + \mathcal{O}(\beta_n^{-1}). \tag{A.25}$$

Since $\theta_n \in \mathfrak{J}_{\phi^{-1}}(\alpha_n(p^*_{\alpha_n} - p^*))$ it follows from the definition of $\mathfrak{J}_{\phi^{-1}}$ that

$$\|\theta_n\| = \phi^{-1}\left(\alpha_n \left\|p^*_{\alpha_n} - p^*\right\|\right) = \beta_n, \quad \text{for all } n \geq 1.$$

Hence (A.25) and the weak lower semicontinuity of the dual norm in Y^* imply

$$\|p^*_\infty - p^*\| \leq \liminf_{n\to\infty}\left\|p^*_{\alpha_n} - p^*\right\| \leq \|\zeta^* - p^*\|.$$

Since $\overline{D(\partial F^*)} = \overline{D(F^*)}$ and $\zeta^* \in D(\partial F^*(\cdot; y))$ was chosen arbitrarily, this implies that $p^*_\infty = p^*$ whenever $p^* \in \overline{D(F^*)}$.

Summarizing we note that for each subsequence of $\left\{p^*_{\alpha_n}\right\}_{n\in\mathbb{N}}$ one can drop a further subsequence that weakly converges to p^*, which shows

$$\underset{n\to\infty}{w\text{-}\lim}\, p^*_{\alpha_n} = p^*.$$

Weak lower semicontinuity of F^* hence gives

$$\lim_{n\to\infty} F^*(p^*_{\alpha_n}) \geq F^*(p^*).$$

From Lemma A.3.4 it follows that $F^*(p^*_{\alpha_n}) \leq F^*(p^*)$ for all $n \in \mathbb{N}$. Thus it follows that

$$\lim_{\alpha\to\infty} \inf_{q^*\in R^\alpha_{F^*}(p^*)} F^*(q^*) = \lim_{n\to\infty} F^*(p^*_{\alpha_n}) = F^*(p^*).$$

We complete the proof by observing that for all $n \in \mathbb{N}$

$$F^*(p^*_{\alpha_n}) \leq \alpha_n^{-1}\psi_{\phi^{-1}}(\alpha_n \left\|p^*_{\alpha_n} - p^*\right\|) + F^*(p^*_{\alpha_n}) = g(\alpha_n, p^*) \leq F^*(p^*),$$

where the last inequality follows from Lemma A.3.3. □

Remark A.3.7. If $p^* \notin \overline{D(F^*)}$, one has

$$\lim_{\alpha\to\infty} g(\alpha, p^*) = +\infty \ (= F^*(p^*)).$$

The next Lemma shows that for each fixed $p^* \in Y^*$, the mapping $\alpha \mapsto g(\alpha, p^*)$ is differentiable a.e. in $[0, \infty)$. The assertion corresponds to [9, Lem. 3.1.4].

Lemma A.3.8. For given $p^* \in Y^*$ the mapping $\alpha \mapsto g(\alpha, p^*)$ is differentiable a.e. in $(0, \infty)$ and

$$\frac{d}{d\alpha} g(\alpha, p^*) = \frac{1}{\alpha^2}\psi_\phi\left(\phi^{-1}(\alpha\, \mathrm{d}^\pm_\alpha(p^*))\right),$$

where $\mathrm{d}^\pm_\alpha$ is defined in (A.20).

Proof. Let $p^* \in Y^*$ and $\alpha, \beta \in (0, \infty)$. Moreover, choose $p^*_\alpha \in R^\alpha_{F^*}(p^*)$ and $p^*_\beta \in R^\beta_{F^*}(p^*)$. Recall from (2.18) in Chapter 2 that for all $s \geq 0$

$$\phi^{-1}(s)s - \psi_{\phi^{-1}}(s) = \psi_\phi(\phi^{-1}(s)). \tag{A.26}$$

First, we remark that from the definition of g in (A.15) it follows that

$$g(\alpha, p^*) = \inf_{q^*\in Y^*} \alpha^{-1}\psi_{\phi^{-1}}(\alpha \|q^* - p^*\|) + F^*(q^*) \leq \alpha^{-1}\psi_{\phi^{-1}}(\alpha \left\|p^*_\beta - p^*\right\|) + F^*(p^*_\beta).$$

This yields

$$\begin{aligned} g(\alpha,p^*)-g(\beta,p^*) &\le \alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\beta-p^*\right\|)+F^*(p^*_\beta)-g(\beta,p^*)\\ &=\alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\beta-p^*\right\|)+F^*(p^*_\beta)-\left(\beta^{-1}\psi_{\phi^{-1}}(\beta\left\|p^*_\beta-p^*\right\|)+F^*(p^*_\beta)\right)\\ &=\alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\beta-p^*\right\|)-\beta^{-1}\psi_{\phi^{-1}}(\beta\left\|p^*_\beta-p^*\right\|). \end{aligned}$$

If $\beta<\alpha$ this implies

$$\frac{g(\alpha,p^*)-g(\beta,p^*)}{\alpha-\beta}\le\frac{\alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\beta-p^*\right\|)-\beta^{-1}\psi_{\phi^{-1}}(\beta\left\|p^*_\beta-p^*\right\|)}{\alpha-\beta} \tag{A.27}$$

After passing $\beta\to\alpha^-$ it follows from the fact that $\frac{d}{ds}\psi_{\phi^{-1}}(s)=\phi^{-1}(s)$ and (A.26)

$$\begin{aligned} \frac{\mathrm{d}^-}{\mathrm{d}\alpha}g(\alpha,p^*) &\le \frac{\mathrm{d}^-}{\mathrm{d}\alpha}\left(\alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\beta-p^*\right\|)\right)\\ &=\frac{\phi^{-1}(\alpha\|p^*_\beta-p^*\|)\alpha\|p^*_\beta-p^*\|-\psi_{\phi^{-1}}(\alpha\|p^*_\beta-p^*\|)}{\alpha^2}\\ &=\frac{\psi_\phi(\phi^{-1}(\alpha\|p^*_\beta-p^*\|))}{\alpha^2}. \end{aligned}$$

Since the mapping $s\mapsto\psi_\phi(\phi^{-1}(s))$ is nondecreasing, we can take the infimum over all $p^*_\beta\in R^\beta_{F^*}(p^*)$ in above formula and obtain (recall the definition of $\mathrm{d}^-_\alpha(p^*)$ in A.16)

$$\frac{\mathrm{d}^-}{\mathrm{d}\alpha}g(\alpha,p^*)\le\frac{1}{\alpha^2}\psi_\phi(\phi^{-1}(\alpha\,\mathrm{d}^-_\alpha(p^*))). \tag{A.28}$$

Vice versa, if $\beta>\alpha$ we end up with

$$\frac{g(\alpha,p^*)-g(\beta,p^*)}{\alpha-\beta}\ge\frac{\alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\beta-p^*\right\|)-\beta^{-1}\psi_{\phi^{-1}}(\beta\left\|p^*_\beta-p^*\right\|)}{\alpha-\beta}$$

and consequently after passing $\beta\to\alpha^+$ and taking the supremum over all $p^*_\beta\in R^\beta_{F^*}(p^*)$ this gives

$$\frac{\mathrm{d}^+}{\mathrm{d}\alpha}g(\alpha,p^*)\ge\frac{1}{\alpha^2}\psi_\phi(\phi^{-1}(\alpha\,\mathrm{d}^+_\alpha(p^*))). \tag{A.29}$$

According to Lemma A.3.3 the mapping $\alpha\mapsto g(\alpha,p^*)$ is nondecreasing and therefore differentiable for λ^1-a.e. $\alpha\ge0$. Therefore it follows from (A.28) and (A.29) that

$$\frac{1}{\alpha^2}\psi_\phi(\phi^{-1}(\alpha\,\mathrm{d}^+_\alpha(p^*)))\le\frac{d}{d\alpha}g(\alpha,p^*)\le\frac{1}{\alpha^2}\psi_\phi(\phi^{-1}(\alpha\,\mathrm{d}^-_\alpha(p^*))),\quad\text{for }\lambda^1\text{-a.e. }\alpha\ge0.$$

The assertion now follows from (A.20). □

Combining Lemma A.3.6 and Lemma A.3.8 and taking into account Remark A.3.7 shows

Proposition A.3.9. *Let $p^*\in Y^*$ and $\alpha>0$. Then, for all $p^*_\alpha\in R^\alpha_{F^*}(p^*)$ one has that*

$$F^*(p^*)-F^*(p^*_\alpha)=\alpha^{-1}\psi_{\phi^{-1}}(\alpha\left\|p^*_\alpha-p^*\right\|)+\int_\alpha^\infty\frac{1}{\omega^2}\psi_\phi(\phi^{-1}(\omega\,d^\pm_\omega(p^*)))\,d\omega. \tag{A.30}$$

Proof. According to Lemma A.3.8 one has for λ^1-almost all $\beta > 0$ that

$$g(\beta, p^*) - g(\alpha, p^*) = \int_\alpha^\beta \frac{d}{d\omega} g(\omega, p^*)\,\mathrm{d}\omega = \int_\alpha^\beta \frac{1}{\omega^2}\psi_\phi(\phi^{-1}(\omega\,\mathrm{d}_\omega^\pm(p^*)))\,\mathrm{d}\omega.$$

Since $p_\alpha^* \in R_{F^*}^\alpha(p^*)$ this implies (recall the definition of $g(\alpha, p^*)$ in (A.15)) that

$$g(\beta, p^*) - F^*(p_\alpha^*) = \alpha^{-1}\psi_{\phi^{-1}}(\alpha\,\|p_\alpha^* - p^*\|) + \int_\alpha^\beta \frac{1}{\omega^2}\psi_\phi(\phi^{-1}(\omega\,\mathrm{d}_\omega^\pm(p^*)))\,\mathrm{d}\omega.$$

When passing to the limit $\beta \to \infty$ the assertion follows from Lemma A.3.6 (see also Remark A.3.7). □

We close this section with an estimate for the slope $|\partial F^*(\cdot; y)|$ (cf. Definition A.2.2).

Lemma A.3.10. Let $p^* \in Y^*$ and $\alpha > 0$. Then the estimate

$$|\partial F^*(\cdot; y)|\,(p_\alpha^*) \le \phi^{-1}(\alpha\,\|p_\alpha^* - p^*\|)$$

holds for all $p_\alpha^* \in R_{F^*}^\alpha(p^*)$.

Proof. Since $p_\alpha^* \in R_{F^*}^\alpha(p^*)$, it follows from Lemma A.15 that there exist elements $\xi, \theta \in Y$ with

$$\xi \in \mathfrak{J}_{\phi^{-1}}(\alpha(p_\alpha^* - p^*)), \quad \theta \in \partial F^*(p_\alpha^*) \quad \text{and} \quad \xi + \theta = 0.$$

Hence from the definition of the subgradient it follows that

$$F^*(p_\alpha^*) - F^*(q^*) \le \langle \theta, p_\alpha^* - q^* \rangle\,, \qquad \text{for all } q^* \in Y^*.$$

Since $\xi \in \mathfrak{J}_{\phi^{-1}}(\alpha(p_\alpha^* - p^*))$ it follows that $\|\theta\| = \|\xi\| = \phi^{-1}(\alpha\,\|p_\alpha^* - p^*\|)$ and thus the previous estimate together with Lemma A.2.2 (2) gives

$$|\partial F^*|\,(p_\alpha^*) = \sup_{q^* \in Y^* \setminus \{p_\alpha^*\}} \frac{F^*(p_\alpha^*) - F^*(q^*)}{\|p_\alpha^* - q^*\|} \le \phi^{-1}(\alpha\,\|p^* - p_\alpha^*\|).$$

□

Bibliography

[1] R. Acar and C. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10:1217–1229, 1994.

[2] R. A. Adams. *Sobolev Spaces*, volume 65 of *Pure and Applied Mathematics*. Academic Press, New York - London, 1975.

[3] Y. Alber and D. Butnariu. Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces. *J. Optim. Theory Appl.*, 92(1):33–61, 1997.

[4] Y. Alber and I. Ryazantseva. *Nonlinear ill-posed problems of monotone type*. Springer, Dordrecht, 2006.

[5] Y. I. Alber, R. S. Burachik, and A. N. Iusem. A proximal point method for nonsmooth convex optimization problems in Banach spaces. *Abstr. Appl. Anal.*, 2(1-2):97–120, 1997.

[6] F. Alter, V. Caselles, and A. Chambolle. A characterization of convex calibrable sets in $\mathbb{R}^N$. *Math. Ann.*, 332(2):329–366, 2005.

[7] F. Alter, V. Caselles, and A. Chambolle. Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow. *Interfaces Free Bound.*, 7(1):29–53, 2005.

[8] L. Ambrosio. Minimizing movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)*, 19:191–246, 1995.

[9] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows*. Lectures in mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[10] F. Andreu, C. Ballester, V. Caselles, J. I. Díaz, and J. M. Mazón. Minimizing total variation flow. *Differential Integral Equations*, 14(3):347–403, 2001.

[11] F. Andreu-Vaillo, V. Caselles, and J. M. Mazón. *Parabolic quasilinear equations minimizing linear growth functionals*, volume 223 of *Progress in Mathematics*. Birkhäuser, Basel, 2004.

[12] G. Anzellotti. Pairings between measures and bounded functions and compensated compactness. *Ann. Math. Pura Appl.*, 135(4):293–318, 1983.

[13] G. Aubert and J.-F. Aujol. Modeling very oscillationg signals. application to image processing. *Appl. Math. Optim.*, 51(2):163–182, 2005.

[14] J.-P. Aubin. *Optima and equilibria. An introduction to nonlinear analysis.*, volume 140 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 1993.

[15] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition: Application to sar images. In L. Griffin and M. Lillhol, editors, *Scale Space*, volume 2695 of *Lecture Notes in Computer Science*, pages 297–312, Berlin, 2003. Springer-Verlag.

[16] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *J. Math. Imaging Vision*, 22(1):71–88, 2005.

[17] V. Barbu. *Nonlinear semigroups and differential equations in Banach spaces.* Editura Academiei Republicii Socialiste Romnia, Bucharest, 1976.

[18] V. Barbu and T. Precupanu. *Convexity and optimization in Banach spaces.* Editura Academiei, Bucharest, revised edition, 1978. Translated from the Romanian.

[19] H. H. Bauschke, E. Matoušková, and S. Reich. Projection and proximal point methods: convergence results and counterexamples. *Nonlinear Anal.*, 56(5):715–738, 2004.

[20] G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in $\mathbb{R}^n$. *J. Differential Equations*, 184(2):475–525, 2002.

[21] G. Bellettini, V. Caselles, and M. Novaga. Explicit solutions of the eigenvalue problem $-\mathrm{div}\left(\frac{Du}{|du|}\right) = u$ in $\mathbb{R}^2$. *SIAM J. Math. Anal.*, 36(4):1095–1129, 2005.

[22] D. P. Bertsekas. Combined primal-dual and penalty methods for constrained minimization. *SIAM J. Control*, 13:521–544, 1975.

[23] D. P. Bertsekas. On penalty and multiplier methods for constrained minimization. *SIAM J. Control Optimization*, 14(2):216–235, 1976.

[24] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Computer Science and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982.

[25] J. Bourgain and H. Brézis. On the equation $\mathrm{div}\, y = f$ and application to control of phases. *J. Amer. Math. Soc.*, 16(2):393–426, 2003.

[26] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* Number 5 in North-Holland Mathematics Studies. North-Holland Publishing Co, Amsterdam-London, 1973.

[27] H. Brézis and P.-L. Lions. Produits infinis de résolvantes. *Israel J. Math.*, 29(4):329–345, 1978.

[28] M. Brill and E. Schock. Iterative solution of ill-posed problems—a survey. In *Model optimization in exploration geophysics (Berlin, 1986)*, volume 1 of *Theory Practice Appl. Geophys.*, pages 13–37. Vieweg, Braunschweig, 1987.

[29] R. S. Burachik and A. N. Iusem. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM J. Optim.*, 8(1):197–216 (electronic), 1998.

[30] M. Burger, K. Frick, S. Osher, and O. Scherzer. Inverse total variation flow. *Multiscale Model. Simul.*, 6(2):365–395 (electronic), 2007.

[31] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Commun. Math. Sci.*, 4(1):179–212, 2006.

[32] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

[33] M. Burger, S. Osher, J. Xu, and G. Gilboa. Nonlinear inverse scale space methods for image restoration. In N. Paragios, O. D. Faugeras, T. Chan, and C. Schnörr, editors, *VLSM*, volume 3752 of *Lecture Notes in Computer Science*, pages 25–36. Springer, 2005.

[34] D. Butnariu, Y. Censor, and S. Reich. Iterative averaging of entropic projections for solving stochastic convex feasibility problems. *Comput. Optim. Appl.*, 8(1):21–39, 1997.

[35] D. Butnariu and A. N. Iusem. On a proximal point method for convex optimization in Banach spaces. *Numer. Funct. Anal. Optim.*, 18(7-8):723–744, 1997.

[36] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.

[37] Y. Censor and S. A. Zenios. Proximal minimization algorithm with D-functions. *J. Optim. Theory Appl.*, 73(3):451–464, 1992.

[38] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1–2):89–97, 2004.

[39] T. F. Chan and S. Esedoḡlu. Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.*, 65(5):1817–1837, 2005.

[40] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.*, 3(3):538–543, 1993.

[41] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Math. Programming*, 64(1, Ser. A):81–101, 1994.

[42] I. Cioranescu. *Geometry of Banach spaces, duality mappings and nonlinear problems.* Kluwer Academic Publishers Group, Dordrecht, 1990.

[43] J. A. Clarkson. Uniformly convex spaces. *Trans. Amer. Math. Soc.*, 40(3):396–414, 1936.

[44] M. G. Crandall and T. M. Liggett. Generation of semi-groups of nonlinear transformations on general Banach spaces. *Amer. J. Math.*, 93:265–298, 1971.

[45] E. De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 81–98. Masson, Paris, 1993.

[46] E. De Giorgi, A. Marino, and M. Tosques. Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 68(3):180–187, 1980.

[47] J. Diestel and J. J. Uhl, Jr. *Vector measures.* American Mathematical Society, Providence, R.I., 1977. With a foreword by B. J. Pettis, Mathematical Surveys, No. 15.

[48] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.

[49] N. Dunford and J. T. Schwartz. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[50] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.*, 18(1):202–226, 1993.

[51] I. Ekeland and R. Temam. *Convex analysis and variational problems*, volume 1 of *Studies in Mathematics and its Applications.* North-Holland Publishing Co., Amsterdam-Oxford, 1976.

[52] J. Elstrodt. *Maß- und Integrationstheorie.* Springer, Berlin - Heidelberg - New York, 1996.

[53] H. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems.* Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1996.

[54] S. Esedoḡlu and S. J. Osher. Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. *Comm. Pure Appl. Math.*, 57(12):1609–1626, 2004.

[55] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[56] K. Fan and I. Glicksberg. Some geometric properties of the spheres in a normed linear space. *Duke Math. J.*, 25:553–568, 1958.

[57] H. Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[58] A. V. Fiacco and G. P. McCormick. The slacked unconstrained minimization technique for convex programming. *SIAM J. Appl. Math.*, 15:505–515, 1967.

[59] M. Fortin. Minimization of some non-differentiable functionals by the augmented Lagrangian method of Hestenes and Powell. *Appl. Math. Optim.*, 2(3):236–250, 1975/76.

[60] M. Fortin and R. Glowinski. *Augmented Lagrangian methods*, volume 15 of *Studies in Mathematics and its Applications.* North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.

[61] K. Frick and O. Scherzer. Convex inverse scale spaces. In F. Sgallari, A. Murli, and N. Paragios, editors, *SSVM*, volume 4485 of *Lecture Notes in Computer Science*, pages 313–325. Springer, 2007.

[62] M. Fuchs and K. Frick. Imaging C++ library, 2005. http://www.stochastik.math.uni-goettingen.de/~frick/imaging/index.html.

[63] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.*, 4(7):932–946, 1995.

[64] U. Gianazza and G. Savaré. Some results on minimizing movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. (5)*, 18:57–80, 1994.

[65] U. Gianazza and G. Savaré. Abstract evolution equations on variable domains: an approach by minimizing movements. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 23(1):149–178, 1996.

[66] R. Glowinski. *Numerical methods for nonlinear variational problems.* Springer Series in Computational Physics. Springer-Verlag, New York, 1984.

[67] R. Glowinski and A. Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *Rev. Française Automat. Informat. Recherche Opérationnelle RAIRO Analyse Numérique*, 9(R-2):41–76, 1975.

[68] C. W. Groetsch. An iterative stabilization method for the evaluation of unbounded operators. *Proc. Amer. Math. Soc.*, 134(4):1173–1181 (electronic), 2006.

[69] C. W. Groetsch. *Stable approximate evaluation of unbounded operators.* Number 1894 in Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2007.

[70] C. W. Groetsch and O. Scherzer. Non-stationary iterated Tikhonov-Morozov method and third-order differential equations for the evaluation of unbounded operators. *Math. Methods Appl. Sci.*, 23(15):1287–1300, 2000.

[71] F. Guichard, J. M. Morel, and R. Ryan. *Contrast invariant image analysis and PDE's.*

[72] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29(2):403–419, 1991.

[73] O. Güler. New proximal point algorithms for convex minimization. *SIAM J. Optim.*, 2(4):649–664, 1992.

[74] J. S. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–53, 1902.

[75] A. Haddad and Y. Meyer. An improvement of Rudin-Osher-Fatemi model. *Appl. Comput. Harmon. Anal.*, 22(3):319–334, 2007.

[76] M. Hanke and C. W. Groetsch. Nonstationary iterated Tikhonov regularization. *J. Optim. Theory Appl.*, 98(1):37–53, 1998.

[77] M. R. Hestenes. Multiplier and gradient methods. *J. Optimization Theory Appl.*, 4:303–320, 1969.

[78] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[79] R. B. Holmes. *A course on optimization and best approximation.* Springer-Verlag, Berlin, 1972. Lecture Notes in Mathematics, Vol. 257.

[80] S. Hu and N. S. Papageorgiou. *Handbook of multivalued analysis. Vol. I*, volume 419 of *Mathematics and its Applications.* Kluwer Academic Publishers, Dordrecht, 1997. Theory.

[81] A. D. Ioffe and V. M. Tihomirov. *Theorie der Extremalaufgaben.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1979. Translated from the Russian by Bernd Luderer.

[82] K. Ito and K. Kunisch. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Math. Programming*, 46(3, (Ser. A)):341–360, 1990.

[83] K. Ito and K. Kunisch. An augmented Lagrangian technique for variational inequalities. *Appl. Math. Optim.*, 21(3):223–241, 1990.

[84] K. Ito and K. Kunisch. An active set strategy based on the augmented Lagrangian formulation for image restoration. *M2AN Math. Model. Numer. Anal.*, 33(1):1–21, 1999.

[85] K. Ito and K. Kunisch. Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces. *Nonlinear Anal.*, 41(5-6, Ser. A: Theory Methods):591–616, 2000.

[86] A. Iusem. Augmented Lagrangian methods and proximal point methods for convex optimization. *Inverstigación Operativa*, 8:11–50, 1999.

[87] K. C. Kiwiel. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim.*, 35(4):1142–1168, 1997.

[88] B. W. Kort and D. P. Bertsekas. Combined primal-dual and penalty methods for convex programming. *SIAM J. Control Optimization*, 14(2):268–294, 1976.

[89] M. A. Krasnosel′skiĭ. Solution of equations involving adjoint operators by successive approximations. *Uspehi Mat. Nauk*, 15(3 (93)):161–165, 1960.

[90] A. V. Krjanev. Solution of ill-posed problems by successive approximation methods. *Dokl. Akad. Nauk SSSR*, 210:20–22, 1973.

[91] L. J. Lardy. A series representation for the generalized inverse of a closed linear operator. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 58(2):152–157, 1975.

[92] B.-L. Lin and P.-K. Lin. Property (H) in Lebesgue-Bochner function spaces. *Proc. Amer. Math. Soc.*, 95(4):581–584, 1985.

[93] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Ser. R-3):154–158, 1970.

[94] R. E. Megginson. *An introduction to Banach space theory*, volume 183 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1998.

[95] Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[96] A. Miele, E. E. Cragg, R. R. Iyer, and A. V. Levy. Use of the augmented penalty function in mathematical programming problems, part 1. *J. Optim. Theory Appl.*, 8(2):115–130, 1971.

[97] A. Miele, E. E. Cragg, and A. V. Levy. Use of the augmented penalty function in mathematical programming problems, part 2. *J. Optim. Theory Appl.*, 8(2):131–153, 1971.

[98] A. Miele, P. E. Moseley, A. V. Levy, and G. M. Coggins. On the method of multipliers for mathematical programming problems. *J. Optimization Theory Appl.*, 10:1–33, 1972.

[99] I. Miyadera. *Nonlinear semigroups*, volume 109 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1992. Translated from the 1977 Japanese original by Choong Yun Cho.

[100] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[101] J. v. Neumann. Über adjungierte Funktionaloperatoren. *Ann. of Math. (2)*, 33(2):294–310, 1932.

[102] R. H. Nochetto, G. Savaré, and C. Verdi. A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations. *Comm. Pure Appl. Math.*, 53(5):525–589, 2000.

[103] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4(2):460–489 (electronic), 2005.

[104] N. H. Pavel. *Nonlinear evolution operators and semigroups*, volume 1260 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1987. Applications to partial differential equations.

[105] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[106] D. L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.

[107] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization (Sympos., Univ. Keele, Keele, 1968)*, pages 283–298. Academic Press, London, 1969.

[108] E. Radmoser, O. Scherzer, and J. Weickert. Scale-space properties of nonstationary iterative regularization methods. *Journal of Visual Communication and Image Representation*, 11:96–114, 2000.

[109] E. Resmerita. On total convexity, Bregman projections and stability in Banach spaces. *J. Convex Anal.*, 11(1):1–16, 2004.

[110] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[111] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

[112] R. T. Rockafellar. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math. Programming*, 5:354–373, 1973.

[113] R. T. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming. *J. Optimization Theory Appl.*, 12:555–562, 1973.

[114] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

[115] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[116] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.

[117] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *J. Math. Imaging Vision*, 12(1):43–63, 2000.

[118] M. A. Smith and B. Turett. Rotundity in Lebesgue-Bochner function spaces. *Trans. Amer. Math. Soc.*, 257(1):105–118, 1980.

[119] A. N. Tikhonov. On the regularization of ill-posed problems. *Dokl. Akad. Nauk SSSR*, 153:49–52, 1963.

[120] A. N. Tikhonov. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, 151:501–504, 1963.

[121] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems.* V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[122] L. Vese and S. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *J. Sci. Comput*, 19(1-3):553–572, 2003.

[123] J. Weickert. *Anisotropic Diffusion in Image Processing.* ECMI. B.G. Teubner, Stuttgart, 1998.

[124] D. Werner. *Funktionalanalysis.* Springer Verlag, Berlin, 2002.

[125] W. P. Ziemer. *Weakly differentiable functions.* Springer Verlag, New York, 1989.