CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Multimedia ONline ARchiv CHemnitz

TECHNISCHE UNIVERSITÄT CHEMNITZ

Fakultät für Mathematik

Professur Analysis – Inverse Probleme

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)

Generalized Tikhonov regularization

Basic theory and comprehensive results on convergence rates

vorgelegt von

Dipl.-Math. Jens Flemming

geboren am 1. Oktober 1985 in Gera

Chemnitz, den 28. Oktober 2011

Eingereicht am 31. Mai 2011, verteidigt am 27. Oktober 2011

Betreuer/Gutachter:	Prof. Dr. Bernd Hofmann
Zweitgutachter:	Prof. Dr. Thorsten Hohage (Universität Göttingen)

Flemming, Jens
Generalized Tikhonov regularization
Basic theory and comprehensive results on convergence rates
Dissertation, Fakultät für Mathematik
Technische Universität Chemnitz, Oktober 2011

Contents

Preface

Mathematical models of practical problems usually are designed to fit into well-known existing theory. At the same time new theoretic frameworks have to cope with criticism for lacking in practical relevance. To avoid such criticism, new theoretic results should come bundled with suggestions for improved mathematical models offered by the widened theory. Delivering such a bundle is one objective of this thesis.

In Part I we investigate ill-posed inverse problems formulated as operator equation in topological spaces. Such problems require regularization techniques for obtaining a stable approximate solution. Classical Tikhonov regularization allows for extensions to very general settings. We suggest one such extension and discuss its properties.

To fill the theoretic framework of Part I with life we consider a concrete inverse problem in Part II, which exploits the great freedom in modeling provided by the theory developed in the first part. Numerical examples show that alternative Tikhonov approaches yield improved regularized solutions in comparison with more classical Tikhonov-type methods.

Next to describing a general framework for Tikhonov-type regularization the emphasis of Part I lies on convergence rates theory. The sufficient condition for convergence rates proposed there is quite abstract and requires further explanation. Since there are several different formulations of such sufficient conditions in the literature, we embed the discussion into a wider context and present a number of cross connections between various conditions already known in the literature. This is the objective of Part III.

Some mathematical preliminaries are collected in the appendix. We recommend to have a look at the appendix and at the list of symbols and notations since the latter contains some remarks on notational conventions used throughout the thesis.

Only few weeks before the final version of this thesis was ready for printing the preprint [HW11] was published. It contains a very general framework for analyzing iteratively regularized Newton methods, which show some similarity with Tikhonov-type methods. Even though this preprint contains many interesting ideas which should be cited and commented in this thesis, we decided to leave it at a note in the preface. Changes and extensions in this late stage of the thesis would inevitably lead to mistakes and inconsistencies. The reader interested in this thesis should also have a look at the preprint [HW11] since the fundamental difficulties treated in the thesis (replacement for the triangle inequality, sufficient conditions for convergence rates) are handled in a different way.

This thesis was written under the supervision of Professor Bernd Hofmann (Chemnitz) during the last two years. The author thanks Bernd Hofmann for its constant support in mathematical as well as personal development and for the great freedom in research he offers his graduate students. Further, the author thanks Peter Mathé (Berlin) for pleasing collaboration, Radu Ioan Boţ (Chemnitz) for several hints concerning convex analysis, and Frank Werner (Göttingen) for discussions on regulariza-

tion with Poisson distributed data and his questions on logarithmic source conditions which motivated parts of the thesis. Research was partly supported by DFG under grant HO 1454/8-1.

Chemnitz, May 2011
Jens Flemming

Part I.

Theory of generalized Tikhonov regularization

1. Introduction

In this first part of the thesis we provide a comprehensive analysis of generalized Tikhonov regularization. We start with a short discussion of the problem to be analyzed in the sections of the present introductory chapter. Chapter 2 gives details on the setting and poses the basic assumptions on which all results of the subsequent chapters are based. We start the analysis of Tikhonov-type regularization methods in Chapter 3 and formulate the main result (Theorem 4.11) of this part of the thesis in Chapter 4. In the last chapter, Chapter 5, a second setting for Tikhonov-type regularization is presented.

The core of Chapters 2, 3, and 4 (without the section on the discrepancy principle) has already been published in [Fle10a].

1.1. Statement of the problem

Let (X, τ_X), (Y, τ_Y), and (Z, τ_Z) be Hausdorff spaces (see Definition A.9) and let $F : X \to Y$ be a mapping defined on X and taking values in Y. Justified by the considerations below we refer to X as the *solution space*, to Y as the *space of right-hand sides*, and to Z as the *data space*. Our aim is to approximate a solution of the ill-posed equation

$$F(x) = y, \quad x \in X, \tag{1.1}$$

with given right-hand side $y \in Y$ numerically. Here, ill-posedness means that the solutions do not depend continuously on the right-hand side. More precisely, if a sequence of right-hand sides converges with respect to τ_Y then a sequence of solutions of the corresponding equations need not converge with respect to τ_X. The popular definition of ill-posedness in the sense of Hadamard (see [EHN96, Chap. 2]) additionally includes the question of existence and uniqueness of solutions. In our setting existence of solutions will be assumed and uniqueness is not of interest since the developed theory is capable of handling multiple solutions.

Solving ill-posed equations numerically without making an effort to overcome the ill-posedness is impossible because even very small discretization or rounding errors can lead to arbitrarily large deviations of the calculated solution from the exact solution. In practice one also has to cope with the problem that the exact right-hand side y is unknown. Usually one only has some, often discrete, noisy measurement $z \in Z$ of y at hand. This lack of exact data can be dealt with by considering the minimization problem

$$S(F(x), z) \to \min_{x \in X} \tag{1.2}$$

instead of the original equation (1.1), where $S : Y \times Z \to [0, \infty]$ is some *fitting functional*, that is, $S(\tilde{y}, \tilde{z})$ should be the smaller the better a data element $\tilde{z} \in Z$ represents a right-

hand side $\tilde{y} \in Y$. Typical examples of such fitting functionals are given in Section 1.4 and a more ambitious one is proposed in Part II.

Because equation (1.1) is ill-posed, the minimization problem (1.2) typically is ill-posed, too. This means that the minimizers do not depend continuously on the data z. The idea of Tikhonov-type regularization methods is to stabilize the minimization problem by adding an appropriate *regularizing* or *stabilizing functional* $\Omega : X \to (-\infty, \infty]$. To control the influence of the stabilizing functional we introduce the *regularization parameter* $\alpha \in (0, \infty)$. Thus, we consider the minimization problem

$$T_\alpha^z(x) := S(F(x), z) + \alpha\Omega(x) \to \min_{x \in X}. \tag{1.3}$$

The functional T_α^z is referred to as *Tikhonov-type functional*. Its properties and the behavior of its minimizers are the core subject of the first part of this work.

1.2. Aims and scope

Concerning the minimization problem (1.3) there are four fundamental questions to be answered in the subsequent chapters:

Existence We have to guarantee that for each data element z and for each regularization parameter α there exist minimizers of T_α^z.

Stability Having in mind that (1.3) shall be solved numerically, well-posedness of (1.3) has to be shown. That is, small perturbations of the data z should not alter the minimizers too much. In addition we have to take into account that numerical minimization methods provide only an approximation of the true minimizer.

Convergence Because the minimizers of T_α^z are only approximate solutions of the underlying equation (1.1), we have to ensure that the approximation becomes the more exact the better the data z fits to the right-hand side y. This can be achieved by adapting the stability of (1.3) to the data, that is, we choose the regularization parameter α depending on z.

Convergence rates Convergence itself is more or less only of theoretic interest because in general it could be arbitrarily slow. For practical purposes we have to give convergence rates, that is, we have to bound the discrepancy between the minimizers in (1.3) and the exact solutions of (1.1) in terms of the misfit between the data z and the given right-hand side y.

Analogously to convergence rates, estimates for the stability of the minimization problem (1.3) are sometimes given in the literature (see, e.g., [SGG$^+$09, Theorem 3.46]). But since only numerical errors (which are typically small) are of interest there, stability estimates are not of such an importance as convergence rates, which describe the influence of possibly large measurement errors.

Concerning the choice of the regularization parameter there are two fundamental variants: *a priori choices* and *a posteriori choices*. In the first case the regularization parameter depends only on some noise bound, whereas a posteriori choices take into

account the concrete data element z. Usually one uses a priori choices to show which convergence rates can be obtained from the theoretical point of view. In practice, data dependent choices are applied. In this thesis we consider both variants.

To avoid confusion, we note that instead of *Tikhonov-type regularization* sometimes also the term *variational regularization* is given in the literature. But, more precisely, we use this term to denote the class of all non-iterative regularization methods. For information about *iterative regularization* we refer to the books [EHN96, KNS08].

1.3. Other approaches for generalized Tikhonov regularization

We want to mention four approaches for Tikhonov-type regularization which differ slightly from ours but are quite general, too. All four can be seen, more or less, as special cases of our setting.

- The authors of [JZ10] consider fitting functionals which implicitly contain the operator. With our notation and $Z := Y$ they set $\tilde{S}(x, y) := S(F(x), y)$ and formulate all assumptions and theorems with $\tilde{S}$ instead of S and F. Corrupting the interpretation of the spaces X, Y, and Z somewhat one sees that our approach is more general than the one in [JZ10]: we simply have to set $Y := X$ and F to be the identity on X, which results in the Tikhonov-type functional $S(x, z) + \alpha\Omega(x)$. The operator of the equation to be solved has to be contained in the fitting functional S.

- In [TLY98] a similar approach is chosen. The only difference is an additional function $f : [0, \infty] \to [0, \infty]$ which is applied to the fitting functional, that is, the authors use $f \circ S$ instead of S itself in the Tikhonov-type functional. Such a decomposition allows to fine-tune the assumptions on the fitting term, but besides this it provides no essential advantages.

- The present work has been inspired by the thesis [Pös08]. Thus, our approach is quite similar to the one considered there, except that we take into account a third space Z for the data. In [Pös08] only the case $Z := Y$ has been investigated. Another improvement is the weakening of the assumptions the fitting functional has to satisfy for proving convergence rates.

- In [Gei09] and [FH10] the setting coincides up to minor improvements with the one considered in [Pös08].

Especially the distinction between the space Y of right-hand sides and the space Z of data elements is considered for the first time in this thesis. On the one hand this additional feature forces us to develop new proofs instead of simply replacing norms by the fitting functional S in existing proofs, and thus encourages a deeper understanding of Tikhonov-type regularization methods. On the other hand the combination with general topological spaces (instead of using Banach or Hilbert spaces) allows to establish extended mathematical models for practical problems. The benefit of this new degree of freedom will be demonstrated in Part II of the thesis.

1.4. The standard examples

Investigation of Tikhonov-type regularization methods (1.3), at least for special cases, has been started in the nineteen sixties by Andrei Nikolaevich Tikhonov. The reader interested in the historical development of Tikhonov-type regularization may find information and references to the original papers in [TA76].

Analytic investigations mainly concentrated on two settings, which are described below. We refer to these two settings as the *standard Hilbert space setting* and the *standard Banach space setting*. They serve as illustrating examples several times in the sequel.

Example 1.1 (standard Hilbert space setting)**.** Let X and Y be Hilbert spaces and set $Z := Y$. The topologies τ_X and τ_Y shall be the corresponding weak topologies and τ_Z shall be the norm topology on $Y = Z$. Further assume that $A := F : X \to Y$ is a bounded linear operator. Setting $S(y_1, y_2) := \frac{1}{2}\|y_1 - y_2\|^2$ for $y_1, y_2 \in Y$ and $\Omega(x) := \frac{1}{2}\|x\|^2$ for $x \in X$ the objective function in (1.3) becomes the well-known *Tikhonov functional*

$$T_\alpha^{y^\delta}(x) = \tfrac{1}{2}\|Ax - y^\delta\|^2 + \tfrac{\alpha}{2}\|x\|^2$$

with $y^\delta \in Y$, $\delta \geq 0$, being some noisy measurement of the exact right-hand side y in (1.1) and satisfying

$$\|y^\delta - y\| \leq \delta.$$

The number $\delta \geq 0$ is called *noise level*. The distance between minimizers $x_\alpha^{y^\delta} \in X$ of $T_\alpha^{y^\delta}$ and a solution $x^\dagger \in X$ of (1.1) is typically expressed by $\|x_\alpha^{y^\delta} - x^\dagger\|$.

This special case of variational regularization has been extensively investigated and is well understood. For a detailed treatment we refer to [EHN96].

The Hilbert space setting has been extended in two ways: Instead of linear also nonlinear operators are considered and the Hilbert spaces may be replaced by Banach spaces.

Example 1.2 (standard Banach space setting)**.** Let X and Y be Banach spaces and set $Z := Y$. The topologies τ_X and τ_Y shall be the corresponding weak topologies and τ_Z shall be the norm topology on $Y = Z$. Further assume that $F : D(F) \subseteq X \to Y$ is a nonlinear operator with domain $D(F)$ and that the stabilizing functional Ω is convex. Note, that in Section 1.1 we assume $D(F) = X$. How to handle the situation $D(F) \subsetneq X$ within our setting is shown in Proposition 2.9. Setting $S(y_1, y_2) := \frac{1}{p}\|y_1 - y_2\|^p$ for $y_1, y_2 \in Y$ with $p \in (0, \infty)$ the objective function in (1.3) becomes

$$T_\alpha^{y^\delta}(x) = \tfrac{1}{p}\|F(x) - y^\delta\|^p + \alpha\Omega(x)$$

with $y^\delta \in Y$, $\delta \geq 0$, being some noisy measurement of the exact right-hand side y in (1.1) and satisfying

$$\|y^\delta - y\| \leq \delta.$$

The number $\delta \geq 0$ is called *noise level*. The distance between minimizers $x_\alpha^{y^\delta} \in X$ of $T_\alpha^{y^\delta}$ and a solution $x^\dagger \in X$ of (1.1) is typically expressed by the *Bregman distance*

$B^{\Omega}_{\xi^\dagger}(x_\alpha^{y^\delta}, x^\dagger)$ with respect to some subgradient $\xi^\dagger \in \partial\Omega(x^\dagger) \subseteq X^*$ (Bregman distances are introduced in Definition B.3).

Tikhonov-type regularization methods for mappings from a Banach into a Hilbert space in connection with Bregman distances have been considered for the first time in [BO04]. The same setting is analyzed in [Res05]. Further results on the standard Banach space setting may be found in [HKPS07] and also in [HH09, BH10, NHH$^+$10].

2. Assumptions

2.1. Basic assumptions and definitions

The following assumptions are fundamental for all the results in this part of the thesis and will be used in subsequent chapters without further notice. Within this chapter we explicitly indicate their usage to avoid confusion.

Assumption 2.1. The mapping $F : X \to Y$, the fitting functional $S : Y \times Z \to [0, \infty]$ and the stabilizing functional $\Omega : X \to (-\infty, \infty]$ have the following properties:

(i) F is sequentially continuous with respect to τ_X and τ_Y.

(ii) S is sequentially lower semi-continuous with respect to $\tau_Y \otimes \tau_Z$.

(iii) For each $y \in Y$ and each sequence $(z_k)_{k\in\mathbb{N}}$ in Z with $S(y, z_k) \to 0$ there exists some $z \in Z$ such that $z_k \to z$.

(iv) For each $y \in Y$, each $z \in Z$ satisfying $S(y, z) < \infty$, and each sequence $(z_k)_{k\in\mathbb{N}}$ in Z the convergence $z_k \to z$ implies $S(y, z_k) \to S(y, z)$.

(v) The sublevel sets $M_\Omega(c) := \{x \in X : \Omega(x) \leq c\}$ are sequentially compact with respect to τ_X for all $c \in \mathbb{R}$.

Remark 2.2. Equivalently to item (v) we could state that Ω is sequentially lower semi-continuous and that the sets $M_\Omega(c)$ are relatively sequentially compact.

Items (i), (ii), and (v) guarantee that the Tikhonov-type functional T_α^z in (1.3) is lower semi-continuous, which is an important prerequisite to show existence of minimizers. The desired stability of the minimization problem (1.3) comes from (v), and items (iii) and (iv) of Assumption 2.1 regulate the interplay of right-hand sides and data elements.

Remark 2.3. To avoid longish formulations, in the sequel we write 'closed', 'continuous', and so on instead of 'sequentially closed', 'sequentially continuous', and so on. Since we do not need the non-sequential versions of these topological terms confusion can be excluded.

An interesting question is, whether for given F, S, and Ω one can always find topologies τ_X, τ_Y, and τ_Z such that Assumption 2.1 is satisfied. Table 2.1 shows that for each of the three topologies there is at least one item in Assumption 2.1 bounding the topology below, that is, the topology must not be too weak, and at least one item bounding the topology above, that is, it must not be too strong. If for one topology the lower bound lies above the upper bound, then Assumption 2.1 cannot be satisfied by simply choosing the right topologies.

	(i)	(ii)	(iii)	(iv)	(v)
τ_X	$\mapsto$	$\bullet$	$\bullet$	$\bullet$	$\overset{\mapsto}{\underset{\mapsto}{}}$
τ_Y	$\mapsto$	$\mapsto$	$\bullet$	$\bullet$	$\bullet$
τ_Z	$\bullet$	$\mapsto$	$\mapsto$	$\mapsto$	$\bullet$

Table 2.1.: Lower and upper bounds for the topologies τ_X, τ_Y, and τ_Z ('$\mapsto$' means that the topology must not be too weak but can be arbitrarily strong to satisfy the corresponding item of Assumption 2.1 and '$\mapsto$' stands for the converse assertion, that is, the topology must not be too strong)

Unlike most other works on convergence theory of Tikhonov-type regularization methods we allow Ω to attain negative values. Thus, entropy functionals as an important class of stabilizing functionals are covered without modifications (see [EHN96, Section 5.3] for an introduction to maximum entropy methods). However, the assumptions on Ω guarantee that Ω is bounded below.

Proposition 2.4. *Let Assumption 2.1 be true. Then the stabilizing functional Ω is bounded below.*

Proof. Set $c := \inf_{x\in X} \Omega(x)$ and exclude the trivial case $c = \infty$. Then we find a sequence $(x_k)_{k\in\mathbb{N}}$ in X with $\Omega(x_k) \to c$ and thus, for sufficiently large k we have $\Omega(x_k) \leq c + 1$, which is equivalent to $x_k \in M_\Omega(c + 1)$. Therefore there exists a subsequence $(x_{k_l})_{l\in\mathbb{N}}$ converging to some $\tilde{x} \in X$ and the estimate $\Omega(\tilde{x}) \leq \liminf_{l\to\infty} \Omega(x_{k_l}) = c$ implies $c = \Omega(\tilde{x}) > -\infty$. □

The lower semi-continuity of the fitting functional S provides the following conclusion from Assumption 2.1 (iii).

Remark 2.5. Let Assumption 2.1 be satisfied. If for $y \in Y$ there exists a sequence $(z_k)_{k\in\mathbb{N}}$ with $S(y, z_k) \to 0$ then there exists a data element $z \in Z$ satisfying $S(y, z) = 0$.

Since in general $Y \neq Z$, the fitting functional S provides no direct way to check whether two elements $y_1 \in Y$ and $y_2 \in Y$ coincide. Thus, we have to introduce a notion of weak equality.

Definition 2.6. If for two elements $y_1, y_2 \in Y$ there exists some $z \in Z$ such that $S(y_1, z) = 0$ and $S(y_2, z) = 0$, we say that y_1 and y_2 are *S-equivalent with respect to z*. Further, we say that $x_1, x_2 \in X$ are S-equivalent if $F(x_1)$ and $F(x_2)$ are S-equivalent.

The notion of S-equivalence can be extended to a kind of distance on $Y \times Y$ induced by S.

Definition 2.7. We define the *distance* $S_Y : Y \times Y \to [0, \infty]$ *induced by* S via

$$S_Y(y_1, y_2) := \inf_{z\in Z} \big(S(y_1, z) + S(y_2, z)\big) \quad \text{for all } y_1, y_2 \in Y.$$

Proposition 2.8. *Let Assumption 2.1 be satisfied. Then the functional S_Y defined in Definition 2.7 is symmetric and $S_Y(y_1, y_2) = 0$ if and only if y_1 and y_2 are S-equivalent.*

Proof. We only show that y_1, y_2 are S-equivalent if $S_Y(y_1, y_2) = 0$ (the remaining assertions are obviously true). If $S_Y(y_1, y_2) = 0$, then there is a sequence $(z_k)_{k\in\mathbb{N}}$ such that $S(y_1, z_k) + S(y_2, z_k) \to 0$, in particular $S(y_1, z_k) \to 0$. Thus, by item (iii) of Assumption 2.1 there is some $\tilde{z} \in Z$ with $z_k \to \tilde{z}$ and therefore, by the lower semi-continuity of S, we have $S(y_1, \tilde{z}) = 0$ and $S(y_2, \tilde{z}) = 0$. This shows that y_1, y_2 are S-equivalent with respect to $\tilde{z}$. □

The functional S_Y plays an important role in Chapter 4 as part of a sufficient condition for convergence rates.

2.2. Special cases and examples

Sometimes it is favorable to consider a space X which is larger than the domain of the mapping F, for example to make X a Banach space. The following proposition shows how to reduce such a situation to the one described in this thesis.

Proposition 2.9. *Let $(\tilde{X}, \tau_{\tilde{X}})$ be a topological space and assume that the continuous mapping $\tilde{F} : D(\tilde{F}) \subseteq \tilde{X} \to Y$ has a closed domain $D(\tilde{F})$. Further, assume that $\tilde{\Omega} : \tilde{X} \to (-\infty, \infty]$ has compact sublevel sets. If we set $X := D(\tilde{F})$ and if τ_X is the topology on X induced by $\tau_{\tilde{X}}$ then the restrictions $F := \tilde{F}|_X$ and $\Omega := \tilde{\Omega}|_X$ satisfy items (i) and (v) of Assumption 2.1.*

Proof. Only the compactness of the sublevel sets of Ω needs a short comment: obviously $M_\Omega(c) = M_{\tilde{\Omega}}(c) \cap D(\tilde{F})$ and thus $M_\Omega(c)$ is closed as an intersection of closed sets. In addition we see $M_\Omega(c) \subseteq M_{\tilde{\Omega}}(c)$. Because closed subsets of compact sets are compact, $M_\Omega(c)$ is compact. □

Assumption 2.1 and Definition 2.6 can be simplified if the measured data lie in the same space as the right-hand sides. This is the standard assumption in inverse problems literature (see, e.g., [EHN96, SGG+09]) and with respect to non-metric fitting functionals this setting was already investigated in [Pös08].

Proposition 2.10. *Set $Z := Y$ and assume that $S : Y \times Y \to [0, \infty]$ satisfies the following properties:*

(i) Two elements $y_1, y_2 \in Y$ coincide if and only if $S(y_1, y_2) = 0$.

(ii) S is lower semi-continuous with respect to $\tau_Y \otimes \tau_Y$.

(iii) For each $y \in Y$ and each sequence $(y_k)_{k\in\mathbb{N}}$ the convergence $S(y, y_k) \to 0$ implies $y_k \to y$.

(iv) For each $y \in Y$, each $\tilde{y} \in Y$ satisfying $S(y, \tilde{y}) < \infty$, and each sequence $(y_k)_{k\in\mathbb{N}}$ in Y the convergence $S(\tilde{y}, y_k) \to 0$ implies $S(y, y_k) \to S(y, \tilde{y})$.

Then the following assertions are true:

- *S induces a topology on Y such that a sequence $(y_k)_{k\in\mathbb{N}}$ in Y converges to some $y \in Y$ with respect to this topology if and only if $S(y, y_k) \to 0$. This topology is stronger than τ_Y.*

- *Defining τ_Z to be the topology on $Z = Y$ induced by S, items (ii), (iii), and (iv) of Assumption 2.1 are satisfied.*
- *Two elements $y_1, y_2 \in Y$ are S-equivalent if and only if $y_1 = y_2$.*

Proof. We define the topology

$$\tau_S := \Big\{ G \subseteq Y : \big(y \in G,\ y_k \in Y,\ S(y, y_k) \to 0\big) \Rightarrow \big(\exists k_0 \in \mathbb{N} : y_k \in G\ \forall k \geq k_0\big) \Big\}$$

to be the family of all subsets of Y consisting only of 'interior points'. This is indeed a topology: Obviously $\emptyset \in \tau_S$ and $Y \in \tau_S$. For $G_1, G_2 \in \tau_S$ we have

$$\begin{aligned} & y \in G_1 \cap G_2,\ y_k \in Y,\ S(y, y_k) \to 0 \\ \Rightarrow\quad & \exists k_1, k_2 \in \mathbb{N} : y_k \in G\ \forall k \geq k_1,\ y_k \in G_2\ \forall k \geq k_2 \\ \Rightarrow\quad & y_k \in G_1 \cap G_2\ \forall k \geq k_0 := \max\{k_1, k_2\}, \end{aligned}$$

that is, $G_1 \cap G_2 \in \tau_S$. In a similar way we find that τ_S is closed under infinite unions.

We now prove that $y_k \overset{\tau_S}{\to} y$ if and only if $S(y, y_k) \to 0$. First, note that $y_k \overset{\tau}{\to} y$ with some topology τ on Y holds if and only if for each $G \in \tau$ with $y \in G$ also $y_k \in G$ is true for sufficiently large k. From this observation we immediately get $y_k \overset{\tau_S}{\to} y$ if $S(y, y_k) \to 0$. Now assume that $y_k \overset{\tau_S}{\to} y$ and set

$$M_\varepsilon(y) := \{\tilde{y} \in Y : S(y, \tilde{y}) < \varepsilon\}$$

for $\varepsilon > 0$. Because

$$\begin{aligned} & \tilde{y} \in M_\varepsilon(y),\ \tilde{y}_k \in Y,\ S(\tilde{y}, \tilde{y}_k) \to 0 \\ \Rightarrow\quad & S(y, \tilde{y}_k) \to S(y, \tilde{y}) < \varepsilon \quad \text{(Assumption (iv))} \\ \Rightarrow\quad & \exists k_0 \in \mathbb{N} : S(y, \tilde{y}_k) < \varepsilon\ \forall k \geq k_0 \\ \Rightarrow\quad & \tilde{y}_k \in M_\varepsilon(y)\ \forall k \geq k_0 \end{aligned}$$

we have $M_\varepsilon(y) \in \tau_S$. By Assumption (i) we know $y \in M_\varepsilon(y)$ and therefore $y_k \overset{\tau_S}{\to} y$ implies $y_k \in M_\varepsilon(y)$ for all sufficiently large k. Thus, $S(y, y_k) < \varepsilon$ for all sufficiently large k, that is, $S(y, y_k) \to 0$.

The next step is to show that τ_S is stronger than τ_Y. So let $G \in \tau_Y$. Then we have

$$\begin{aligned} & y \in G,\ y_k \in Y,\ S(y, y_k) \to 0 \\ \Rightarrow\quad & y_k \to y \quad \text{(Assumption (iii))} \\ \Rightarrow\quad & \exists k_0 \in \mathbb{N} : y_k \in G\ \forall k \geq k_0, \end{aligned}$$

which is exactly the property in the definition of τ_S, that is, $G \in \tau_S$.

Now set $Z := Y$ and $\tau_Z := \tau_S$. Then item (ii) of Assumption 2.1 follows from Assumption (ii) because τ_Z is stronger than τ_Y. With $z := y$ Assumption 2.1 (iii) is an immediate consequence of Assumption (iii). And (iv) of Assumption 2.1 is simply a reformulation of Assumption (iv).

The third and last assertion of the proposition is quite obvious: If y_1 and y_2 are S-equivalent with respect to some $z \in Z$ then by Assumption (i) we know $y_1 = z = y_2$. Conversely, if $y_1 = y_2$ then y_1 and y_2 are S-equivalent with respect to $z := y_1$. □

Eventually, we show that the standard settings for Tikhonov-type regularization described in Example 1.1 and in Example 1.2 fit well into our general framework.

Proposition 2.11. *Let $Y = Z$ be a normed vector space and let τ_Y be the corresponding weak topology. Then the fitting functional $S : Y \times Y \to [0, \infty)$ defined by*

$$S(y_1, y_2) := \tfrac{1}{p}\|y_1 - y_2\|^p$$

with $p > 0$ satisfies the assumptions of Proposition 2.10.

Proof. Item (i) of Propostion 2.10 is trivial, the lower semi-continuity of S follows from [BP86, Chapter 1, Corollary 2.5], and the remaining two assumptions of Propostion 2.10 are obviously satisfied. □

Proposition 2.12. *Let X be a normed vector space and let τ_X be the corresponding weak topology. Then the stabilizing functional $\Omega : X \to [0, \infty)$ defined by*

$$\Omega(x) := \tfrac{1}{q}\|x - \bar{x}\|^q$$

with $q > 0$ and fixed $\bar{x} \in X$ is weakly lower semi-continuous and its sublevel sets are relatively weakly compact if and only if X is reflexive.

Proof. The weak lower semi-continuity of Ω is a consequence of [BP86, Chapter 1, Corollary 2.5] and the assertion about the sublevel sets of Ω follows immediately from [BP86, Chapter 1, Theorem 2.4]. □

The setting of Example 1.1 is obtained from Propositions 2.11 and 2.12 by setting $p = 2$ and $q = 2$.

If the fitting functional S is the power of a norm we can give an explicit formula for the distance S_Y induced by S.

Proposition 2.13. *Let $Y = Z$ be a normed vector space and define the fitting functional $S : Y \times Y \to [0, \infty)$ by*

$$S(y_1, y_2) := \tfrac{1}{p}\|y_1 - y_2\|^p$$

for some $p > 0$. Then the corresponding distance $S_Y : Y \times Y \to [0, \infty]$ defined in Definition 2.7 reads as

$$S_Y(y_1, y_2) = \frac{c}{p}\|y_1 - y_2\|^p \quad \text{for all } y_1, y_2 \in Y$$

with $c := \min\{1, 2^{1-p}\}$.

Proof. Applying the triangle inequality and the inequality

$$(a + b)^p \leq \max\{1, 2^{p-1}\}(a^p + b^p) \quad \text{for all } a, b \geq 0$$

(cf. [SGG+09, Lemma 3.20]) gives $\frac{c}{p}\|y_1 - y_2\|^p \leq S(y_1, y) + S(y_2, y)$ for all $y \in Y = Z$, that is, $\frac{c}{p}\|y_1 - y_2\|^p \leq S_Y(y_1, y_2)$. And setting $y := \frac{1}{2}(y_1 + y_2)$ for $p \geq 1$ and $y := y_1$ for $p < 1$ proves equality. □

3. Fundamental properties of Tikhonov-type minimization problems

In this chapter we prove that the minimization problem (1.3) has solutions, that these solutions are stable with respect to the data z and with respect to inexact minimization, and that problem (1.3) provides arbitrarily accurate approximations of the solutions to (1.1). In addition, we provide some remarks on the discretization of (1.3). Throughout this chapter we assume that Assumption 2.1 is satisfied.

The proofs of the three main theorems in this chapter use standard techniques and are quite similar to the corresponding proofs given in [HKPS07] or [Pös08].

3.1. Generalized solutions

Since in general $Y \neq Z$, it may happen that the data elements z contain not enough information about the exact right-hand side y to reconstruct a solution of (1.1) from the measured data. Therefore we can only expect to find some $x \in X$ such that $F(x)$ is S-equivalent to y (cf. Definition 2.6). We denote such x as *S-generalized solutions*. Obviously, each true solution is also an S-generalized solution if there is some $z \in Z$ such that $S(y, z) = 0$.

As we will see in Section 3.4, the minimizers of (1.3) are approximations of a specific type of S-generalized solutions, so called *Ω-minimizing S-generalized solutions*, which are introduced by the following proposition.

Proposition 3.1. *If for $y \in Y$ there exists an S-generalized solution $\bar{x} \in X$ of (1.1) with $\Omega(\bar{x}) < \infty$, then there exists an Ω-minimizing S-generalized solution of (1.1), that is, there exists an S-generalized solution $x^\dagger \in X$ which satisfies*

$$\Omega(x^\dagger) = \inf\{\Omega(x) : x \in X,\ F(x) \text{ is } S\text{-equivalent to } y\}.$$

Proof. Set $c := \inf\{\Omega(x) : x \in X,\ F(x)$ is S-equivalent to $y\}$ and take a sequence $(x_k)_{k\in\mathbb{N}}$ in X such that $F(x_k)$ is S-equivalent to y and $\Omega(x_k) \to c$. Because the sublevel sets of Ω are compact and $\Omega(x_k) \leq c + 1$ for sufficiently large k, there is a subsequence $(x_{k_l})_{l\in\mathbb{N}}$ of (x_k) converging to some $\tilde{x} \in X$. The continuity of F implies $F(x_{k_l}) \to F(\tilde{x})$.

Now take a sequence $(z_k)_{k\in\mathbb{N}}$ in Z such that $F(x_k)$ is S-equivalent to y with respect to z_k, that is, $S(F(x_k), z_k) = 0$ and $S(y, z_k) = 0$. Because $S(y, z_k) \to 0$, we find some $z \in Z$ with $z_k \to z$. The lower semi-continuity of S implies $S(y, z) = 0$ and $S(F(\tilde{x}), z) \leq \liminf_{l\to\infty} S(F(x_{k_l}), z_{k_l}) = 0$. Thus, $\tilde{x}$ is an S-generalized solution. The estimate

$$\Omega(\tilde{x}) \leq \liminf_{l\to\infty} \Omega(x_{k_l}) = \lim_{l\to\infty} \Omega(x_{k_l}) = c$$

shows that $\tilde{x}$ is Ω-minimizing. □

3.2. Existence of minimizers

The following result on the existence of minimizers of the Tikhonov-type functional T_α^z is sometimes denoted as 'well-posedness' of (1.3) and the term 'existence' is used for the assertion of Proposition 3.1 (see, e.g., [HKPS07, Pös08]). Since well-posedness for us means the opposite of ill-posedness in the sense described in Section 1.1, we avoid using this term for results not directly connected to the ill-posedness phenomenon.

Theorem 3.2 (existence)**.** *For all $z \in Z$ and all $\alpha \in (0, \infty)$ the minimization problem (1.3) has a solution. A minimizer $x^* \in X$ satisfies $T_\alpha^z(x^*) < \infty$ if and only if there exists an element $\bar{x} \in X$ with $S(F(\bar{x}), z) < \infty$ and $\Omega(\bar{x}) < \infty$.*

Proof. Set $c := \inf_{x \in X} T_\alpha^z(x)$ and take a sequence $(x_k)_{k \in \mathbb{N}}$ in X with $T_\alpha^z(x_k) \to c$. To avoid trivialities we exclude the case $c = \infty$, which occurs if and only if there is no $\bar{x} \in X$ with $S(F(\bar{x}), z) < \infty$ and $\Omega(\bar{x}) < \infty$. Then

$$\Omega(x_k) \leq \frac{1}{\alpha} T_\alpha^z(x_k) \leq \frac{1}{\alpha}(c+1)$$

for sufficiently large k and by the compactness of the sublevel sets of Ω there is a subsequence $(x_{k_l})_{l \in \mathbb{N}}$ converging to some $\tilde{x} \in X$. The continuity of F implies $F(x_{k_l}) \to F(\tilde{x})$ and the lower semi-continuity of S and Ω gives

$$T_\alpha^z(\tilde{x}) \leq \liminf_{l \to \infty} T_\alpha^z(x_k) = c,$$

that is, $\tilde{x}$ is a minimizer of T_α^z. □

3.3. Stability of the minimizers

For the numerical treatment of the minimization problem (1.3) it is of fundamental importance that the minimizers are not significantly affected by numerical inaccuracies. Examples for such inaccuracies are discretization and rounding errors. In addition, numerical minimization procedures do not give real minimizers of the objective function; instead, they provide an element at which the objective function is only very close to its minimal value. The following theorem states that (1.3) is stable in this sense. More precisely, we show that reducing numerical inaccuracies yields arbitrarily exact approximations of the true minimizers.

Theorem 3.3 (stability)**.** *Let $z \in Z$ and $\alpha \in (0, \infty)$ be fixed and assume that $(z_k)_{k \in \mathbb{N}}$ is a sequence in Z satisfying $z_k \to z$, that $(\alpha_k)_{k \in \mathbb{N}}$ is a sequence in $(0, \infty)$ converging to α, and that $(\varepsilon_k)_{k \in \mathbb{N}}$ is a sequence in $[0, \infty)$ converging to zero. Further assume that there exists an element $\bar{x} \in X$ with $S(F(\bar{x}), z) < \infty$ and $\Omega(\bar{x}) < \infty$.*

Then each sequence $(x_k)_{k \in \mathbb{N}}$ with $T_{\alpha_k}^{z_k}(x_k) \leq \inf_{x \in X} T_{\alpha_k}^{z_k}(x) + \varepsilon_k$ has a τ_X-convergent subsequence and for sufficiently large k the elements x_k satisfy $T_{\alpha_k}^{z_k}(x_k) < \infty$. Each limit $\tilde{x} \in X$ of a τ_X-convergent subsequence $(x_{k_l})_{l \in \mathbb{N}}$ is a minimizer of T_α^z and we have $T_{\alpha_{k_l}}^{z_{k_l}}(x_{k_l}) \to T_\alpha^z(\tilde{x})$, $\Omega(x_{k_l}) \to \Omega(\tilde{x})$ and thus also $S(F(x_{k_l}), z_{k_l}) \to S(F(\tilde{x}), z)$. If T_α^z has only one minimizer $\hat{x}$, then (x_k) converges to $\hat{x}$.

Proof. The convergence $z_k \to z$ together with $S(F(\bar{x}), z) < \infty$ implies $S(F(\bar{x}), z_k) \to S(F(\bar{x}), z)$, that is, $S(F(\bar{x}), z_k) < \infty$ for sufficiently large k. Thus, without loss of generality we assume $S(F(\bar{x}), z_k) < \infty$ for all $k \in \mathbb{N}$. From Theorem 3.2 we obtain the existence of minimizers $x_k^* \in \operatorname{argmin}_{x \in X} T_{\alpha_k}^{z_k}(x)$ and that $T_{\alpha_k}^{z_k}(x_k^*) < \infty$ for all $k \in \mathbb{N}$. Further, $S(F(\bar{x}), z_k) \to S(F(\bar{x}), z)$ implies

$$\begin{aligned}\Omega(x_k) &\le \frac{1}{\alpha_k} T_{\alpha_k}^{z_k}(x_k) \le \frac{1}{\alpha_k} T_{\alpha_k}^{z_k}(x_k^*) + \frac{\varepsilon_k}{\alpha_k} \le \frac{1}{\alpha_k} T_{\alpha_k}^{z_k}(\bar{x}) + \frac{\varepsilon_k}{\alpha_k} \\ &= \frac{1}{\alpha_k} S(F(\bar{x}), z_k) + \Omega(\bar{x}) + \frac{\varepsilon_k}{\alpha_k} \le \frac{2}{\alpha}\big(S(F(\bar{x}), z) + 1\big) + \Omega(\bar{x}) + \frac{2}{\alpha} \sup_{l \in \mathbb{N}} \varepsilon_l =: c_\Omega\end{aligned}$$

for sufficiently large k and therefore, by the compactness of the sublevel sets of Ω, the sequence (x_k) has a τ_X-convergent subsequence.

Now let $(x_{k_l})_{l \in \mathbb{N}}$ be an arbitrary τ_X-convergent subsequence of (x_k) and let $\tilde{x} \in X$ be the limit of (x_{k_l}). Then for all $x_\alpha^z \in \operatorname{argmin}_{x \in X} T_\alpha^z(x)$ Theorem 3.2 shows $S(F(x_\alpha^z), z) < \infty$ and $\Omega(x_\alpha^z) < \infty$, and thus we get

$$\begin{aligned}T_\alpha^z(\tilde{x}) &\le \liminf_{l \to \infty} T_\alpha^{z_{k_l}}(x_{k_l}) \le \limsup_{l \to \infty} T_\alpha^{z_{k_l}}(x_{k_l}) = \limsup_{l \to \infty} \big(T_{\alpha_{k_l}}^{z_{k_l}}(x_{k_l}) + (\alpha - \alpha_{k_l})\Omega(x_{k_l})\big) \\ &\le \limsup_{l \to \infty} \big(T_{\alpha_{k_l}}^{z_{k_l}}(x_{k_l}^*) + \varepsilon_{k_l} + |\alpha - \alpha_{k_l}||\Omega(x_{k_l})|\big) \\ &\le \limsup_{l \to \infty} \big(T_{\alpha_{k_l}}^{z_{k_l}}(x_\alpha^z) + \varepsilon_{k_l} + |\alpha - \alpha_{k_l}||c_\Omega|\big) \\ &= \lim_{l \to \infty} \big(S(F(x_\alpha^z), z_{k_l}) + \alpha_{k_l}\Omega(x_\alpha^z) + \varepsilon_{k_l} + |\alpha - \alpha_{k_l}||c_\Omega|\big) = T_\alpha^z(x_\alpha^z),\end{aligned}$$

that is, $\tilde{x}$ minimizes T_α^z. In addition, with $x_\alpha^z = \tilde{x}$, we obtain

$$T_\alpha^z(\tilde{x}) = \lim_{l \to \infty} T_{\alpha_{k_l}}^{z_{k_l}}(x_{k_l}).$$

Assume $\Omega(x_{k_l}) \nrightarrow \Omega(\tilde{x})$. Then

$$\Omega(\tilde{x}) \ne \liminf_{l \to \infty} \Omega(x_{k_l}) \quad \text{or} \quad \Omega(\tilde{x}) \ne \limsup_{l \to \infty} \Omega(x_{k_l})$$

must hold, which implies

$$c := \limsup_{l \to \infty} \Omega(x_{k_l}) > \Omega(\tilde{x}).$$

If $(x_{l_m})_{m \in \mathbb{N}}$ is a subsequence of (x_{k_l}) with $\Omega(x_{l_m}) \to c$ we get

$$\begin{aligned}\lim_{m \to \infty} S(F(x_{l_m}), z_{l_m}) &= \lim_{m \to \infty} T_{\alpha_{l_m}}^{z_{l_m}}(x_{l_m}) - \lim_{m \to \infty} \big(\alpha_{l_m}\Omega(x_{l_m})\big) = T_\alpha^z(\tilde{x}) - \alpha c \\ &= S(F(\tilde{x}), z) - \alpha(c - \Omega(\tilde{x})) < S(F(\tilde{x}), z),\end{aligned}$$

which contradicts the lower semi-continuity of S.

Finally, assume that T_α^z has only one minimizer $\hat{x}$. Since $(x_k)_{k \in \mathbb{N}}$ has a convergent subsequence and each convergent subsequence has the limit $\hat{x}$, by Proposition A.27 the whole sequence (x_k) converges to $\hat{x}$. ∎

3.4. Convergence to generalized solutions

Now, that we know that (1.3) can be solved numerically, we have to prove that the minimizers of (1.3) are approximations of the solutions to (1.1). In fact, we show that the better a data element z represents the exact right-hand side y the more accurate the approximation. To achieve this, the regularization parameter α has to be chosen appropriately.

Theorem 3.4 (convergence)**.** *Let $y \in Y$ and let $(z_k)_{k\in\mathbb{N}}$ be a sequence in Z satisfying $S(y, z_k) \to 0$. Further, let $(\alpha_k)_{k\in\mathbb{N}}$ be a sequence in $(0,\infty)$ and let $(x_k)_{k\in\mathbb{N}}$ be a sequence in X with $x_k \in \operatorname{argmin}_{x\in X} T_{\alpha_k}^{z_k}(x)$.*

If the choice of (α_k) guarantees $\Omega(x_k) \le c$ for some $c \in \mathbb{R}$ and sufficiently large k and also $S(F(x_k), z_k) \to 0$, then (x_k) has a τ_X-convergent subsequence and each limit of a τ_X-convergent subsequence $(x_{k_l})_{l\in\mathbb{N}}$ is an S-generalized solution of (1.1).

If in addition $\limsup_{k\to\infty} \Omega(x_k) \le \Omega(\hat{x})$ for all S-generalized solutions $\hat{x} \in X$ of (1.1), then each such limit $\tilde{x}$ is an Ω-minimizing S-generalized solution and $\Omega(\tilde{x}) = \lim_{l\to\infty} \Omega(x_{k_l})$. If (1.1) has only one Ω-minimizing S-generalized solution $x^\dagger$, then (x_k) has the limit $x^\dagger$.

Proof. The existence of a τ_X-convergent subsequence of (x_k) is guaranteed by $\Omega(x_k) \le c$ for sufficiently large k. Let $(x_{k_l})_{l\in\mathbb{N}}$ be an arbitrary subsequence of (x_k) converging to some element $\tilde{x} \in X$. Since $S(y, z_k) \to 0$, there exists some $z \in Z$ with $z_k \to z$, and the lower semi-continuity of S together with the continuity of F implies

$$S(F(\tilde{x}), z) \le \liminf_{l\to\infty} S(F(x_{k_l}), z_{k_l}) = 0,$$

that is, $S(F(\tilde{x}), z) = 0$. Thus, from $S(y, z) \le \liminf_{k\to\infty} S(y, z_k) = 0$ we see that $\tilde{x}$ is an S-generalized solution of (1.1).

If $\limsup_{k\to\infty} \Omega(x_k) \le \Omega(\hat{x})$ for all S-generalized solutions $\hat{x} \in X$, then for each S-generalized solution $\hat{x}$ we get

$$\Omega(\tilde{x}) \le \liminf_{l\to\infty} \Omega(x_{k_l}) \le \limsup_{l\to\infty} \Omega(x_{k_l}) \le \Omega(\hat{x}).$$

This estimate shows that $\tilde{x}$ is an Ω-minimizing S-generalized solution, and setting $\hat{x} := \tilde{x}$ we find $\Omega(\tilde{x}) = \lim_{l\to\infty} \Omega(x_{k_l})$.

Finally, assume there is only one Ω-minimizing S-generalized solution $x^\dagger$. Then for each limit $\tilde{x}$ of a convergent subsequence of (x_k) we have $\tilde{x} = x^\dagger$ and therefore, by Propostion A.27 the whole sequence (x_k) converges to $x^\dagger$. □

Having a look at Proposition 4.15 in Section 4.3 we see that the conditions $\Omega(x_k) \le c$ and $\limsup_{k\to\infty} \Omega(x_k) \le \Omega(\hat{x})$ in Theorem 3.4 constitute a lower bound for α_k. On the other hand, this proposition suggests that $S(F(x_k), z_k) \to 0$ bounds α_k from above.

Remark 3.5. Since it is not obvious that a sequence $(\alpha_k)_{k\in\mathbb{N}}$ satisfying the assumptions of Theorem 3.4 exists, we have to show that there is always such a sequence: Let y, $(z_k)_{k\in\mathbb{N}}$, and $(x_k)_{k\in\mathbb{N}}$ be as in Theorem 3.4 and assume that there exists an S-generalized solution $\bar{x} \in X$ of (1.1) with $\Omega(\bar{x}) < \infty$ and $S(F(\bar{x}), z_k) \to 0$. If we set

$$\alpha_k := \begin{cases} \frac{1}{k}, & \text{if } S(F(\bar{x}), z_k) = 0, \\ \sqrt{S(F(\bar{x}), z_k)}, & \text{if } S(F(\bar{x}), z_k) > 0, \end{cases}$$

then

$$\begin{aligned} S(F(x_k), z_k) &= T_{\alpha_k}^{z_k}(x_k) - \alpha_k \Omega(x_k) \leq T_{\alpha_k}^{z_k}(\bar{x}) - \alpha_k \Omega(x_k) \\ &= S(F(\bar{x}), z_k) + \alpha_k \big(\Omega(\bar{x}) - \Omega(x_k)\big) \\ &\leq S(F(\bar{x}), z_k) + \alpha_k \big(\Omega(\bar{x}) - \inf_{x \in X} \Omega(x)\big) \to 0 \end{aligned}$$

and

$$\Omega(x_k) \leq \frac{1}{\alpha_k} T_{\alpha_k}^{z_k}(x_k) \leq \frac{1}{\alpha_k} T_{\alpha_k}^{z_k}(\bar{x}) = \Omega(\bar{x}) + \frac{S(F(\bar{x}), z_k)}{\alpha_k} \to \Omega(\bar{x}).$$

An a priori parameter choice satisfying the assumptions of Theorem 3.4 is given in Corollary 4.2 and an a posteriori parameter choice applicable in practice is the subject of Section 4.3 (see Corollary 4.22 for the verification of the assumptions on (α_k)).

3.5. Discretization

In this section we briefly discuss how to discretize the minimization problem (1.3). The basic ideas for appropriately extending Assumption 2.1 are taken from [PRS05], where discretization of Tikhonov-type methods in nonseparable Banach spaces is discussed.

Note that we do not discretize the spaces Y and Z here. The space Y of right-hand sides is only of analytic interest and the data space Z is allowed to be finite-dimensional in the setting investigated in the previous chapters. Only the elements of the space X, which in general is infinite-dimensional in applications, have to be approximated by elements from finite-dimensional spaces. Therefore let $(X_n)_{n \in \mathbb{N}}$ be an increasing sequence of τ_X-closed subspaces $X_1 \subseteq X_2 \subseteq \ldots \subseteq X$ of X equipped with the topology induced by τ_X. Typically, the X_n are finite-dimensional spaces. Thus, the minimization of T_α^z over X_n can be carried out numerically.

The main result of this section is based on the following assumption.

Assumption 3.6. For each $x \in X$ there is a sequence $(x_n)_{n \in \mathbb{N}}$ with $x_n \in X_n$ such that $S(F(x_n), z) \to S(F(x), z)$ for all $z \in Z$ and $\Omega(x_n) \to \Omega(x)$.

This assumption is rather technical. The next assumption formulates a set of sufficient conditions which are more accessible.

Assumption 3.7. Let τ_X^+ and τ_Y^+ be topologies on X and Y, respectively, and assume that the following items are satisfied:

(i) $S(\bullet, z) : Y \to [0, \infty]$ is continuous with respect to τ_Y^+ for all $z \in Z$.

(ii) Ω is continuous with respect to τ_X^+.

(iii) F is continuous with respect to τ_X^+ and τ_Y^+.

(iv) The union of (X_n) is dense in X with respect to τ_X^+, that is,

$$\overline{\bigcup_{n \in \mathbb{N}} X_n}^{\tau_X^+} = X.$$

Obviously Assumption 3.6 is a consequence of Assumption 3.7.

Remark 3.8. Note that we always find topologies τ_X^+ and τ_Y^+ such that items (i), (ii), and (iii) of Assumption 3.7 are satisfied. Usually, these topologies will be stronger than τ_X and τ_Y (indicated by the '+' sign). The crucial limitation is item (iv). Especially if X is a nonseparable space one has to be very careful in choosing appropriate topologies τ_X^+ and τ_Y^+ and suitable subspaces X_n. For concrete examples we refer to [PRS05].

Remark 3.9. As noted above we assume that the subspaces X_n are τ_X-closed. Therefore Assumption 2.1 is still satisfied if X is replaced by X_n (with the topology induced by τ_X). Consequently, the results on existence (Theorem 3.2) and stability (Theorem 3.3) also apply if T_α^z in (1.3) is minimized over X_n instead of X. For the sake of completeness we mention that we can guarantee the existence of elements $\bar{x}_n \in X_n$ with $S(F(\bar{x}_n), z) < \infty$ and $\Omega(\bar{x}_n) < \infty$ if there is $\bar{x} \in X$ with $S(F(\bar{x}), z) < \infty$ and $\Omega(\bar{x}) < \infty$ and if n is sufficiently large. This is an immediate consequence of Assumption 3.6.

The following corollary of Theorem 3.3 (stability) states that the minimizers of T_α^z over X can be approximated arbitrarily exact by the minimizers of T_α^z over X_n if n is chosen large enough.

Corollary 3.10. *Let Assumption 3.6 by satisfied, let $z \in Z$ and $\alpha \in (0, \infty)$ be fixed, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence of minimizers $x_n \in \operatorname{argmin}_{x\in X_n} T_\alpha^z(x)$. Further assume that there exists an element $\bar{x} \in X$ with $S(F(\bar{x}), z) < \infty$ and $\Omega(\bar{x}) < \infty$.*

Then (x_n) has a τ_X-convergent subsequence and for sufficiently large n the elements x_n satisfy $T_\alpha^z(x_n) < \infty$. Each limit $\tilde{x} \in X$ of a τ_X-convergent subsequence $(x_{n_l})_{l\in\mathbb{N}}$ is a minimizer of T_α^z over X and we have $T_\alpha^z(x_{n_l}) \to T_\alpha^z(\tilde{x})$, $\Omega(x_{n_l}) \to \Omega(\tilde{x})$, and thus also $S(F(x_{n_l}), z) \to S(F(\tilde{x}), z)$. If T_α^z has only one minimizer $\hat{x}$, then (x_n) converges to $\hat{x}$.

Proof. From Theorem 3.2 in combination with Remark 3.9 we obtain the existence of minimizers $x_n \in \operatorname{argmin}_{x\in X_n} T_\alpha^z(x)$ and that $T_\alpha^z(x_n) < \infty$ for sufficiently large $n \in \mathbb{N}$. We verify the assumptions of Theorem 3.3 for $k := n$, $z_n := z$, $\alpha_n := \alpha$, and $\varepsilon_n := T_\alpha^z(x_n) - T_\alpha^z(x^*)$ with some minimizer $x^* \in \operatorname{argmin}_{x\in X} T_\alpha^z(x)$. The only thing we have to show is $\varepsilon_n \to 0$. For this purpose take a sequence $(x_n^*)_{n\in\mathbb{N}}$ of approximations $x_n^* \in X_n$ such that $S(F(x_n^*), z) \to S(F(x^*), z)$ and $\Omega(x_n^*) \to \Omega(x^*)$ (cf. Assumption 3.6). Then

$$0 \le \varepsilon_n = T_\alpha^z(x_n) - T_\alpha^z(x^*) \le T_\alpha^z(x_n^*) - T_\alpha^z(x^*) \to 0.$$

Now all assertions follow from Theorem 3.3. □

4. Convergence rates

We consider equation (1.1) with fixed right-hand side $y := y^0 \in Y$. By $x^\dagger \in X$ we denote one fixed Ω-minimizing S-generalized solution of (1.1), where we assume that there exists an S-generalized solution $\bar{x} \in X$ with $\Omega(\bar{x}) < \infty$ (then Proposition 3.1 guarantees the existence of Ω-minimizing S-generalized solutions).

4.1. Error model

Convergence rate results describe the dependence of the *solution error* on the *data error* if the data error is small. So at first we have to decide how to measure these errors. For this purpose we introduce a functional $D_{y^0} : Z \to [0, \infty]$ measuring the distance between the right-hand side y^0 and a data element $z \in Z$. On the solution space X we introduce a functional $E_{x^\dagger} : X \to [0, \infty]$ measuring the distance between the Ω-minimizing S-generalized solution $x^\dagger$ and an approximate solution $x \in X$. We want to obtain bounds for the solution error $E_{x^\dagger}(x_\alpha^z)$ in terms of $D_{y^0}(z)$, where z is the given data and x_α^z is a corresponding minimizer of the Tikhonov-type functional (1.3).

4.1.1. Handling the data error

In practice we do not know the exact data error $D_{y^0}(z)$, but at least we should have some upper bound at hand, the so called *noise level* $\delta \in [0, \infty)$. Given δ, by

$$Z_{y^0}^\delta := \{z \in Z : D_{y^0}(z) \leq \delta\}$$

we denote the set of all data elements adhering to the noise level δ. To guarantee $Z_{y^0}^\delta \neq \emptyset$ for each $\delta \geq 0$, we assume that there is some $z \in Z$ with $D_{y^0}(z) = 0$.

Of course, we have to connect D_{y^0} to the Tikhonov-type functional (1.3) to obtain any useful result on the influence of data errors on the minimizers of the functional. This connection is established by the following assumption, which we assume to hold throughout this chapter.

Assumption 4.1. There exists a monotonically increasing function $\psi : [0, \infty) \to [0, \infty)$ satisfying $\psi(\delta) \to 0$ if $\delta \to 0$, $\psi(\delta) = 0$ if and only if $\delta = 0$, and

$$S(\tilde{y}, z) \leq \psi(D_{y^0}(z))$$

for all $\tilde{y} \in Y$ which are S-equivalent to y^0 and for all $z \in Z$ with $D_{y^0}(z) < \infty$.

This assumption provides the estimate

$$S(F(x), z^\delta) \leq \psi(D_{y^0}(z^\delta)) \leq \psi(\delta) < \infty \tag{4.1}$$

for all S-generalized solutions x of (1.1) and for all $z^\delta \in Z^\delta_{y^0}$. Thus, we obtain the following version of Theorem 3.4 with an a priori parameter choice (that is, the regularization parameter α does not depend on the concrete data element z but only on the noise level δ).

Corollary 4.2. *Let $(\delta_k)_{k\in\mathbb{N}}$ be a sequence in $[0,\infty)$ converging to zero, take an arbitrary sequence $(z_k)_{k\in\mathbb{N}}$ with $z_k \in Z^{\delta_k}_{y^0}$, and choose a sequence $(\alpha_k)_{k\in\mathbb{N}}$ in $(0,\infty)$ such that $\alpha_k \to 0$ and $\frac{\psi(\delta_k)}{\alpha_k} \to 0$. Then for each sequence $(x_k)_{k\in\mathbb{N}}$ with $x_k \in \operatorname{argmin}_{x\in X} T^{z_k}_{\alpha_k}(x)$ all the assertions of Theorem 3.4 about subsequences of (x_k) and their limits are true.*

Proof. We show that the assumptions of Theorem 3.4 are satisfied for $y := y^0$ and $\bar{x} := x^\dagger$. Obviously $S(y^0, z_k) \to 0$ by Assumption 4.1. Further, $S(F(x^\dagger), z_k) \to 0$ by (4.1), which implies

$$\begin{aligned} S(F(x_k), z_k) &= T^{z_k}_{\alpha_k}(x_k) - \alpha_k\Omega(x_k) \le T^{z_k}_{\alpha_k}(x^\dagger) - \alpha_k\Omega(x_k) \\ &= S(F(x^\dagger), z_k) + \alpha_k\big(\Omega(x^\dagger) - \Omega(x_k)\big) \\ &\le S(F(x^\dagger), z_k) + \alpha_k\big(\Omega(x^\dagger) - \inf_{x\in X}\Omega(x)\big) \to 0. \end{aligned}$$

Thus, it only remains to show $\limsup_{k\to\infty}\Omega(x_k) \le \Omega(\hat{x})$ for all S-generalized solutions $\hat{x}$. But this follows immediately from

$$\Omega(x_k) \le \frac{1}{\alpha_k}T^{z_k}_{\alpha_k}(x_k) \le \frac{1}{\alpha_k}T^{z_k}_{\alpha_k}(\hat{x}) = \frac{S(F(\hat{x}), z_k)}{\alpha_k} + \Omega(\hat{x}) \le \frac{\psi(\delta_k)}{\alpha_k} + \Omega(\hat{x}),$$

where we used (4.1). □

The following example shows how to specialize our general data model to the typical Banach space setting.

Example 4.3. Consider the standard Banach space setting introduced in Example 1.2, that is, $Y = Z$ and $S(y_1, y_2) = \frac{1}{p}\|y_1 - y_2\|^p$ for some $p \in (0,\infty)$. Setting $D_{y^0}(y) := \|y - y^0\|$ and $\psi(\delta) := \frac{1}{p}\delta^p$ Assumption 4.1 is satisfied and the estimate $S(y^0, z^\delta) \le \psi(\delta)$ for $z^\delta \in Z^\delta_{y^0}$ (cf. (4.1)) is equivalent to the standard assumption $\|y^\delta - y^0\| \le \delta$. Note that in this setting by Proposition 2.10 two elements $y_1, y_2 \in Y$ are S-equivalent if and only if $y_1 = y_2$.

4.1.2. Handling the solution error

To obtain bounds for the solution error $E_{x^\dagger}(x^z_\alpha)$ in terms of the data error $D_{y^0}(z)$ or in terms of the corresponding noise level δ we have to connect $E_{x^\dagger}$ to the Tikhonov-type functional (1.3). Establishing such a connection requires some preparation.

At first we note that for the distance $S_Y : Y \times Y \to [0,\infty]$ introduced in Definition 2.7 we have the triangle-type inequality

$$S_Y(y_1, y_2) \le S(y_1, z) + S(y_2, z) \quad \text{for all } y_1, y_2 \in Y \text{ and all } z \in Z. \tag{4.2}$$

The additional functional S_Y, as well as the notion of S-equivalence (see Definition 2.6), is the price to pay when considering the case $Y \neq Z$. But it is the triangle-type

inequality (4.2) which allows to handle non-metric fitting functionals. Even in the case $Y = Z$ this inequality is weaker (or at least not stronger) than the usual triangle inequality $S(y_1, y_2) \leq S(y_1, y_3) + S(y_2, y_3)$; but it is strong enough to pave the way for obtaining convergence rates results.

The key to controlling the very general error measure $E_{x^\dagger}$ is the following observation.

Lemma 4.4. *Let $\delta \in [0, \infty)$, $z^\delta \in Z^\delta_{y^0}$, $\alpha \in (0, \infty)$, and $x_\alpha^{z^\delta} \in \operatorname{argmin}_{x \in X} T_\alpha^{z^\delta}(x)$. Further, let $\varphi : [0, \infty) \to [0, \infty)$ be a monotonically increasing function. Then*

$$\begin{aligned}\Omega(x_\alpha^{z^\delta}) - \Omega(x^\dagger) + \varphi\big(S_Y(F(x_\alpha^{z^\delta}), F(x^\dagger))\big) \\ \leq \frac{1}{\alpha}\big(\psi(\delta) - S(F(x_\alpha^{z^\delta}), z^\delta)\big) + \varphi\big(\psi(\delta) + S(F(x_\alpha^{z^\delta}), z^\delta)\big).\end{aligned}$$

Proof. We simply use the minimizing property of $x_\alpha^{z^\delta}$ and estimates (4.1) and (4.2):

$$\begin{aligned}&\Omega(x_\alpha^{z^\delta}) - \Omega(x^\dagger) + \varphi\big(S_Y(F(x_\alpha^{z^\delta}), F(x^\dagger))\big) \\ &\quad = \frac{1}{\alpha}\big(T_\alpha^{z^\delta}(x_\alpha^{z^\delta}) - \alpha\Omega(x^\dagger) - S(F(x_\alpha^{z^\delta}), z^\delta)\big) + \varphi\big(S_Y(F(x_\alpha^{z^\delta}), F(x^\dagger))\big) \\ &\quad \leq \frac{1}{\alpha}\big(S(F(x^\dagger), z^\delta) - S(F(x_\alpha^{z^\delta}), z^\delta)\big) + \varphi\big(S(F(x_\alpha^{z^\delta}), z^\delta) + S(F(x^\dagger), z^\delta)\big) \\ &\quad \leq \frac{1}{\alpha}\big(\psi(\delta) - S(F(x_\alpha^{z^\delta}), z^\delta)\big) + \varphi\big(S(F(x_\alpha^{z^\delta}), z^\delta) + \psi(\delta)\big).\end{aligned}$$

□

The left-hand side in the lemma does not depend directly on the data z^δ or on the noise level δ, whereas the right-hand side is independent of the exact solution $x^\dagger$ and of the exact right-hand side y^0. Of course, at the moment it is not clear whether the estimate in Lemma 4.4 is of any use. But we will see below that it is exactly this estimate which provides us with convergence rates.

In the light of Lemma 4.4 we would like to have

$$E_{x^\dagger}(x_\alpha^{z^\delta}) \leq \Omega(x_\alpha^{z^\delta}) - \Omega(x^\dagger) + \varphi\big(S_Y(F(x_\alpha^{z^\delta}), F(x^\dagger))\big)$$

for all minimizers $x_\alpha^{z^\delta}$. But since the connection between $E_{x^\dagger}$ and the Tikhonov-type functional should be independent of the a priori unknown data z^δ and therefore also of the minimizers $x_\alpha^{z^\delta}$, we have to demand such an estimate from all x in a sufficiently large set $M \subseteq X$. Here, 'sufficiently large' has to be understood in the following sense:

Assumption 4.5. Given a parameter choice $(\delta, z^\delta) \mapsto \alpha(\delta, z^\delta)$ let $M \subseteq X$ be a set such that $S_Y(F(x), F(x^\dagger)) < \infty$ for all $x \in M$ and such that there is some $\bar{\delta} > 0$ with

$$\bigcup_{z^\delta \in Z^\delta_{y^0}} \operatorname*{argmin}_{x \in X} T^{z^\delta}_{\alpha(\delta, z^\delta)}(x) \subseteq M \quad \text{for all } \delta \in (0, \bar{\delta}].$$

Obviously $M = X$ satisfies Assumption 4.5. The following proposition gives an example of a smaller set M if the regularization parameter α is chosen a priori as in Corollary 4.2. A similar construction for a set containing the minimizers of the Tikhonov-type functional is used in [HKPS07, Pös08].

Proposition 4.6. *Let $\bar{\alpha} > 0$ and $\varrho > \Omega(x^\dagger)$ and let $\delta \mapsto \alpha(\delta)$ be a parameter choice satisfying $\alpha(\delta) \to 0$ and $\frac{\psi(\delta)}{\alpha(\delta)} \to 0$ if $\delta \to 0$. Then*

$$M := \{x \in X : S_Y(F(x), F(x^\dagger)) + \bar{\alpha}\Omega(x) \leq \varrho\bar{\alpha}\}$$

satisfies Assumption 4.5.

Proof. Because $\alpha(\delta) \to 0$ and $\frac{\psi(\delta)}{\alpha(\delta)} \to 0$ if $\delta \to 0$ there exists some $\bar{\delta} > 0$ with $\alpha(\delta) \leq \bar{\alpha}$ and $\frac{\psi(\delta)}{\alpha(\delta)} \leq \frac{1}{2}(\varrho - \Omega(x^\dagger))$ for all $\delta \in (0, \bar{\delta}]$. For the sake of brevity we now write α instead of $\alpha(\delta)$.

Using (4.2), (4.1), $1 \leq \frac{\bar{\alpha}}{\alpha}$, and $2\frac{\psi(\delta)}{\alpha} + \Omega(x^\dagger) \leq \varrho$, for each $\delta \in (0, \bar{\delta}]$, each $z^\delta \in Z^\delta_{y^0}$, and each minimizer $x_\alpha^{z^\delta}$ of $T_\alpha^{z^\delta}$ we have

$$\begin{aligned}
&S_Y(F(x_\alpha^{z^\delta}), F(x^\dagger)) + \bar{\alpha}\Omega(x_\alpha^{z^\delta}) \\
&\quad \leq S(F(x_\alpha^{z^\delta}), z^\delta) + \psi(\delta) + \bar{\alpha}\Omega(x_\alpha^{z^\delta}) = \psi(\delta) + \frac{\bar{\alpha}}{\alpha}T_\alpha^{z^\delta}(x_\alpha^{z^\delta}) - \Big(\frac{\bar{\alpha}}{\alpha} - 1\Big)S(F(x_\alpha^{z^\delta}), z^\delta) \\
&\quad \leq \psi(\delta) + \frac{\bar{\alpha}}{\alpha}T_\alpha^{z^\delta}(x^\dagger) \leq \Big(1 + \frac{\bar{\alpha}}{\alpha}\Big)\psi(\delta) + \bar{\alpha}\Omega(x^\dagger) \leq \bar{\alpha}\Big(2\frac{\psi(\delta)}{\alpha} + \Omega(x^\dagger)\Big) \leq \varrho\bar{\alpha},
\end{aligned}$$

that is, $x_\alpha^{z^\delta} \in M$. □

Now we are in the position to state the connection between the solution error $E_{x^\dagger}$ and the Tikhonov-type functional (1.3).

Assumption 4.7. Let $M \subseteq X$ and assume that there exist a constant $\beta \in (0, \infty)$ and a monotonically increasing function $\varphi : [0, \infty) \to [0, \infty)$ such that

$$\beta E_{x^\dagger}(x) \leq \Omega(x) - \Omega(x^\dagger) + \varphi\big(S_Y(F(x), F(x^\dagger))\big) \quad \text{for all } x \in M. \tag{4.3}$$

Showing that the exact solution $x^\dagger$ satisfies an inequality of the form (4.3) on a set fulfilling Assumption 4.5 will be the main task when proving convergence rates. Using, e.g., norms or Bregman distances for the solution error $E_{x^\dagger}$, inequality (4.3) becomes a kind of smoothness assumption on $x^\dagger$. We have a detailed look at such inequalities in Part III.

Inequalities of type (4.3) appeared first in [HKPS07] as 'variational inequalities'. There they have been introduced for the standard Banach space setting of Example 1.2 with $\varphi(t) = ct^{1/p}$, $c > 0$ (see also [SGG+09, Section 3.2] for results on this special case). Details on how to specialize our general model for measuring the solution error to that setting will be given below in Example 4.8. An extension of variational inequalities to more general Tikhonov-type methods, again only for one fixed function φ, has been given in [Pös08]. For the standard Banach space setting the technique has been extended in [HH09] using arbitrary monomials φ, that is, the restriction of φ to the reciprocal of the norm exponent in the fitting functional has been dropped. In [Gei09, FH10] the two mentioned extensions are combined. Variational inequalities with more general φ are considered in [BH10] for a slightly extended Banach space setting and first results using the, to our knowledge, up to now most general version (4.3) of variational inequalities have been published in [Fle10a]. The setting and proof

techniques considered in [Gra10a] differ only slightly from [Fle10b], but the formulation of the results there is not as general as ours. Here we should mention that the corresponding preprints [Fle10b] and [Gra10b] appeared almost at the same time. Examples of variational inequalities with non-standard error measures $E_{x^\dagger}$ are given in [BL09, Lemmas 4.4 and 4.6] and [Gra10a].

The cited literature only considers Tikhonov-type regularization methods. But the technique of variational inequalities has also been used, even though rarely, to obtain convergence rates for other regularization methods (see, e.g., [Hei08a, GHS09]). In Part III we will see that at least for linear problems in Hilbert spaces variational inequalities provide convergence rates for a wide class of linear regularization methods.

The following example shows how to specialize our general error model to the standard Banach space setting.

Example 4.8. Consider the standard Banach space setting introduced in Example 1.2 (see also Example 4.3), that is, X and $Y = Z$ are Banach spaces and $S(y_1, y_2) = \frac{1}{p}\|y_1 - y_2\|^p$ for some $p \in (0, \infty)$. As error measure $E_{x^\dagger}$ in Banach spaces one usually uses the Bregman distance

$$E_{x^\dagger}(x) := B_{\xi^\dagger}^{\Omega}(x, x^\dagger) = \Omega(x) - \Omega(x^\dagger) - \langle \xi^\dagger, x - x^\dagger \rangle$$

with respect to some subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ (cf. Example 1.2). Note that a Bregman distance can only be defined if $\partial\Omega(x^\dagger) \neq \emptyset$.

In Proposition 2.13 we have seen $S_Y(y_1, y_2) = \frac{1}{p}\min\{1, 2^{1-p}\}\|y_1 - y_2\|^p$. Together with $\varphi(t) := ct^{\kappa/p}$ for $\kappa \in (0, \infty)$ and some $c > 0$ the variational inequality (4.3) attains the form

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \tilde{c}\|F(x) - F(x^\dagger)\|^\kappa$$

with $\tilde{c} > 0$. This variational inequality is equivalent to

$$-\langle \xi^\dagger, x - x^\dagger \rangle \leq \beta_1 B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \beta_2\|F(x) - F(x^\dagger)\|^\kappa$$

with constants $\beta_1 := 1 - \beta$ and $\beta_2 := \tilde{c}$. The last inequality is of the type introduced in [HKPS07] for $\kappa = 1$ and in [HH09] for general κ.

4.2. Convergence rates with a priori parameter choice

In this section we give a first convergence rates result using an a priori parameter choice. The obtained rate only depends on the function φ in the variational inequality (4.3). For future reference we formulate the following properties of this function φ.

Assumption 4.9. The function $\varphi : [0, \infty) \to [0, \infty)$ satisfies:

(i) φ is monotonically increasing, $\varphi(0) = 0$, and $\varphi(t) \to 0$ if $t \to 0$;

(ii) there exists a constant $\gamma > 0$ such that φ is concave and strictly monotonically increasing on $[0, \gamma]$;

(iii) the inequality

$$\varphi(t) \leq \varphi(\gamma) + \left(\inf_{\tau \in [0,\gamma)} \frac{\varphi(\gamma) - \varphi(\tau)}{\gamma - \tau} \right) (t - \gamma)$$

is satisfied for all $t > \gamma$.

If φ satisfies items (i) and (ii) of Assumption 4.9 and if φ is differentiable in γ, then item (iii) is equivalent to

$$\varphi(t) \leq \varphi(\gamma) + \varphi'(\gamma)(t - \gamma) \quad \text{for all } t > \gamma.$$

For example, $\varphi(t) = t^\mu$, $\mu \in (0, 1]$, satisfies Assumption 4.9 for each $\gamma > 0$. The function

$$\varphi(t) = \begin{cases} (-\ln t)^{-\mu}, & t \leq \mathrm{e}^{-\mu-1}, \\ \left(\frac{1}{\mu+1}\right)^\mu + \mu\left(\frac{\mathrm{e}}{\mu+1}\right)^{\mu+1}(t - \mathrm{e}^{-\mu-1}), & \text{else} \end{cases}$$

with $\mu > 0$ has a sharper cusp at zero than monomials and satisfies Assumption 4.9 for $\gamma \in (0, \mathrm{e}^{-\mu-1}]$.

In preparation of the main convergence rates result of this thesis we prove the following error estimates. The idea to use conjugate functions (see Definition B.4) for expressing the error bounds comes from [Gra10a].

Lemma 4.10. *Let $x^\dagger$ satisfy Assumption 4.7. Then*

$$\beta E_{x^\dagger}(x_\alpha^{z^\delta}) \leq 2\frac{\psi(\delta)}{\alpha} + (-\varphi)^*\Big(-\frac{1}{\alpha}\Big) \quad \text{if } x_\alpha^{z^\delta} \in M,$$

where $\alpha > 0$, $\delta \geq 0$, $z^\delta \in Z_{y^0}^\delta$, and $x_\alpha^{z^\delta} \in \operatorname{argmin}_{x \in X} T_\alpha^{z^\delta}(x)$. If φ is invertible with inverse $\varphi^{-1} : \mathcal{R}(\varphi) \to [0, \infty)$, then

$$\beta E_{x^\dagger}(x_\alpha^{z^\delta}) \leq 2\frac{\psi(\delta)}{\alpha} + \frac{1}{\alpha}\big(\varphi^{-1}\big)^*(\alpha) \quad \text{if } x_\alpha^{z^\delta} \in M.$$

Proof. By Lemma 4.4 and inequality (4.3) we have

$$\beta E_{x^\dagger}(x_\alpha^{z^\delta}) \leq \frac{1}{\alpha}\big(\psi(\delta) - S(F(x_\alpha^{z^\delta}), z^\delta)\big) + \varphi\big(S(F(x_\alpha^{z^\delta}), z^\delta) + \psi(\delta)\big)$$

and therefore

$$\begin{aligned} \beta E_{x^\dagger}(x_\alpha^{z^\delta}) &\leq 2\frac{\psi(\delta)}{\alpha} + \varphi\big(S(F(x_\alpha^{z^\delta}), z^\delta) + \psi(\delta)\big) - \frac{1}{\alpha}\big(S(F(x_\alpha^{z^\delta}), z^\delta) + \psi(\delta)\big) \\ &\leq 2\frac{\psi(\delta)}{\alpha} + \sup_{t \geq 0}\Big(\varphi(t) - \frac{1}{\alpha}t\Big) \end{aligned}$$

Now the first estimate in the lemma follows from

$$\sup_{t \geq 0}\Big(\varphi(t) - \frac{1}{\alpha}t\Big) = \sup_{t \geq 0}\Big(-\frac{1}{\alpha}t - (-\varphi)(t)\Big) = (-\varphi)^*\Big(-\frac{1}{\alpha}\Big)$$

and the second from

$$\sup_{t \geq 0}\Big(\varphi(t) - \frac{1}{\alpha}t\Big) = \sup_{s \in \mathcal{R}(\varphi)}\Big(s - \frac{1}{\alpha}\varphi^{-1}(s)\Big) = \frac{1}{\alpha}\sup_{s \in \mathcal{R}(\varphi)}\big(\alpha s - \varphi^{-1}(s)\big) = \frac{1}{\alpha}\big(\varphi^{-1}\big)^*(\alpha).$$

□

The following main result is an adaption of Theorem 4.3 in [BH10] to our generalized setting. A similar result for the case $Z := Y$ can be found in [Gra10a]. Unlike the corresponding proofs given in [BH10] and [Gra10a] our proof avoids the use of Young-type inequalities or differentiability assumptions on $(-\varphi)^*$ or $\left(\varphi^{-1}\right)^*$. Therefore it works also for non-differentiable functions φ in the variational inequality (4.3).

Theorem 4.11 (convergence rates)**.** *Let $x^\dagger$ satisfy Assumption 4.7 such that the associated function φ satisfies Assumption 4.9 and let $\delta \mapsto \alpha(\delta)$ be a parameter choice such that*

$$\inf_{\tau\in[0,\psi(\delta))} \frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} \geq \frac{1}{\alpha(\delta)} \geq \sup_{\tau\in(\psi(\delta),\gamma]} \frac{\varphi(\tau) - \varphi(\psi(\delta))}{\tau - \psi(\delta)} \tag{4.4}$$

for all $\delta > 0$ with $\psi(\delta) < \gamma$, where ψ and γ come from Assumptions 4.1 and 4.9, respectively. Further, let M satisfy Assumption 4.5. Then there is some $\bar{\delta} > 0$ such that

$$E_{x^\dagger}\left(x^{z^\delta}_{\alpha(\delta)}\right) \leq \frac{2}{\beta}\varphi(\psi(\delta)) \quad \text{for all } \delta \in (0, \bar{\delta}],$$

where $z^\delta \in Z^\delta_{y^0}$ and $x^{z^\delta}_{\alpha(\delta)} \in \operatorname{argmin}_{x\in X} T^{z^\delta}_{\alpha(\delta)}(x)$. The constant β comes from Assumption 4.7.

Before we prove the theorem the a priori parameter choice (4.4) requires some comments.

Remark 4.12. By item (ii) of Assumption 4.9 a parameter choice satisfying (4.4) exists, that is,

$$\infty > \inf_{\tau\in[0,t)} \frac{\varphi(t) - \varphi(\tau)}{t - \tau} \geq \sup_{\tau\in(t,\gamma]} \frac{\varphi(\tau) - \varphi(t)}{\tau - t} > 0$$

for all $t \in (0, \gamma)$.

By (4.4) we have

$$\frac{\psi(\delta)}{\alpha(\delta)} \leq \psi(\delta)\frac{\varphi(\psi(\delta)) - \varphi(0)}{\psi(\delta) - 0} = \varphi(\psi(\delta)) \to 0 \quad \text{if } \delta \to 0 \ (\delta > 0)$$

and if $\sup_{\tau\in(t,\gamma]} \frac{\varphi(\tau)-\varphi(t)}{\tau-t} \to \infty$ if $t \to 0$ $(t > 0)$, then $\alpha(\delta) \to 0$ if $\delta \to 0$ $(\delta > 0)$. Thus, if the supremum goes to infinity, the parameter choice satisfies the assumptions of Corollary 4.2 and of Proposition 4.6.

If $\sup_{\tau\in(t,\gamma]} \frac{\varphi(\tau)-\varphi(t)}{\tau-t} \leq c$ for some $c > 0$ and all $t \in (0, t_0]$, $t_0 > 0$, then $\varphi(s) \leq cs$ for all $s \in [0, \infty)$. To see this, take $s \in (0, \gamma]$ and $t \in (0, \min\{s, t_0\})$. Then

$$\frac{\varphi(s) - \varphi(t)}{s - t} \leq \sup_{\tau\in(t,\gamma]} \frac{\varphi(\tau) - \varphi(t)}{\tau - t} \leq c$$

and thus, $\varphi(s) \leq cs + \varphi(t) - ct$. Letting $t \to 0$ we obtain $\varphi(s) \leq cs$ for all $s \in (0, \gamma]$. For $s > \gamma$ by item (iii) we have

$$\varphi(s) \leq \varphi(\gamma) + \left(\inf_{\tau\in[0,\gamma)} \frac{\varphi(\gamma) - \varphi(\tau)}{\gamma - \tau}\right)(s - \gamma) \leq \varphi(\gamma) + \frac{\varphi(\gamma)}{\gamma}(s - \gamma) \leq \frac{c\gamma}{\gamma}s = cs.$$

The case $\varphi(s) \leq cs$ for all $s \in [0, \infty)$ is a singular situation. Details are given in Proposition 4.14 below.

Remark 4.13. If φ is differentiable in $(0,\gamma)$ then the parameter choice (4.4) is equivalent to

$$\alpha(\delta) = \frac{1}{\varphi'(\psi(\delta))}.$$

Now we prove the theorem.

Proof of Theorem 4.11. We write α instead of $\alpha(\delta)$.

By Assumption 4.5 we have $x_\alpha^{z^\delta} \in M$ for sufficiently small $\delta > 0$. Thus, Lemma 4.10 gives

$$\beta E_{x^\dagger}\big(x_\alpha^{z^\delta}\big) \le 2\frac{\psi(\delta)}{\alpha} + \sup_{\tau\in[0,\infty)}\Big(\varphi(\tau) - \frac{1}{\alpha}\tau\Big).$$

If we can show, for sufficiently small δ and α as proposed in the theorem, that

$$\varphi(\tau) - \frac{1}{\alpha}\tau \le \varphi(\psi(\delta)) - \frac{1}{\alpha}\psi(\delta) \quad \text{for all } \tau \ge 0, \tag{4.5}$$

then we obtain

$$\begin{aligned}\beta E_{x^\dagger}\big(x_\alpha^{z^\delta}\big) &\le \frac{\psi(\delta)}{\alpha} + \varphi(\psi(\delta)) \le \psi(\delta) \inf_{\tau\in[0,\psi(\delta))} \frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} + \varphi(\psi(\delta)) \\ &\le \psi(\delta)\frac{\varphi(\psi(\delta)) - \varphi(0)}{\psi(\delta) - 0} + \varphi(\psi(\delta)) = 2\varphi(\psi(\delta)).\end{aligned}$$

Thus, it remains to show (4.5).

First we note that for fixed $t \in (0,\gamma)$ and all $\tau > \gamma$ item (iii) of Assumption 4.9 implies

$$\begin{aligned}\frac{\varphi(\tau) - \varphi(t)}{\tau - t} &\le \frac{1}{\tau - t}\left(\varphi(\gamma) + \left(\inf_{\sigma\in[0,\gamma)} \frac{\varphi(\gamma) - \varphi(\sigma)}{\gamma - \sigma}\right)(\tau - \gamma) - \varphi(t)\right) \\ &\le \frac{1}{\tau - t}\left(\varphi(\gamma) + \frac{\varphi(\gamma) - \varphi(t)}{\gamma - t}(\tau - \gamma) - \varphi(t)\right) = \frac{\varphi(\gamma) - \varphi(t)}{\gamma - t}.\end{aligned}$$

Using this estimate with $t = \psi(\delta)$ we can extend the supremum in the lower bound for $\frac{1}{\alpha}$ in (4.4) from $\tau \in (\psi(\delta),\gamma]$ to $\tau \in (\psi(\delta),\infty)$, that is,

$$\inf_{\tau\in[0,\psi(\delta))} \frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} \ge \frac{1}{\alpha} \ge \sup_{\tau\in(\psi(\delta),\infty)} \frac{\varphi(\tau) - \varphi(\psi(\delta))}{\tau - \psi(\delta)}$$

or, equivalently,

$$\begin{aligned}&\frac{\varphi(\psi(\delta)) - \varphi(\tau)}{\psi(\delta) - \tau} \ge \frac{1}{\alpha} \quad \text{for all } \tau \in [0,\psi(\delta)) \quad \text{and} \\ &\frac{\varphi(\tau) - \varphi(\psi(\delta))}{\tau - \psi(\delta)} \le \frac{1}{\alpha} \quad \text{for all } \tau \in (\psi(\delta),\infty).\end{aligned}$$

These two inequalities together are equivalent to

$$\frac{1}{\alpha}(\psi(\delta) - \tau) \le \varphi(\psi(\delta)) - \varphi(\tau) \quad \text{for all } \tau \ge 0$$

and simple rearrangements yield (4.5). □

Note that Theorem 4.11 covers only the case $\delta > 0$. If $\delta = 0$, then we would intuitively choose $\alpha(\delta) = 0$. But for $\alpha = 0$ the minimization problem (1.3) is not stable and it may happen that it has no solution. Thus, also in the case $\delta = 0$ we have to choose $\alpha > 0$. Lemma 4.10 then provides the estimate

$$E_{x^\dagger}(x_\alpha^{z^0}) \le \frac{1}{\beta}(-\varphi)^*\Big(-\frac{1}{\alpha}\Big).$$

In connection with $\delta = 0$ the authors of [BO04] observed a phenomenon called *exact penalization*. This means that for $\delta = 0$ and sufficiently small $\alpha > 0$ the solution error $E_{x^\dagger}(x_\alpha^{z^0})$ is zero. In [BO04] a Banach space setting similar to Example 4.8 with $p = 1$ and $\kappa = 1$ was considered. In our more general setting we observe the same phenomenon:

Proposition 4.14. *Let $x^\dagger$ satisfy Assumption 4.7. If $\varphi(t) \le ct$ for some $c > 0$ and all $t \in [0, \infty)$, then*

$$E_{x^\dagger}(x_\alpha^{z^\delta}) \le \frac{2}{\beta}\frac{\psi(\delta)}{\alpha} \quad \text{if } x_\alpha^{z^\delta} \in M \text{ and } \alpha \in \big(0, \tfrac{1}{c}\big],$$

where $\delta \ge 0$, $z^\delta \in Z_{y^0}^\delta$, and $x_\alpha^{z^\delta} \in \operatorname{argmin}_{x\in X} T_\alpha^{z^\delta}(x)$. That is, choosing a fixed $\tilde\alpha \in (0, \frac{1}{c}]$ we obtain

$$E_{x^\dagger}(x_{\tilde\alpha}^{z^\delta}) = \mathcal{O}(\psi(\delta)) \quad \text{if } \delta \to 0.$$

Proof. By Lemma 4.10 we have

$$\beta E_{x^\dagger}(x_\alpha^{z^\delta}) \le 2\frac{\psi(\delta)}{\alpha} + (-\varphi)^*\Big(-\frac{1}{\alpha}\Big) \quad \text{if } x_\alpha^{z^\delta} \in M$$

and using $\varphi(t) \le ct$ and $\alpha \le \frac{1}{c}$ we see

$$(-\varphi)^*\Big(-\frac{1}{\alpha}\Big) = \sup_{t\ge 0}\Big(\varphi(t) - \frac{1}{\alpha}t\Big) \le \sup_{t\ge 0}\Big(ct - \frac{1}{\alpha}t\Big) \le 0.$$

Thus, the assertion is true. □

4.3. The discrepancy principle

The convergence rate obtained in Theorem 4.11 is based on an a priori parameter choice. That is, we have proven that there is some parameter choice yielding the asserted rate. For practical purposes the proposed choice is useless because it is based on the function φ in the variational inequality (4.3) and therefore on the unknown Ω-minimizing S-generalized solution $x^\dagger$.

In this section we propose another parameter choice strategy, which requires only the knowledge of the noise level δ, the noisy data z^δ, and the fitting functional $S(F(x_\alpha^{z^\delta}), z^\delta)$ at the regularized solutions $x_\alpha^{z^\delta}$ for all $\alpha > 0$. The expression $S(F(x_\alpha^{z^\delta}), z^\delta)$ is also known as *discrepancy* and therefore the parameter choice which we introduce and analyze in this section is known as the *discrepancy principle* or the *Morozow discrepancy principle*.

Choosing the regularization parameter according to the discrepancy principle is a standard technique in Hilbert space settings (see, e.g., [EHN96]) and some extensions to Banach spaces, nonlinear operators, and general stabilizing functionals Ω are available (see, e.g., [AR11, Bon09, KNS08, TLY98]). To our knowledge the most general analysis of the discrepancy principle in connection with Tikhonov-type regularization methods is given in [JZ10]. Motivated by this paper we show that the discrepancy principle is also applicable in our more general setting. The major contribution will be that in contrast to [JZ10] we do not assume uniqueness of the minimizers of the Tikhonov-type functional (1.3). Another a posteriori parameter choice which works for very general Tikhonov-type regularization methods and which is capable of handling multiple minimizers of the Tikhonov-type functional is proposed in [IJT10]. The basic ideas of this section are taken from this paper and [JZ10].

4.3.1. Motivation and definition

Remember that $x^\dagger$ is one fixed Ω-minimizing S-generalized solution to the original equation (1.1) with right-hand side $y^0 \in Y$. Given a fixed data element $z^\delta \in Z^\delta_{y^0}$, $\delta \geq 0$, for each $\alpha \in (0, \infty)$ the Tikhonov-type regularization method (1.3) provides a nonempty set

$$X_\alpha^{z^\delta} := \operatorname*{argmin}_{x \in X} T_\alpha^{z^\delta}(x)$$

of regularized solutions. We would like to choose a regularization parameter α^* in such a way that $X_{\alpha^*}^{z^\delta}$ contains an element $x_{\alpha^*}^{z^\delta}$ for which the solution error $E_{x^\dagger}$ becomes minimal with respect to α, that is,

$$E_{x^\dagger}(x_{\alpha^*}^{z^\delta}) = \inf\{E_{x^\dagger}(x_\alpha^{z^\delta}) : \alpha > 0,\ x_\alpha^{z^\delta} \in X_\alpha^{z^\delta}\}.$$

In this sense α^* is the *optimal regularization parameter*. But in practice we do not know $x^\dagger$ and therefore α^* is not accessible.

The only thing we know about $x^\dagger$ is that $F(x^\dagger)$ is S-equivalent to y^0 and that $D_{y^0}(z^\delta) \leq \delta$ (cf. Subsection 4.1.1). Avoiding the use of the intermediate functional D_{y^0}, Assumption 4.1 provides the possibly weaker condition $S(F(x^\dagger), z^\delta) \leq \psi(\delta)$. Therefore it is reasonable to choose an α such that at least for one $x_\alpha^{z^\delta} \in X_\alpha^{z^\delta}$ the element $F(x_\alpha^{z^\delta})$ lies 'near' the set $\{y \in Y : S(y, z^\delta) \leq \psi(\delta)\}$. Of course, many different parameters α could satisfy such a condition. We use this degree of freedom and choose the most regular $x_\alpha^{z^\delta}$, where 'regular' means that the stabilizing functional Ω is small. How to realize this is shown in the following proposition (cf. [JZ10, Lemma 2.1]).

Proposition 4.15. *Let $0 < \alpha_1 < \alpha_2 < \infty$ be two regularization parameters and take corresponding minimizers $x_{\alpha_1}^{z^\delta} \in X_{\alpha_1}^{z^\delta}$ and $x_{\alpha_2}^{z^\delta} \in X_{\alpha_2}^{z^\delta}$. Then*

$$S(F(x_{\alpha_1}^{z^\delta}), z^\delta) \leq S(F(x_{\alpha_2}^{z^\delta}), z^\delta) \quad \text{and} \quad \Omega(x_{\alpha_1}^{z^\delta}) \geq \Omega(x_{\alpha_2}^{z^\delta}).$$

Proof. Because $x_{\alpha_1}^{z^\delta}$ and $x_{\alpha_2}^{z^\delta}$ are minimizers of the Tikhonov-type functionals $T_{\alpha_1}^{z^\delta}$ and $T_{\alpha_2}^{z^\delta}$, respectively, we have

$$\begin{aligned} S(F(x_{\alpha_1}^{z^\delta}), z^\delta) + \alpha_1 \Omega(x_{\alpha_1}^{z^\delta}) &\leq S(F(x_{\alpha_2}^{z^\delta}), z^\delta) + \alpha_1 \Omega(x_{\alpha_2}^{z^\delta}), \\ S(F(x_{\alpha_2}^{z^\delta}), z^\delta) + \alpha_2 \Omega(x_{\alpha_2}^{z^\delta}) &\leq S(F(x_{\alpha_1}^{z^\delta}), z^\delta) + \alpha_2 \Omega(x_{\alpha_1}^{z^\delta}). \end{aligned}$$

Adding these two inequalities the S-terms can be eliminated and simple rearrangements yield $(\alpha_2 - \alpha_1)\big(\Omega(x_{\alpha_2}^{z^\delta}) - \Omega(x_{\alpha_1}^{z^\delta})\big) \leq 0$. Dividing the first inequality by α_1 and the second by α_2, the sum of the resulting inequalities allows to eliminate the Ω-terms. Simple rearrangements now give $(\frac{1}{\alpha_1} - \frac{1}{\alpha_2})\big(S(F(x_{\alpha_2}^{z^\delta}), z^\delta) - S(F(x_{\alpha_1}^{z^\delta}), z^\delta)\big) \geq 0$ and multiplication by $\alpha_1\alpha_2$ shows $(\alpha_2 - \alpha_1)\big(S(F(x_{\alpha_2}^{z^\delta}), z^\delta) - S(F(x_{\alpha_1}^{z^\delta}), z^\delta)\big) \geq 0$. Dividing the two derived inequalities by $\alpha_2 - \alpha_1$ proves the assertion. □

Proposition 4.15 shows that the larger α the more regular the corresponding regularized solutions $x_\alpha^{z^\delta}$, but $S(F(x_\alpha^{z^\delta}), z^\delta)$ increases as α becomes larger. Thus, choosing α such that $S(F(x_\alpha^{z^\delta}), z^\delta) \approx \psi(\delta)$ meets both requirements, $F(x_\alpha^{z^\delta})$ lies 'near' the set $\{y \in Y : S(y, z^\delta) \leq \psi(\delta)\}$ and $\Omega(x_\alpha^{z^\delta})$ is as small as possible.

The '$\approx$' symbol can be made even more precise: Because our data model allows $S(F(x^\dagger), z^\delta) = \psi(\delta)$, there is no reason to choose α with $S(F(x_\alpha^{z^\delta}), z^\delta) < \psi(\delta)$. On the other hand a regularized solution $x_{\alpha^*}^{z^\delta}$ corresponding to the optimal regularization parameter α^* has not to coincide with $x^\dagger$ and therefore $S(F(x_{\alpha^*}^{z^\delta}), z^\delta) > \psi(\delta)$ is possible. But since the regularized solutions are approximations of $x^\dagger$, the difference $S(F(x_{\alpha^*}^{z^\delta}), z^\delta) - \psi(\delta)$ cannot become arbitrarily large, in fact it has to go to zero if $\delta \to 0$. Following this reasoning, we should demand

$$\psi(\delta) \leq S(F(x_\alpha^{z^\delta}), z^\delta) \leq c\psi(\delta) \tag{4.6}$$

with some $c > 1$.

If we choose α in such a way that (4.6) is satisfied for at least one $x_\alpha^{z^\delta} \in X_\alpha^{z^\delta}$, then we say that α is chosen according to the *discrepancy principle*.

4.3.2. Properties of the discrepancy inequality

The aim of this subsection is to show that the discrepancy inequality (4.6) has a solution and that this inequality is also manageable if the sets $X_\alpha^{z^\delta}$, $\alpha > 0$, contain more than one element.

To keep explanations accessible we first introduce some notation. By $u = (u_\alpha)_{\alpha>0}$ we denote a family of regularized solutions $u_\alpha \in X_\alpha^{z^\delta}$ and

$$U := \{(u_\alpha)_{\alpha>0} : u_\alpha \in X_\alpha^{z^\delta}\}$$

contains all such families. To each $u \in U$ we assign a function $S_u : (0, \infty) \to [0, \infty)$ via $S_u(\alpha) := S(F(u_\alpha), z^\delta)$. Remember that $S(F(x^\dagger), z^\delta) \leq \psi(\delta)$ by Assumption 4.1 and that $\Omega(x^\dagger) < \infty$ by assumption. Therefore, Theorem 3.2 (existence) guarantees $X_\alpha^{z^\delta} \neq \emptyset$ and $S(F(x_\alpha^{z^\delta}), z^\delta) < \infty$ for all $x_\alpha^{z^\delta} \in X_\alpha^{z^\delta}$. If all $X_\alpha^{z^\delta}$, $\alpha > 0$, contain only one element $x_\alpha^{z^\delta}$, then $U = \{\tilde{u}\}$ with $\tilde{u}_\alpha = x_\alpha^{z^\delta}$. In this case solving the discrepancy inequality (4.6) is equivalent to finding some α with

$$\psi(\delta) \leq S_{\tilde{u}}(\alpha) \leq c\psi(\delta).$$

In general, the sets $X_\alpha^{z^\delta}$ contain more than one element and the question of the influence of $u \in U$ on S_u arises. By Proposition 4.15 we know that S_u is monotonically increasing for all $u \in U$. Thus, a well-known result from real analysis says that for each

$\alpha > 0$ the one-sided limits $S_u(\alpha - 0)$ and $S_u(\alpha + 0)$ exist (and are finite) and that S_u has at most countably many points of discontinuity (see, e.g., [Zor04, Proposition 3, Corollary 2]). The following lemma formulates the fundamental observation for analyzing the influence of $u \in U$ on S_u.

Lemma 4.16. *Let $\tilde{\alpha} > 0$. Then $S_{u^1}(\tilde{\alpha} - 0) = S_{u^2}(\tilde{\alpha} - 0)$ for all $u^1, u^2 \in U$ and there is some $\tilde{u} \in U$ with $S_{\tilde{u}}(\tilde{\alpha}) = S_{\tilde{u}}(\tilde{\alpha} - 0)$. The same is true if $\tilde{\alpha} - 0$ is replaced by $\tilde{\alpha} + 0$.*

Proof. Let $u^1, u^2 \in U$ and take sequences $(\alpha_k^1)_{k\in\mathbb{N}}$ and $(\alpha_k^2)_{k\in\mathbb{N}}$ in $(0, \infty)$ converging to $\tilde{\alpha}$ and satisfying $\alpha_k^1 < \alpha_k^2 < \tilde{\alpha}$. Proposition 4.15 gives $S_{u^1}(\alpha_k^1) \le S_{u^2}(\alpha_k^2)$ for all $k \in \mathbb{N}$. Thus,

$$S_{u^1}(\tilde{\alpha} - 0) = \lim_{k\to\infty} S_{u^1}(\alpha_k^1) \le \lim_{k\to\infty} S_{u^2}(\alpha_k^2) = S_{u^2}(\tilde{\alpha} - 0).$$

Interchanging the roles of u^1 and u^2 shows $S_{u^1}(\tilde{\alpha} - 0) = S_{u^2}(\tilde{\alpha} - 0)$.

For proving the existence of $\tilde{u}$ we apply Theorem 3.3 (stability) with $z := z^\delta$, $z_k := z^\delta$, $\alpha := \tilde{\alpha}$, $\varepsilon_k := 0$, and $\bar{x} := x^\dagger$. Take a sequence $(\alpha_k)_{k\in\mathbb{N}}$ in $(0, \tilde{\alpha})$ converging to $\tilde{\alpha}$. Then, by Theorem 3.3, a corresponding sequence of minimizers $(x_k)_{k\in\mathbb{N}}$ of $T_{\alpha_k}^{z^\delta}$ has a subsequence $(x_{k_l})_{l\in\mathbb{N}}$ converging to a minimizer $\tilde{x}$ of $T_{\tilde{\alpha}}^{z^\delta}$. Theorem 3.3 also gives $S(F(x_{k_l}), z^\delta) \to S(F(\tilde{x}), z^\delta)$. Thus, choosing $\tilde{u}$ such that $\tilde{u}_{\alpha_{k_l}} = x_{k_l}$ and $\tilde{u}_{\tilde{\alpha}} = \tilde{x}$, we get $S_{\tilde{u}}(\alpha_{k_l}) \to S_{\tilde{u}}(\tilde{\alpha})$ and therefore $S_{\tilde{u}}(\tilde{\alpha}) = S_{\tilde{u}}(\tilde{\alpha} - 0)$.

Analog arguments apply if $\tilde{\alpha} - 0$ is replaced by $\tilde{\alpha} + 0$ in the lemma. □

Exploiting Proposition 4.15 and Lemma 4.16 we obtain the following result on the continuity of the functions S_u, $u \in U$.

Proposition 4.17. *For fixed $\tilde{\alpha} > 0$ the following assertions are equivalent:*

(i) There exists some $\tilde{u} \in U$ such that $S_{\tilde{u}}$ is continuous in $\tilde{\alpha}$.

(ii) For all $u \in U$ the function S_u is continuous in $\tilde{\alpha}$.

(iii) $S_{u^1}(\tilde{\alpha}) = S_{u^2}(\tilde{\alpha})$ for all $u^1, u^2 \in U$.

Proof. Obviously, (ii) implies (i). We show '(i) ⇒ (iii)'. So let (i) be satisfied, that is $S_{\tilde{u}}(\tilde{\alpha} - 0) = S_{\tilde{u}}(\tilde{\alpha}) = S_{\tilde{u}}(\tilde{\alpha} + 0)$, and let $u \in U$. Then, by Proposition 4.15, $S_{\tilde{u}}(\alpha_1) \le S_u(\tilde{\alpha}) \le S_{\tilde{u}}(\alpha_2)$ for all $\alpha_1 \in (0, \tilde{\alpha})$ and all $\alpha_2 \in (\tilde{\alpha}, \infty)$. Thus, $S_{\tilde{u}}(\tilde{\alpha} - 0) \le S_u(\tilde{\alpha}) \le S_{\tilde{u}}(\tilde{\alpha} + 0)$, which implies (iii).

It remains to show '(iii) ⇒ (ii)'. Let (iii) be satisfied and let $u \in U$. By Lemma 4.16 there is some $\tilde{u}$ with $S_{\tilde{u}}(\tilde{\alpha}) = S_{\tilde{u}}(\tilde{\alpha} - 0)$ and the lemma also provides $S_u(\tilde{\alpha} - 0) = S_{\tilde{u}}(\tilde{\alpha} - 0)$. Thus,

$$S_u(\tilde{\alpha} - 0) = S_{\tilde{u}}(\tilde{\alpha} - 0) = S_{\tilde{u}}(\tilde{\alpha}) = S_u(\tilde{\alpha}).$$

Analog arguments show $S_u(\tilde{\alpha} + 0) = S_u(\tilde{\alpha})$, that is, S_u is continuous in $\tilde{\alpha}$. □

By Proposition 4.17 the (at most countable) set of points of discontinuity coincides for all S_u, $u \in U$, and at the points of continuity all S_u are identical. If the S_u are not continuous at some point $\tilde{\alpha} > 0$, then, by Lemma 4.16, there are $u^1, u^2 \in U$ such that S_{u^1} is continuous from the left and S_{u^2} is continuous from the right in $\tilde{\alpha}$. For all other $u \in U$ Proposition 4.15 states $S_{u^1}(\tilde{\alpha}) \le S_u(\tilde{\alpha}) \le S_{u^2}(\tilde{\alpha})$.

The following example shows that discontinuous functions S_u really occur.

Example 4.18. Let $X = Y = Z = \mathbb{R}$ and define

$$S(y,z) := |y-z|, \quad \Omega(x) := x^2 + x, \quad F(x) := x^4 - 5x^2 - x, \quad z^\delta := -8.$$

Then $T_\alpha^{z^\delta}(x) = x^4 + (\alpha-5)x^2 + (\alpha-1)x + 8$ (cf. upper row in Figure 4.1). For $\alpha \neq 1$ the function $T_\alpha^{z^\delta}$ has exactly one global minimizer, but $T_1^{z^\delta}$ has two global minimizers: $X_1^{z^\delta} = \{-\sqrt{2}, \sqrt{2}\}$ (cf. lower row in Figure 4.1). Thus, $U = \{u^1, u^2\}$ with $u_1^1 = -\sqrt{2}$, $u_1^2 = \sqrt{2}$, and $u_\alpha^1 = u_\alpha^2$ for $\alpha \neq 1$. The corresponding functions S_{u^1} and S_{u^2} satisfy $S_{u^1}(1) = 2 + \sqrt{2}$ and $S_{u^2}(1) = 2 - \sqrt{2}$.

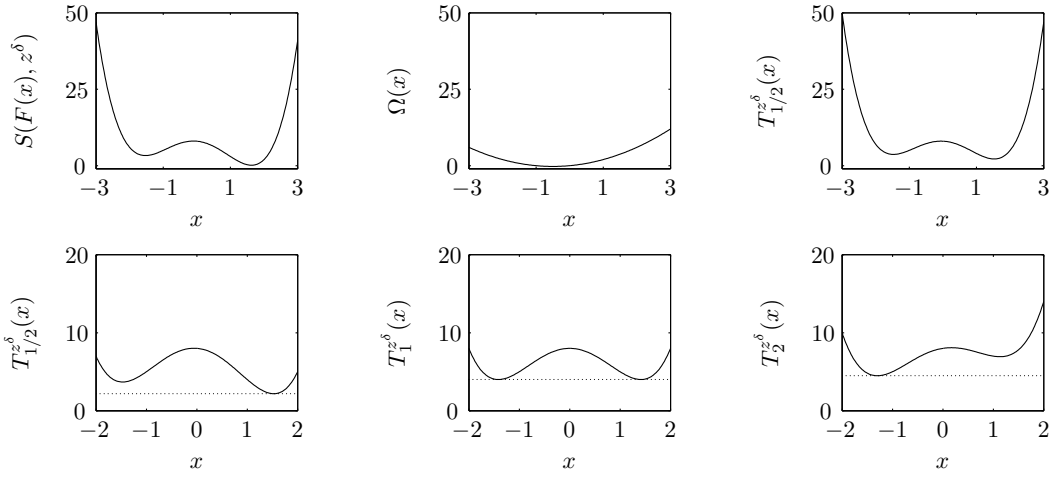

Figure 4.1.: Upper row from left to right: fitting functional, stabilizing functional, Tikhonov-type functional for $\alpha = \frac{1}{2}$. Lower row from left to right: Tikhonov-type functional for $\alpha = \frac{1}{2}$, $\alpha = 1$, $\alpha = 2$.

Since we know that the S_u, $u \in U$, differ only slightly, we set $S_* := S_u$ for an arbitrary $u \in U$. Then solving the discrepancy inequality is (nearly) equivalent to finding $\alpha \in (0, \infty)$ with

$$\psi(\delta) \leq S_*(\alpha) \leq c\psi(\delta). \tag{4.7}$$

Two questions remain to be answered: For which $\delta \geq 0$ inequality (4.7) has a solution? And if there is a solution, is it unique (at least for $c = 1$)?

To answer the first question we formulate the following proposition.

Proposition 4.19. *Let $D(\Omega) := \{x \in X : \Omega(x) < \infty\}$ and let $M(\Omega) := \{x_M \in X : \Omega(x_M) \leq \Omega(x)$ for all $x \in X$ with $S(F(x), z^\delta) < \infty\}$. Then we have*

$$\lim_{\alpha \to 0} S_*(\alpha) = \inf_{x \in D(\Omega)} S(F(x), z^\delta) \quad \text{and} \quad \lim_{\alpha \to \infty} S_*(\alpha) = \min_{x_M \in M(\Omega)} S(F(x_M), z^\delta).$$

Proof. For all $\alpha > 0$ and all $x \in X$ we have

$$\begin{aligned} S(F(x_\alpha^{z^\delta}), z^\delta) &= T_\alpha^{z^\delta}(x_\alpha^{z^\delta}) - \alpha\Omega(x_\alpha^{z^\delta}) \leq T_\alpha^{z^\delta}(x) - \alpha\Omega(x_\alpha^{z^\delta}) \\ &\leq S(F(x), z^\delta) + \alpha\big(\Omega(x) - \inf \Omega\big), \end{aligned}$$

that is, $\lim_{\alpha\to 0} S_*(\alpha) \le S(F(x), z^\delta)$ for all $x \in D(\Omega)$. On the other hand,

$$\inf_{x\in D(\Omega)} S(F(x), z^\delta) \le S_*(\alpha) \quad \text{for all } \alpha > 0,$$

which yields $\lim_{\alpha\to 0} S_*(\alpha) \ge \inf_{x\in D(\Omega)} S(F(x), z^\delta)$. Thus, the first assertion is true.

We prove the second one. Let $(\alpha_k)_{k\in\mathbb{N}}$ be an arbitrary sequence in $(0,\infty)$ with $\alpha_k \to \infty$ and take a sequence $(x_k)_{k\in\mathbb{N}}$ in X with $x_k \in \operatorname{argmin}_{x\in X} T_{\alpha_k}^{z^\delta}(x)$. Because

$$\Omega(x_k) \le \frac{1}{\alpha_k} T_{\alpha_k}^{z^\delta}(x_k) \le \frac{1}{\alpha_k} S(F(x^\dagger), z^\delta) + \Omega(x^\dagger) \to \Omega(x^\dagger),$$

the sequence (x_k) has a convergent subsequence. Let $(x_{k_l})_{l\in\mathbb{N}}$ be such a convergent subsequence and let $\tilde{x} \in X$ be one of its limits. Then

$$\Omega(\tilde{x}) \le \liminf_{l\to\infty} \Omega(x_{k_l}) \le \liminf_{l\to\infty} \Big(\frac{1}{\alpha_{k_l}} S(F(x), z^\delta) + \Omega(x) \Big) = \Omega(x)$$

for all $x \in X$ with $S(F(x), z^\delta) < \infty$, that is, $\tilde{x} \in M(\Omega)$. In addition, for each $x_M \in M(\Omega)$ we get

$$\begin{aligned} S(F(\tilde{x}), z) &\le \liminf_{l\to\infty} S(F(x_{k_l}), z^\delta) \le \limsup_{l\to\infty} S(F(x_{k_l}), z^\delta) \\ &= \limsup_{l\to\infty} \big(T_{\alpha_{k_l}}^{z^\delta}(x_{k_l}) - \alpha_{k_l}\Omega(x_{k_l}) \big) \le \limsup_{l\to\infty} \big(T_{\alpha_{k_l}}^{z^\delta}(x_{k_l}) - \alpha_{k_l}\Omega(x_M) \big) \\ &\le \limsup_{l\to\infty} S(F(x_M), z^\delta) = S(F(x_M), z^\delta). \end{aligned}$$

Thus, $S(F(\tilde{x}), z) = \min_{x_M\in M(\Omega)} S(F(x_M), z)$ and setting $x_M := \tilde{x}$ in the estimate above, we obtain $S(F(\tilde{x}), z) = \lim_{l\to\infty} S(F(x_{k_l}), z^\delta)$. Now one easily sees that

$$\lim_{k\to\infty} S(F(x_k), z^\delta) = \min_{x_M\in M(\Omega)} S(F(x_M), z).$$

Because this reasoning works for all sequences (α_k) with $\alpha_k \to \infty$, the second assertion of the proposition is true. □

Example 4.20. Consider the standard Hilbert space setting introduced in Example 1.1, that is, X and $Y = Z$ are Hilbert spaces, $F = A$ is a bounded linear operator, $S(y,z) = \frac{1}{2}\|y - z\|^2$, and $\Omega(x) = \frac{1}{2}\|x\|^2$. Then $T_\alpha^{z^\delta}$ has exactly one minimizer $x_\alpha^{z^\delta}$ for each $\alpha > 0$ and $S_*(\alpha) = \frac{1}{2}\|Ax_\alpha^{z^\delta} - z^\delta\|^2$ is continuous in α. In Proposition 4.19 we have $D(\Omega) = X$ and $M(\Omega) = \{0\}$. Thus, $\lim_{\alpha\to 0} S_*(\alpha)$ is the distance from z^δ to $\overline{\mathcal{R}(A)}$ and $\lim_{\alpha\to\infty} S_*(\alpha) = \frac{1}{2}\|z^\delta\|^2$. Details on the discrepancy principle in Hilbert spaces can be found, for instance, in [EHN96].

Proposition 4.19 shows that

$$\frac{1}{c} \inf_{x\in D(\Omega)} S(F(x), z^\delta) \le \psi(\delta) \le \min_{x_M\in M(\Omega)} S(F(x_M), z^\delta)$$

with $D(\Omega)$ and $M(\Omega)$ as defined in the proposition is a necessary condition for the solvability of (4.7).

If we in addition assume that the jumps of S_* are not too high, that is,

$$S_*(\alpha + 0) \leq cS(\alpha - 0) \quad \text{for all } \alpha > 0, \tag{4.8}$$

then (4.7) has a solution. This follows from the implication

$$\psi(\delta) > S_*(\alpha - 0) \quad \Rightarrow \quad c\psi(\delta) > cS_*(\alpha - 0) \geq S_*(\alpha + 0).$$

In other words, it is not possible that $\psi(\delta) > S_*(\alpha - 0)$ and $c\psi(\delta) < S_*(\alpha + 0)$. Note that the sum of all jumps in a finite interval is finite. Therefore, in the case of infinitely many jumps, the height of the jumps has to tend to zero. That is, there are only few relevant jumps.

Now we come to the second open question: assume that (4.7) has a solution for $c = 1$; could there be another solution? More precisely, are there $\alpha_1 \neq \alpha_2$ with $S_*(\alpha_1) = S_*(\alpha_2)$?

In general the answer is 'yes'. But the next proposition shows that in such a case it does not matter which solution is chosen.

Proposition 4.21. *Assume $0 < \alpha_1 < \alpha_2 < \infty$ and $S_*(\alpha_1) = S_*(\alpha_2)$. Further, let $x_{\alpha_1}^{z^\delta} \in X_{\alpha_1}^{z^\delta}$ such that $S(F(x_{\alpha_1}^{z^\delta}), z^\delta) = S_*(\alpha_1)$ and $x_{\alpha_2}^{z^\delta} \in X_{\alpha_2}^{z^\delta}$ such that $S(F(x_{\alpha_2}^{z^\delta}), z^\delta) = S_*(\alpha_2)$. Then $x_{\alpha_1}^{z^\delta} \in X_{\alpha_2}^{z^\delta}$ and $x_{\alpha_2}^{z^\delta} \in X_{\alpha_1}^{z^\delta}$.*

Proof. Because $S(F(x_{\alpha_1}^{z^\delta}), z^\delta) = S(F(x_{\alpha_2}^{z^\delta}), z^\delta)$, the inequality $T_{\alpha_1}^{z^\delta}(x_{\alpha_1}^{z^\delta}) \leq T_{\alpha_1}^{z^\delta}(x_{\alpha_2}^{z^\delta})$ implies $\Omega(x_{\alpha_1}^{z^\delta}) \leq \Omega(x_{\alpha_2}^{z^\delta})$. Together with Proposition 4.15 this shows $\Omega(x_{\alpha_1}^{z^\delta}) = \Omega(x_{\alpha_2}^{z^\delta})$. Thus, $T_{\alpha_1}^{z^\delta}(x_{\alpha_1}^{z^\delta}) = T_{\alpha_1}^{z^\delta}(x_{\alpha_2}^{z^\delta})$ and $T_{\alpha_2}^{z^\delta}(x_{\alpha_2}^{z^\delta}) = T_{\alpha_2}^{z^\delta}(x_{\alpha_1}^{z^\delta})$. □

4.3.3. Convergence and convergence rates

In this subsection we show that choosing the regularization parameter α according to the discrepancy principle yields convergence rates for the error measure $E_{x^\dagger}$ introduced in Subsection 4.1.2.

First we show that the regularized solutions converge to Ω-minimizing S-generalized solutions.

Corollary 4.22. *Let $(\delta_k)_{k\in\mathbb{N}}$ be a sequence in $[0, \infty)$ converging to zero, take an arbitrary sequence $(z_k)_{k\in\mathbb{N}}$ with $z_k \in Z_{y^0}^{\delta_k}$, and choose a sequence $(\alpha_k)_{k\in\mathbb{N}}$ in $(0, \infty)$ and a sequence $(x_k)_{k\in\mathbb{N}}$ in X with $x_k \in \operatorname{argmin}_{x\in X} T_{\alpha_k}^{z_k}(x)$ such that the discrepancy inequality (4.6) is satisfied. Then all the assertions of Theorem 3.4 (convergence) about subsequences of (x_k) and their limits are true.*

Proof. We show that the assumptions of Theorem 3.4 are satisfied for $y := y^0$ and $\bar{x} := x^\dagger$. Obviously $S(y^0, z_k) \to 0$ by Assumption 4.1 and $S(F(x_k), z_k) \leq c\psi(\delta_k) \to 0$ by (4.6). Thus, it only remains to show $\limsup_{k\to\infty} \Omega(x_k) \leq \Omega(\hat{x})$ for all S-generalized solutions $\hat{x}$. But this follows immediately from

$$\begin{aligned}\Omega(x_k) &= \frac{1}{\alpha_k}\big(T_{\alpha_k}^{z_k}(x_k) - S(F(x_k), z_k)\big) \\ &\leq \frac{1}{\alpha_k}\big(S(F(\hat{x}), z_k) - S(F(x_k), z_k)\big) + \Omega(\hat{x}) \leq \Omega(\hat{x}),\end{aligned}$$

where we used $S(F(\hat{x}), z_k) \leq \psi(\delta)$ and $S(F(x_k), z_k) \geq \psi(\delta)$. □

Before we come to the convergence rates result we want to give an example of a set M (on which a variational inequality (4.3) shall hold) satisfying Assumption 4.5. It is the same set as in Proposition 4.6.

Proposition 4.23. *Let $\bar{\alpha} > 0$ and $\varrho > \Omega(x^\dagger)$. If $\alpha = \alpha(\delta, z^\delta)$ is chosen according to the discrepancy principle* (4.6)*, then*

$$M := \{x \in X : S_Y(F(x), F(x^\dagger)) + \bar{\alpha}\Omega(x) \leq \varrho\bar{\alpha}\}$$

satisfies Assumption 4.5.

Proof. For the sake of brevity we write α instead of $\alpha(\delta, z^\delta)$. Set $\bar{\delta} > 0$ such that $\psi(\bar{\delta}) \leq \frac{\bar{\alpha}}{c+1}(\varrho - \Omega(x^\dagger))$ with $c > 1$ from (4.6). As in the proof of Corollary 4.22 we see $\Omega(x_\alpha^{z^\delta}) \leq \Omega(x^\dagger)$ for each $\delta \in (0, \bar{\delta}]$, each $z^\delta \in Z_{y^0}^\delta$, and each minimizer $x_\alpha^{z^\delta}$ of $T_\alpha^{z^\delta}$. Therefore, we have

$$\begin{aligned} S_Y(F(x_\alpha^{z^\delta}), F(x^\dagger)) + \bar{\alpha}\Omega(x_\alpha^{z^\delta}) &\leq S(F(x_\alpha^{z^\delta}), z^\delta) + S(F(x^\dagger), z^\delta) + \bar{\alpha}\Omega(x_\alpha^{z^\delta}) \\ &\leq (c+1)\psi(\delta) + \bar{\alpha}\Omega(x^\dagger) \leq (c+1)\psi(\bar{\delta}) + \bar{\alpha}\Omega(x^\dagger) \leq \bar{\alpha}\varrho, \end{aligned}$$

that is, $x_\alpha^{z^\delta} \in M$. □

The following theorem shows that choosing the regularization parameter according to the discrepancy principle and assuming that the fixed Ω-minimizing S-generalized solution $x^\dagger$ satisfies a variational inequality (4.3) we obtain the same rates as for the a priori parameter choice in Theorem 4.11. But, in contrast to Theorem 4.11, the convergence rates result based on the discrepancy principle works without Assumption 4.9, that is, this assumption is not an intrinsic prerequisite for obtaining convergence rates from variational inequalities.

Theorem 4.24. *Let $x^\dagger$ satisfy Assumption 4.7 and choose $\alpha = \alpha(\delta, z^\delta)$ and $x_{\alpha(\delta,z^\delta)}^{z^\delta} \in \operatorname{argmin}_{x\in X} T_{\alpha(\delta,z^\delta)}^{z^\delta}(x)$ for $\delta > 0$ and $z^\delta \in Z_{y^0}^\delta$ such that the discrepancy inequality* (4.6) *is satisfied. Further assume that M satisfies Assumption 4.5. Then there is some $\bar{\delta} > 0$ such that*

$$E_{x^\dagger}\big(x_{\alpha(\delta,z^\delta)}^{z^\delta}\big) \leq \frac{1}{\beta}\varphi\big((c+1)\psi(\delta)\big) \quad \text{for all } \delta \in (0, \bar{\delta}].$$

The constant $\beta > 0$ and the function φ come from Assumption 4.7. The constant $c > 1$ is from (4.6)*.*

Proof. We write α instead of $\alpha(\delta, z^\delta)$. By Assumption 4.5 we have $x_\alpha^{z^\delta} \in M$ for sufficiently small $\delta > 0$. Thus, Lemma 4.4 gives

$$\beta E_{x^\dagger}\big(x_\alpha^{z^\delta}\big) \leq \frac{1}{\alpha}\big(\psi(\delta) - S(F(x_\alpha^{z^\delta}), z^\delta)\big) + \varphi\big(S(F(x_\alpha^{z^\delta}), z^\delta) + \psi(\delta)\big).$$

By the left-hand inequality in (4.6) the first summand is nonpositive and by the right-hand one the second summand is bounded by $\varphi((c+1)\psi(\delta))$. □

5. Random data

In Section 1.1 we gave a heuristic motivation for considering minimizers of (1.3) as approximate solutions to (1.1). The investigation of such Tikhonov-type minimization problems culminated in the convergence rates theorems 4.11 and 4.24. Both theorems provide bounds for the solution error $E_{x^\dagger}\big(x^{z^\delta}_{\alpha(\delta,z^\delta)}\big)$ in terms of an upper bound δ for the data error $D_{y^0}(z)$. That setting is completely deterministic, that is, the influence of (random) noise on the data is solely controlled by the bound δ.

In the present chapter we give another motivation for minimizing Tikhonov-type functionals. This second approach is based on the idea of MAP estimation (maximum a posteriori probability estimation) introduced in Section 5.1 and pays more attention to the random nature of noise. A major benefit is that the MAP approach yields a fitting functional S suited to the probability distribution governing the data.

We do not want to go into the details of statistical inverse problems (see, e.g., [KS05]) here. The aims of this chapter are to show that the MAP approach can be made precise also in a very general setting and that the tools for obtaining deterministic convergence rates also work in a stochastic framework.

5.1. MAP estimation

The method of maximum a posteriori probability estimation is an elegant way for motivating the minimization of various Tikhonov-type functionals. It comes from *statistical inversion theory* and is well-known in the inverse problems community. But usually MAP estimation is considered in a finite-dimensional setting (see, e.g., [KS05, Chapter 3] and [BB09]). To show that the approach can be rigorously justified also in our very general (infinite-dimensional) framework, we give a detailed description of it.

Note that Chapter C in the appendix contains all necessary information on random variables taking values in topological spaces. Especially questions arising when using conditional probability densities in such a general setting as ours are addressed there.

5.1.1. The idea

The basic idea is to treat all relevant quantities as random variables over a common probability space $(\Theta, \mathcal{A}, P)$, where $\mathcal{A}$ is a σ-algebra over $\Theta \neq \emptyset$ and $P : \mathcal{A} \to [0,1]$ is a probability measure on $\mathcal{A}$. For this purpose we equip the spaces X, Y, and Z with their Borel-σ-algebras $\mathcal{B}_X$, $\mathcal{B}_Y$, and $\mathcal{B}_Z$. In our setting the relevant quantities are the variables $x \in X$ (solution), $y \in Y$ (right-hand side), and $z \in Z$ (data). By $\xi : \Theta \to X$, $\eta : \Theta \to Y$, and $\zeta : \Theta \to Z$ we denote the corresponding random variables. Since η is determined by $F(\xi) = \eta$, it plays only a minor role and the information of interest about an outcome $\theta \in \Theta$ of the experiment will be contained in $\xi(\theta)$ and $\zeta(\theta)$. After appropriately modeling the three components of the probability space $(\Theta, \mathcal{A}, P)$,

Bayes' formula allows us to maximize with respect to $\theta \in \Theta$ the probability that $\xi(\theta)$ is observed knowing that $\zeta(\theta)$ coincides with some fixed measurement $z \in Z$. Such a maximization problem can equivalently be written as the minimization over $x \in X$ of a Tikhonov-type functional.

5.1.2. Modeling the propability space

The main point in modeling the probability space $(\Theta, \mathcal{A}, P)$ is the choice of the probability measure P, but first we have to think about Θ and $\mathcal{A}$. Assuming that nothing is known about the connection between ξ and ζ, each pair (x, z) of elements $x \in X$ and $z \in Z$ may be an outcome of our experiment, that is, for each such pair there should exist some $\theta \in \Theta$ with $x = \xi(\theta)$ and $z = \zeta(\theta)$. Since ξ and ζ are the only random variables of interest, $\Theta := X \times Z$ is a sufficiently ample sampling set. We choose $\mathcal{A} := \mathcal{B}_X \otimes \mathcal{B}_Z$ to be the corresponding product σ-algebra and we define ξ and ζ by $\xi(\theta) := x_\theta$ and $\zeta(\theta) := z_\theta$ for $\theta = (x_\theta, z_\theta) \in \Theta$.

The probability measure P can be compiled of two components: a weighting of the elements of X and the description of the dependence of ζ on ξ. Both components are accessible in practice. The aim of the first one, the weighting, is to incorporate desirable or a priori known properties of solutions of the underlying equation (1.1) into the model. This weighting can be realized by prescribing the probability distribution of ξ. A common approach, especially having Tikhonov-type functionals in mind, is to use a functional $\Omega : X \to (-\infty, \infty]$ assigning small, possibly negative values to favorable elements x and high values to unfavorable ones. Then, given a measure μ_X on $(X, \mathcal{B}_X)$ and assuming that $0 < \int_X \exp(-\alpha\Omega(\bullet)) \, \mathrm{d}\mu_X < \infty$ (we set $\exp(-\infty) := 0$), the density

$$p_\xi(x) := c \exp(-\alpha\Omega(x)) \quad \text{for } x \in X$$

with $c := \left(\int_X \exp(-\alpha\Omega(\bullet)) \, \mathrm{d}\mu_X\right)^{-1}$ realizes the intended weighting. The parameter $\alpha \in (0, \infty)$ allows to scale the strength of the weighting. Assuming that Ω has τ_X-closed sublevel sets (cf. item (v) of Assumption 2.1) Ω is measurable with respect to $\mathcal{B}_X$. Thus, $c \exp(-\alpha\Omega(\bullet))$ is measureable, too.

The second component in modeling the probability measure P is to prescribe the conditional probability that a data element $z \in Z$ is observed, that is, $\zeta(\theta) = z$, if $\xi(\theta) = x$ is known. In terms of densities this corresponds to prescribing the conditional density $p_{\zeta|\xi=x}$ with respect to some measure μ_Z on $(Z, \mathcal{B}_Z)$. Here, on the one hand we have to incorporate the underlying equation (1.1) and on the other hand we have to take the measurement process yielding the data z into account. Since the data depends only on $F(x)$, and not directly on x, we have

$$p_{\zeta|\xi=x}(z) = p_{\zeta|\eta=F(x)}(z) \quad \text{for } x \in X \text{ and } z \in Z. \tag{5.1}$$

The conditional density $p_{\zeta|\eta=y}$ for $y \in Y$ has to be chosen according to the measurement process.

Assuming that μ_X and μ_Z, and thus also the product measure $\mu_X \otimes \mu_Z$, are σ-finite and that the functional $p_{(\xi,\zeta)} : X \times Z \to [0, \infty)$ defined by

$$p_{(\xi,\zeta)}(x, z) := p_{\zeta|\xi=x}(z) p_\xi(x) \quad \text{for } x \in X \text{ and } z \in Z \tag{5.2}$$

is measurable with respect to the product σ-algebra $\mathcal{B}_X \otimes \mathcal{B}_Z$ on $X \times Z$, there exists a probability measure having $p_{(\xi,\zeta)}$ as a density with respect to $\mu_X \otimes \mu_Z$. We define P to be this probability measure. Equation (C.3), being a consequence of Proposition C.10, guarantees that the notations p_ξ and $p_{\zeta|\xi=x}$ are chosen appropriately.

Note that Section C.2 in the Appendix discusses the question why working with densities is more favorable than directly working with the probability measure P. To guarantee the existence of conditional distributions as the basis of conditional densities we assume that $(X, \mathcal{B}_X)$ and $(Z, \mathcal{B}_Z)$ are Borel spaces (cf. Definition C.6, Proposition C.7, and Theorem C.8).

5.1.3. A Tikhonov-type minimization problem

Now, that we have a probability space, we can attack the problem of maximizing with respect to $\theta \in \Theta$ the probability that $\xi(\theta)$ is observed knowing that $\zeta(\theta)$ coincides with some fixed measurement $z \in Z$. This probability is expressed by the conditional density $p_{\xi|\zeta=z}(x)$, that is, we want to maximize $p_{\xi|\zeta=z}(x)$ over $x \in X$ for some fixed $z \in Z$.

For $x \in X$ with $p_\xi(x) > 0$ and $z \in Z$ with $p_\zeta(z) > 0$ Bayes' formula (C.2) states

$$p_{\xi|\zeta=z}(x) = \frac{p_{\zeta|\xi=x}(z) p_\xi(x)}{p_\zeta(z)}.$$

If $p_\zeta(z) = 0$, then $p_{\xi|\zeta=z}(x) = p_\xi(x)$ for all $x \in X$ by Proposition C.10. In the case $p_\xi(x) = 0$ and $p_\zeta(z) > 0$ the same proposition provides

$$p_{\xi|\zeta=z}(x) = \frac{p_{(\xi,\zeta)}(x,z)}{p_\zeta(z)}$$

and (5.2) shows that the numerator is zero. Thus, $p_{\xi|\zeta=z}(x) = 0$. Summarizing the three cases and using equality (5.1) we obtain

$$p_{\xi|\zeta=z}(x) = \begin{cases} p_\xi(x), & \text{if } p_\zeta(z) = 0, \\ \dfrac{p_{\zeta|\eta=F(x)}(z) p_\xi(x)}{p_\zeta(z)}, & \text{if } p_\zeta(z) > 0. \end{cases} \tag{5.3}$$

We now transform the maximization of $p_{\xi|\zeta=z}$ into an equivalent minimization problem. First observe

$$\operatorname*{argmax}_{x \in X} p_{\xi|\zeta=z}(x) = \operatorname*{argmin}_{x \in X} \bigl(-\ln p_{\xi|\zeta=z}(x)\bigr)$$

(with $-\ln(0) := \infty$) and

$$-\ln p_{\xi|\zeta=z}(x) = \begin{cases} -\ln c + \alpha\Omega(x), & \text{if } p_\zeta(z) = 0, \\ -\ln p_{\zeta|\eta=F(x)}(z) + \ln p_\zeta(z) - \ln c + \alpha\Omega(x), & \text{if } p_\zeta(z) > 0. \end{cases}$$

In the case $p_\zeta(z) = 0$ we can add $\ln c$ without changing the minimizers. The resulting objective function is $\alpha\Omega(x)$. If $p_\zeta(z) > 0$ we can substract $\ln p_\zeta(z) - \ln c$ and also the

infimum of $-\ln p_{\zeta|\eta=y}(z)$ over $y \in Y$ without changing the minimizers, where we assume that the infimum is not $-\infty$. This results in the objective function

$$-\ln p_{\zeta|\eta=F(x)}(z) - \inf_{y\in Y}\big(-\ln p_{\zeta|\eta=y}(z)\big) + \alpha\Omega(x).$$

Setting

$$S(y,z) := \begin{cases} 0, & \text{if } p_\zeta(z) = 0, \\ -\ln p_{\zeta|\eta=y}(z) - \inf_{\tilde{y}\in Y}\big(-\ln p_{\zeta|\eta=\tilde{y}}(z)\big), & \text{if } p_\zeta(z) > 0 \end{cases} \tag{5.4}$$

for all $y \in Y$ and all $z \in Z$ we thus obtain

$$\operatorname*{argmax}_{x\in X} p_{\xi|\zeta=z}(x) = \operatorname*{argmin}_{x\in X}\big(S(F(x),y) + \alpha\Omega(x)\big) \quad \text{for all } z \in Z.$$

That is, the maximization of the conditional probability density $p_{\xi|\zeta=z}$ is equivalent to the Tikhonov-type minimization problem (1.3) with a specific fitting functional S. Note that by construction $S(y,z) \geq 0$ for all $y \in Y$ and $z \in Z$.

5.1.4. Example

To illustrate the MAP approach we give a simple finite-dimensional example. A more complex one is the subject of Part II of this thesis.

Let $X := \mathbb{R}^n$ and $Y = Z := \mathbb{R}^m$ be Euclidean spaces, let μ_X and μ_Z be the corresponding Lebesgue measures, and denote the Euclidean norm by $\|\bullet\|$. We define the weighting functional Ω by $\Omega(x) := \frac{1}{2}\|x\|^2$, that is,

$$p_\xi(x) = \left(\frac{\alpha}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\alpha}{2}\sum_{j=1}^{n} x_j^2\right)$$

is the density of the n-dimensional Gauss distribution with mean vector $(0,\dots,0)$ and covariance matrix $\operatorname{diag}\big(\frac{1}{\alpha},\dots,\frac{1}{\alpha}\big)$. Given a right-hand side $y \in Y$ of (1.1) the data elements $z \in Z$ shall follow the m-dimensional Gauss distribution with mean vector y and covariance matrix $\operatorname{diag}(\sigma^2,\dots,\sigma^2)$ for a fixed constant $\sigma > 0$. Thus,

$$p_{\zeta|\eta=y}(z) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} \exp\left(-\sum_{i=1}^{m}\frac{(z_i - y_i)^2}{2\sigma^2}\right).$$

From $p_\xi(x) > 0$ for all $x \in X$ and $p_{\zeta|\eta=y}(z) > 0$ for all $y \in Y$ and $z \in Z$ we see $p_\zeta(z) > 0$ for all $z \in Z$. Thus, the fitting functional S defined by (5.4) becomes $S(y,z) = \frac{1}{2\sigma^2}\|F(x) - z\|^2$ and the whole Tikhonov-type functional reads as

$$\frac{1}{2\sigma^2}\|F(x) - z\|^2 + \frac{\alpha}{2}\|x\|^2.$$

This functional corresponds to the standard Tikhonov method (in finite-dimensional spaces).

5.2. Convergence rates

Contrary to the situation of Chapter 4, in the stochastic setting of the present chapter we have no upper bound for the data error. In fact, we only assume that data elements are close to the exact right-hand side with high probability and far away from it with low probability. Thus, we cannot derive estimates for the solution error $E_{x^\dagger}(x_\alpha^z)$ (cf. Section 4.1) in terms of some noise level δ. Instead we provide lower bounds for the probability that we observe a data element z for which all corresponding regularized solutions x_α^z lie in a small ball around $x^\dagger$. Here, as before, $x^\dagger \in X$ is a fixed Ω-minimizing S-generalized solution to (1.1) with fixed right-hand side $y^0 \in Y$ and we assume $\Omega(x^\dagger) < \infty$.

For $\alpha > 0$ and $\varepsilon > 0$ we define the set

$$Z_\alpha^\varepsilon(x^\dagger) := \{z \in Z : E_{x^\dagger}(x_\alpha^z) \leq \varepsilon \text{ for all } x_\alpha^z \in \operatorname{argmin} T_\alpha^z\}$$

and we are interested in the probability that an observed data element belongs to this set. The following lemma gives a sufficient condition for the measurability of the sets $Z_\alpha^\varepsilon(x^\dagger)$ with respect to the Borel-σ-algebra $\mathcal{B}_Z$ of (Z, τ_Z).

Lemma 5.1. *Let $E_{x^\dagger} : X \to [0, \infty]$ be lower semi-continuous and assume that the Tikhonov-type functional T_α^z has only one global minimizer for all $\alpha > 0$ and all $z \in Z$. Then for each $\alpha > 0$ and each $\varepsilon > 0$ the set $Z_\alpha^\varepsilon(x^\dagger)$ is closed.*

Proof. Let $(z_k)_{k\in\mathbb{N}}$ be a sequence in $Z_\alpha^\varepsilon(x^\dagger)$ converging to some $z \in Z$. We have to show that $z \in Z_\alpha^\varepsilon(x^\dagger)$, too. For each z_k the corresponding minimizer $x_k \in \operatorname{argmin} T_\alpha^{z_k}$ satisfies $E_{x^\dagger}(x_k) \leq \varepsilon$ and by Theorem 3.3 (stability) the sequence $(x_k)_{k\in\mathbb{N}}$ converges to the minimizer $\tilde{x} \in \operatorname{argmin} T_\alpha^z$. Therefore the lower semi-continuity of $E_{x^\dagger}$ implies $E_{x^\dagger}(\tilde{x}) \leq \liminf_{k\to\infty} E_{x^\dagger}(x_k) \leq \varepsilon$, that is, $z \in Z_\alpha^\varepsilon(x^\dagger)$. □

From now on we assume that the sets $Z_\alpha^\varepsilon(x^\dagger)$ are measurable.

By $P_{\zeta|\xi=x^\dagger}(A)$ we denote the probability of an event $A \in \mathcal{B}_Z$ conditioned on $\xi = x^\dagger$, that is,

$$P_{\zeta|\xi=x^\dagger}(A) = \int_A p_{\zeta|\xi=x^\dagger} \, \mathrm{d}\mu_z.$$

Instead of seeking upper bounds for the solution error $E_{x^\dagger}(x_\alpha^z)$ as done in the deterministic setting of Chapter 4, we are interested in lower bounds for the probability $P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger))$. The higher this probability the better is the chance to obtain well approximating regularized solutions from the measured data.

Lower bounds for $P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger))$ can be derived in analogy to the upper bounds for $E_{x^\dagger}\big(x_{\alpha(\delta)}^{z^\delta}\big)$ in Section 4.2. We only have to replace expressions involving δ by suitable expressions in ε.

The analog of Assumption 4.5 reads as follows.

Assumption 5.2. Given a parameter choice $\varepsilon \mapsto \alpha(\varepsilon)$ let $M \subseteq X$ be a set such that $S_Y(F(x), F(x^\dagger)) < \infty$ for all $x \in M$ and such that there is some $\bar{\varepsilon} > 0$ with

$$\bigcup_{z \in Z_{\alpha(\varepsilon)}^\varepsilon(x^\dagger)} \operatorname*{argmin}_{x\in X} T_{\alpha(\varepsilon)}^z(x) \subseteq M \quad \text{for all } \varepsilon \in (0, \bar{\varepsilon}].$$

For example the set $M = \{x \in X : E_{x^\dagger}(x) \leq \bar{\varepsilon},\ S_Y(F(x), F(x^\dagger)) < \infty\}$ satisfies Assumption 5.2 for each parameter choice $\alpha \mapsto \alpha(\varepsilon)$.

The following lemma is the analog of Lemma 4.10 and provides a basic estimate for $P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger))$.

Lemma 5.3. *Let $x^\dagger$ satisfy Assumption 4.7, let $\alpha > 0$ and $\varepsilon \geq 0$, and assume $\operatorname{argmin} T_\alpha^z \subseteq M$ for all $z \in Z_\alpha^\varepsilon(x^\dagger)$. Then*

$$P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger)) \geq P_{\zeta|\xi=x^\dagger}\left(\left\{z \in Z : 2S(F(x^\dagger), z) \leq \beta\alpha\varepsilon - \alpha(-\varphi)^*\Big(-\frac{1}{\alpha}\Big)\right\}\right).$$

Proof. Using the minimizing property of x_α^z the variational inequality (4.3) implies

$$\begin{aligned}
\beta E_{x^\dagger}(x_\alpha^z) &\leq \Omega(x_\alpha^z) - \Omega(x^\dagger) + \varphi\big(S_Y(F(x_\alpha^z), F(x^\dagger))\big) \\
&= \frac{1}{\alpha}T_\alpha^z(x_\alpha^z) - \Omega(x^\dagger) - \frac{1}{\alpha}S(F(x_\alpha^z), z) + \varphi\big(S_Y(F(x_\alpha^z), F(x^\dagger))\big) \\
&\leq \frac{1}{\alpha}\big(S(F(x^\dagger), z) - S(F(x_\alpha^z), z)\big) + \varphi\big(S(F(x_\alpha^z), z) + S(F(x^\dagger), z)\big) \\
&\leq \frac{2}{\alpha}S(F(x^\dagger), z) + \sup_{t\geq 0}\left(\varphi(t) - \frac{1}{\alpha}t\right) = \frac{2}{\alpha}S(F(x^\dagger), z) + (-\varphi)^*\Big(-\frac{1}{\alpha}\Big)
\end{aligned}$$

for all $z \in Z_\alpha^\varepsilon(x^\dagger)$ and all corresponding regularized solutions x_α^z. Thus,

$$\begin{aligned}
Z_\alpha^\varepsilon(x^\dagger) &\supseteq \left\{z \in Z : \frac{2}{\alpha}S(F(x^\dagger), z) + (-\varphi)^*\Big(-\frac{1}{\alpha}\Big) \leq \beta\varepsilon\right\} \\
&= \left\{z \in Z : 2S(F(x^\dagger), z) \leq \beta\alpha\varepsilon - \alpha(-\varphi)^*\Big(-\frac{1}{\alpha}\Big)\right\},
\end{aligned}$$

proving the assertion. □

The following theorem provides a lower bound for $P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger))$ which can be realized by choosing the regularization parameter α in dependence on ε. The theorem can be proven in analogy to Theorem 4.11 if $\psi(\delta)$ is replaced by $\varphi^{-1}(\beta\varepsilon)$, where β comes from Assumption 4.7.

Theorem 5.4. *Let $x^\dagger$ satisfy Assumption 4.7 such that the associated function φ satisfies Assumption 4.9 and let $\varepsilon \mapsto \alpha(\varepsilon)$ be a parameter choice such that*

$$\inf_{\tau\in[0,\varphi^{-1}(\beta\varepsilon))} \frac{\beta\varepsilon - \varphi(\tau)}{\varphi^{-1}(\beta\varepsilon) - \tau} \geq \frac{1}{\alpha(\varepsilon)} \geq \sup_{\tau\in(\varphi^{-1}(\beta\varepsilon),\gamma]} \frac{\varphi(\tau) - \beta\varepsilon}{\tau - \varphi^{-1}(\beta\varepsilon)} \tag{5.5}$$

for all $\varepsilon \in \big(0, \frac{1}{\beta}\varphi(\gamma)\big)$, where γ comes from Assumption 4.9 and β from Assumption 4.7. Further, let M satisfy Assumption 5.2. Then there is some $\bar{\varepsilon} > 0$ such that

$$P_{\zeta|\xi=x^\dagger}\big(Z_{\alpha(\varepsilon)}^\varepsilon(x^\dagger)\big) \geq P_{\zeta|\xi=x^\dagger}\left(\left\{z \in Z : \frac{1}{\beta}\varphi\big(2S(F(x^\dagger), z)\big) \leq \varepsilon\right\}\right) \quad \text{for all } \varepsilon \in (0, \bar{\varepsilon}].$$

Remark 5.5. By item (ii) of Assumption 4.9 the function φ is invertible as a mapping from $[0, \gamma]$ onto $[0, \varphi(\gamma)]$. Thus $\varphi^{-1}(\beta\varepsilon)$ is well-defined for $\varepsilon \in \big[0, \frac{1}{\beta}\varphi(\gamma)\big]$. In addition, Remarks 4.12 and 4.13 apply if $\psi(\delta)$ is replaced by $\varphi^{-1}(\beta\varepsilon)$.

Proof of Theorem 5.4. We write α instead of $\alpha(\varepsilon)$. From Assumption 5.2 we know $\operatorname{argmin} T_\alpha^z \subseteq M$ for all $z \in Z_\alpha^\varepsilon(x^\dagger)$ if $\varepsilon > 0$ is sufficiently small. Thus, Lemma 5.3 gives

$$P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger)) \geq P_{\zeta|\xi=x^\dagger}\left(\left\{z \in Z : 2S(F(x^\dagger), z) \leq \beta\alpha\varepsilon - \alpha(-\varphi)^*\Big(-\frac{1}{\alpha}\Big)\right\}\right).$$

In analogy to the proof of Theorem 4.11 we can show

$$\varphi(\tau) - \frac{1}{\alpha}\tau \leq \beta\varepsilon - \frac{1}{\alpha}\varphi^{-1}(\beta\varepsilon) \quad \text{for all } \tau \geq 0$$

if ε is small enough (replace $\psi(\delta)$ by $\varphi^{-1}(\beta\varepsilon)$). And using this inequality multiplied by $-\alpha$ we obtain

$$\beta\alpha\varepsilon - \alpha(-\varphi)^*\Big(-\frac{1}{\alpha}\Big) \geq \varphi^{-1}(\beta\varepsilon),$$

yielding

$$P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger)) \geq P_{\zeta|\xi=x^\dagger}\big(\{z \in Z : 2S(F(x^\dagger), z) \leq \varphi^{-1}(\beta\varepsilon)\}\big),$$

which is equivalent to the assertion. □

Note that is the case $\varphi(t) \leq ct$ for some $c > 0$ and all $t \in [0, \infty)$ one can show

$$P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger)) \geq P_{\zeta|\xi=x^\dagger}\left(\left\{z \in Z : \frac{2}{\beta\alpha}S(F(x^\dagger), z) \leq \varepsilon\right\}\right)$$

if $Z_\alpha^\varepsilon(x^\dagger) \subseteq M$ and $\alpha \in \big(0, \frac{1}{c}\big]$ (cf. Proposition 4.14 for details).

Inspecting the lower bound in Theorem 5.4 we see that the faster the function φ decays to zero if the argument goes to zero the faster the set $\big\{z \in Z : \frac{1}{\beta}\varphi\big(2S(F(x^\dagger), z)\big) \leq \varepsilon\big\}$ expands if $\varepsilon \to 0$. Thus, the faster a function φ in a variational inequality (4.3) decays to zero if the argument goes to zero the faster the lower bound for $P_{\zeta|\xi=x^\dagger}(Z_\alpha^\varepsilon(x^\dagger))$ increases if $\varepsilon \to 0$. In other words, if $x^\dagger$ satisfies a variational inequality with a fast decaying function φ, then there is a high probability that the regularized solutions x_α^z corresponding to an observed data element z are close to the exact solution $x^\dagger$.

Part II.

An example: Regularization with Poisson distributed data

6. Introduction

The major noise model in literature on ill-posed inverse problems is Gaussian noise, because for many applications this noise distribution appears quite natural. Due to the fact that more precise noise models promise better reconstruction results from noisy data the interest in advanced approaches for modeling the noisy data is rapidly growing.

One non-standard noise model shall be considered in the present part of the thesis. Poisson distributed data, sometimes also referred to as data corrupted by Poisson noise, occurs especially in imaging applications where the intensity of light or other radiation to be detected is very low. One important application in medical imaging is *positron emission tomography* (PET), where the decay of a radioactive fluid in a human body is recorded. The decay events are recorded in a way such that instead of the concrete point one only knows a straight line through the body where the decay took place. For each line through the body one counts the number of decay events in a fixed time interval. Since the radiation dose has to be very small, the number of decays is small, too. Thus, one may assume that the number of decay events per time interval follows a Poisson distribution. From the mathematical point of view reconstructing PET images consists in inverting the Radon transform with Poisson distributed data. For more information on PET we refer to the literature (e.g. [Eps08])

Other examples for ill-posed imaging problems with Poisson distributed data can be found in the field of astronomical imaging. There, deconvolution and denoising of images which are mostly black with only few bright spots (the stars) are important tasks. As a third application we mention confocal laser scanning microscopy (see [Wil]).

The present part of the thesis is based on Part I. Thus, we use the same notation as in the first part without further notice.

The structure of this part is as follows: at first we motivate a noise adapted Tikhonov-type functional (Chapter 7), which then is specified to a semi-discrete and to a continuous model for Poisson distributed data in Chapters 8 and 9, respectively. In the last chapter, Chapter 10, we present an algorithm for minimizing such Poisson noise adapted Tikhonov-type functionals and compare the results with the results obtained from the usual (Gaussian noise adapted) Tikhonov method.

A preliminary version of some results presented in the next three chapters of this part (that is, excluding Chapter 10) has already been published in [Fle10a].

7. The Tikhonov-type functional

7.1. MAP estimation for imaging problems

We want to apply the considerations of Chapter 5 to the typical setting in imaging applications. Let (X, τ_X) be an arbitrary topological space and let $Y := \{y \in L^1(T, \mu) : y \geq 0 \text{ a.e.}\}$ be the space of real-valued images over $T \subseteq \mathbb{R}^d$ which are integrable with respect to the measure μ (here we implicitly assume that T is equipped with a σ-algebra $\mathcal{A}_T$ on which μ is defined). Usually μ is a multiple of the Lebesgue measure. Assume $\mu(T) < \infty$. The set T on the one hand is the domain of the images and on the other hand it models the surface of the image sensor capturing the image. We equip the space Y with the topology τ_Y induced by the weak $L^1(T, \mu)$-topology.

The operator F assigns to each element $x \in X$ an image $F(x) \in Y$. We assume that the image sensor is a collection of $m \in \mathbb{N}$ sensor cells or pixels $T_1, \dots, T_m \subseteq T$. If we think of an image $y \in Y$ to be an energy density describing the intensity of the image, then each sensor cell counts the number of particles, for example photons, emitted according to the density y and impinging on the cell. The sensor electronics amplify the weak signal caused by the impinging particles, which leads to a scaling and usually also to noise (for details see, e.g., [How06, Chapter 2]). In addition, some preprocessing software could scale the signal to fit into a certain interval. Thus, the data we obtain from a sensor cell is a nonnegative real number, and looking at a large number of measurements we will observe that the values cluster around the points aw for $w \in \mathbb{N}_0$ with a scaling factor $a > 0$. Note that in practice the factor $a > 0$ can be measured, see [How06, Section 3.8]. As a consequence of these considerations we choose $Z := [0, \infty)^m$, although the particle counts are natural numbers. The topology τ_Z will be specified later.

The conditional density $p_{\zeta|\eta=y}$ describing the dependence of the data on the right-hand side $y \in Y$ has to be modeled to represent the specifics of the capturing process of a concrete imaging problem. From the considerations above we only know the mean $Dy \in [0, \infty)^m$ of the data with $D : Y \to [0, \infty)$ given by $D = (D_1, \dots, D_m)$ and

$$D_i y := \int_{T_i} y \, \mathrm{d}\mu$$

for $y \in Y$ and $i = 1, \dots, m$.

Assume that the components $\zeta_i : \Theta \to [0, \infty)$ of ζ are mutually independent and that $p_{\zeta_i|\eta=y}$ can be written as $p_{\zeta_i|\eta=y}(z_i) = g(D_i y, z_i)$ for all $z_i \in [0, \infty)$ with a function $g : [0, \infty) \times [0, \infty) \to [0, \infty)$, that is, $p_{\zeta_i|\eta=y}$ does not depend directly on y but only on $D_i y$. Then we have

$$p_{\zeta|\eta=y}(z) = \prod_{i=1}^{m} g(D_i y, z_i) \quad \text{for all } y \in Y \text{ and } z \in Z \tag{7.1}$$

and the fitting functional S in (5.4) reads as

$$S(y,z) = \begin{cases} 0, & \text{if } p_\zeta(z) = 0, \\ \displaystyle\sum_{i=1}^{m} \Big(-\ln g(D_i y, z_i) - \inf_{v \in [0,\infty)} \big(-\ln g(v, z_i) \big) \Big), & \text{if } p_\zeta(z) > 0. \end{cases}$$

The setting introduced so far will be referred to as *semi-discrete setting*, because the data space Z is finite-dimensional. We can also derive a completely continuous model by temporarily setting $T := \{1, \dots, m\} \subseteq \mathbb{N}$ and μ to be the counting measure on T. Then $Y = [0,\infty)^m$ and $T_i := \{i\}$ leads to $D_i y = y_i$. Regarding y and z as functions over T and writing the sum as an integral over T the fitting functional becomes

$$S(y,z) = \begin{cases} 0, & \text{if } p_\zeta(z) = 0, \\ \displaystyle\int_T \Big(-\ln g(y(t), z(t)) - \inf_{v \in [0,\infty)} \big(-\ln g(v, z(t)) \big) \Big) \, \mathrm{d}\mu(t), & \text{if } p_\zeta(z) > 0. \end{cases}$$

This expression also works for arbitrary T and $Z := Y$ (with Y being the set of nonnegative $L^1(T,\mu)$-functions introduced above) if the infimum is replaced by the essential infimum:

$$S(y,z) = \begin{cases} 0, & \text{if } p_\zeta(z) = 0, \\ \displaystyle\int_T \Big(-\ln g(y(t), z(t)) - \operatorname*{ess\,inf}_{v \in [0,\infty)} \big(-\ln g(v, z(t)) \big) \Big) \, \mathrm{d}\mu(t), & \text{if } p_\zeta(z) > 0. \end{cases}$$

This *continuous setting* has no direct motivation from practice, but it can be regarded as a suitable model for images with very high resolution, that is, if m is very large.

In this thesis we analyze both the semi-discrete and the continuous model. Especially in case of the continuous model we have to ensure that the fitting functional is well-defined (measurability of the integrand). This question will be addressed later when considering a concrete function g.

7.2. Poisson distributed data

In this section we first restrict our attention to data vectors $\tilde{z} \in \mathbb{N}_0^m$ for motivating the use of a certain fitting functional. Then we extend the considerations to data vectors $z \in [0,\infty)^m$ clustering around the vectors $a\tilde{z}$ with a scaling factor $a > 0$ (cf. Section 7.1).

If the number $\tilde{z}_i \in \mathbb{N}_0$ of particles impinging on the sensor cell T_i follows a Poisson distribution with mean $D_i y$ for a given right-hand side $y \in Y$, then

$$g(v,w) := \begin{cases} 1, & \text{if } v = 0,\ w = 0, \\ 0, & \text{if } v = 0,\ w > 0, \\ \dfrac{v^w}{w!} \exp(-v), & \text{if } v > 0 \end{cases}$$

for $v \in [0,\infty)$ and $w \in \mathbb{N}_0$ defines the density $p_{\zeta|\eta=y}$ on $\mathbb{N}_0^m$ with respect to the counting

measure on $\mathbb{N}_0^m$ (cf. (7.1)). Thus, setting

$$s(v,w) := \begin{cases} 0, & \text{if } v = 0,\ w = 0, \\ \infty, & \text{if } v = 0,\ w > 0, \\ v, & \text{if } v > 0,\ w = 0, \\ w \ln \dfrac{w}{v} + v - w, & \text{if } v > 0,\ w > 0, \end{cases} \tag{7.2}$$

the fitting functional reads as $S(y,\tilde{z}) = \sum_{i=1}^m s(D_i y, \tilde{z}_i)$ for all $y \in Y$ and all $\tilde{z} \in \mathbb{N}_0^m$ satisfying $p_\zeta(\tilde{z}) > 0$.

Lemma 7.1. *The function $s : [0,\infty)\times[0,\infty) \to (-\infty,\infty]$ defined by (7.2) is convex and lower semi-continuous and it is continuous outside $(0,0)$ (with respect to the natural topologies). Further, $s(v,w) \geq 0$ for all $v,w \in [0,\infty)$ and $s(v,w) = 0$ if and only if $v = w$.*

Proof. Define the auxiliary function $\tilde{s}(u) := u \ln u - u + 1$ for $u \in (0,\infty)$ and set $\tilde{s}(0) := 1$. This function if strictly convex and continuous on $[0,\infty)$ and it has a global minimum at $u^* = 1$ with $\tilde{s}(u^*) = 0$. Further, $s(v,w) = v\tilde{s}(\frac{w}{v})$ for $v \in (0,\infty)$ and $w \in [0,\infty)$.

We prove convexity and lower semi-continuity of s following an idea in [Gun06, Chapter 1]. It suffices to show that s is a supremum of affine functions $h(v,w) := aw + bv$ with $a,b \in \mathbb{R}$. Such affine functions can be written as $h(v,w) = v\tilde{h}(\frac{w}{v})$ for $v \in (0,\infty)$ and $w \in [0,\infty)$ with $\tilde{h}(u) := au + b$ for $u \in [0,\infty)$. Therefore, $\tilde{h} \leq \tilde{s}$ on $[0,\infty)$ implies $h \leq s$ on $[0,\infty)\times[0,\infty)$, that is, the supremum over all h with $\tilde{h} \leq \tilde{s}$ is at least a lower bound for s.

Now let $\tilde{h}_{\bar{u}} \leq \tilde{s}$ be the tangent of $\tilde{s}$ at $\bar{u} \in (0,\infty)$. Thus, $\tilde{h}_{\bar{u}}(u) = \tilde{s}(\bar{u}) + \tilde{s}'(\bar{u})(u-\bar{u}) = (\ln \bar{u})u + 1 - \bar{u}$ and the associated function $h_{\bar{u}}$ reads as

$$h_{\bar{u}}(v,w) = (\ln \bar{u})w + (1-\bar{u})v.$$

Observing

$$s(v,w) := \begin{cases} h_1(0,0), & \text{if } v = 0,\ w = 0, \\ \lim\limits_{u\to\infty} h_u(0,w), & \text{if } v = 0,\ w > 0, \\ \lim\limits_{u\to 0} h_u(v,0), & \text{if } v > 0,\ w = 0, \\ h_{w/v}(v,w), & \text{if } v > 0,\ w > 0 \end{cases}$$

we see that s is indeed the supremum of all affine functions h with $\tilde{h} \leq \tilde{s}$.

The continuity of s outside $(0,0)$ is a direct consequence of the definition of s. The nonnegativity follows from $\tilde{s} \geq 0$. And for $v,w \in (0,\infty)$ the equality $s(v,w) = 0$ is equivalent to $\tilde{s}(\frac{w}{v}) = 0$, that is, $\frac{w}{v}$ has to coincide with the global minimizer $u^* = 1$ of $\tilde{s}$. □

As discussed in Section 7.1 the data available to us is a scaled version $z = a\tilde{z}$ (more precisely $z \approx a\tilde{z}$) of the particle count $\tilde{z}$. Due to the structure of the fitting functional this scaling is not serious. Obviously $S(y,z) = aS\big(\frac{1}{a}y, \tilde{z}\big)$, that is, working with scaled data is the same as working with scaled right-hand sides. Note that when replacing

$z = a\tilde{z}$ by $z \approx a\tilde{z}$ the question of continuity of S arises. This question will be discussed in detail in Section 8.1 and Section 9.1.

Summing up the above, the MAP approach suggests to use the fitting functional

$$S(y,z) = \sum_{i=1}^{m} s(D_i y, z_i) \quad \text{for } y \in Y \text{ and } z \in [0,\infty)^m$$

in the semi-discrete setting and the fitting functional

$$S(y,z) = \int_T s(y(t), z(t)) \,\mathrm{d}\mu(t) \quad \text{for } y \in Y \text{ and } z \in Y$$

for the continuous setting.

Both fitting functionals are closely related to the *Kullback–Leibler divergence* known in statistics as a distance between two probability densities. In the context of Tikhonov-type regularization methods the Kullback-Leibler divergence appears as fitting functional in [Pös08, Section 2.3] and also in [BB09, RA07]. Properties of entropy functionals similar to the Kullback-Leibler divergence are discussed, e.g., in [Egg93, HK05].

Note that the integral of $t \mapsto s(y(t), z(t))$ is well-defined, because this function is measurable:

Proposition 7.2. *Let $f, g : T \to [0,\infty)$ be two functions which are measurable with respect to the σ-algebra $\mathcal{A}_T$ on T and the Borel σ-algebra $\mathcal{B}_{[0,\infty)}$ on $[0,\infty)$. Then $h : T \to [0,\infty]$ defined by $h(t) := s(f(t), g(t))$ is measurable with respect to $\mathcal{A}_T$ and the Borel σ-algebra $\mathcal{B}_{[0,\infty]}$ on $[0,\infty]$.*

Proof. The proof is standard in introductory lectures on measure theory if the function s is continuous. For the sake of completeness we give a version adapted to lower semi-continuous s.

For $b \geq 0$ define the sets

$$G_b := \{(v,w) \in [0,\infty) \times [0,\infty) : s(v,w) > b\} = \big([0,\infty) \times [0,\infty)\big) \setminus s^{-1}([0,b]).$$

These sets are open (with respect to the natural topology on $[0,\infty) \times [0,\infty)$) because $s^{-1}([0,b])$ is closed due to the lower semi-continuity of s (see Lemma 7.1). Thus, for fixed b there are sequences $(\alpha_k)_{k\in\mathbb{N}}$, $(\beta_k)_{k\in\mathbb{N}}$, $(\gamma_k)_{k\in\mathbb{N}}$, and $(\delta_k)_{k\in\mathbb{N}}$ in $[0,\infty)$ such that

$$G_b = \bigcup_{k\in\mathbb{N}} \big([\alpha_k, \beta_k) \times [\gamma_k, \delta_k)\big).$$

From

$$\begin{aligned} h^{-1}\big((b,\infty]\big) &= \big\{t \in T : (f(t), g(t)) \in G_b\big\} \\ &= \bigcup_{k\in\mathbb{N}} \big\{t \in T : (f(t), g(t)) \in [\alpha_k, \beta_k) \times [\gamma_k, \delta_k)\big\} \\ &= \bigcup_{k\in\mathbb{N}} \Big(f^{-1}\big([\alpha_k,\infty)\big) \cap f^{-1}\big([0,\beta_k)\big) \cap g^{-1}\big([\gamma_k,\infty)\big) \cap g^{-1}\big([0,\delta_k)\big)\Big) \end{aligned}$$

we see $h^{-1}\big((b,\infty]\big) \in \mathcal{A}_T$ for all $b \geq 0$, which is equivalent to the measurability of h. □

7.3. Gamma distributed data

Although this part of the thesis is mainly concerned with Poisson distributed data, we want to give another example of non-metric fitting functionals arising in applications.

In SAR imaging (synthetic aperture radar imaging) a phenomenon called speckle noise occurs. Although speckle noise results from interference phenomenons it behaves like real noise and can be described by the Gamma distribution. For details on SAR imaging and its modeling we refer to the detailed exposition in [OQ04]. Assume that a measurement $z_i \in [0, \infty)$ is the average of $L \in \mathbb{N}$ measurements, which is typical for SAR images, and that each single measurement follows an exponential distribution with mean $D_i y > 0$. Then the averaged measurement z_i follows a Gamma distribution with parameters L and $\frac{L}{D_i y}$, that is

$$g(v, w) := \frac{1}{(L-1)!} \left(\frac{L}{v}\right)^L w^{L-1} \exp\left(-\frac{L}{v} w\right)$$

determines the density $p_{\zeta|\eta=y}$ (cf. (7.1)). Note that the assumption $D_i y > 0$ is quite reasonable since in practice there is always some radiation detected by the sensor (in contrast to applications with Poisson distributed data).

The corresponding fitting functionals are

$$S(y, z) = L \sum_{i=1}^{m} \left(\ln \frac{D_i y}{z_i} + \frac{z_i}{D_i y} - 1\right)$$

in the semi-discrete setting and

$$L \int_T \ln \frac{y(t)}{z(t)} + \frac{z(t)}{y(t)} - 1 \, \mathrm{d}\mu(t)$$

in the continuous setting.

8. The semi-discrete setting

In this chapter we analyze the semi-discrete setting for Tikhonov-type regularization with Poisson distributed data as derived in Chapter 7. The solution space (X, τ_X) is an arbitrary topological space, the space (Y, τ_Y) of right-hand sides is $Y := \{y \in L^1(T, \mu) : y \geq 0 \text{ a.e.}\}$ equipped with the topology τ_Y induced by the weak $L^1(T, \mu)$-topology, and the data space (Z, τ_Z) is given by $Z = [0, \infty)^m$. The topology τ_Z will be specified soon. Remember the definition $D_i y := \int_{T_i} y \, \mathrm{d}\mu$ for $y \in Y$, where $T_1, \ldots, T_m \subseteq T$. For obtaining approximate solutions to (1.1) we minimize a Tikhonov-type functional (1.3) with fitting functional

$$S(y, z) := \sum_{i=1}^{m} s(D_i y, z_i) \quad \text{for all } y \in Y \text{ and } z \in Z,$$

where s is defined by (7.2).

One aim of this chapter is to show that the fitting functional S satisfies items (ii), (iii), and (iv) of the basic Assumption 2.1. The second task consists in deriving variational inequalities (4.3) as a prerequisite for proving convergence rates.

8.1. Fundamental properties of the fitting functional

Before we start to verify Assumption 2.1 we have to specify the topology τ_Z on the data space $Z = [0, \infty)^m$. Choosing the topology induced by the usual topology of $\mathbb{R}^m$ is inadvisable because the fitting functional S is not continuous in the second argument with respect to this topology. Indeed, $s(0, 0) = 0$ but $s(0, \varepsilon) = \infty$ for all $\varepsilon > 0$. Thus, we have to look for a stronger topology.

We started the derivation of S in Section 7.2 by considering data $\tilde{z}_i \in \mathbb{N}_0$ for each pixel T_i, and due to transformations and noise we eventually decided to model the data as elements $z_i \in [0, \infty)$ clustering around the points $a\tilde{z}_i$ with an unknown scaling factor $a > 0$. If we assume that the distance between z_i and $a\tilde{z}_i$ is less than $\frac{a}{2}$, that is, noise is not too large, then we can recover $\tilde{z}_i$ from z_i by rounding z_i and dividing by a. Since a is typically unknown (however it can be measured if really necessary, cf. Section 7.1) we can apply this procedure only in the case $\tilde{z}_i = 0$: if z_i is very small, then we might assume $\tilde{z}_i = 0$ and consequently also $z_i = 0$. Thus, convergence $z_i^k \to 0$ with respect to the natural topology $\tau_{[0,\infty)}$ on $[0, \infty)$ of a sequence $(z_i^k)_{k \in \mathbb{N}}$ means that the underlying sequence $(\tilde{z}_i^k)_{k \in \mathbb{N}}$ of particle counts is zero for sufficiently large k, and in turn we may set z_i^k to zero for large k. In other words, from the particle point of view nontrivial convergence to zero is not possible and thus the discontinuity of $s(0, \bullet)$ at zero is irrelevant. Putting these considerations into the language of topologies we

would like to have a topology $\tau^0_{[0,\infty)}$ on $[0,\infty)$ satisfying

$$z_i^k \overset{\tau^0_{[0,\infty)}}{\longrightarrow} 0 \text{ if } k \to \infty \quad \Leftrightarrow \quad z_i^k = 0 \text{ for sufficiently large } k$$

and providing the same convergence behavior at other points as the usual topology $\tau_{[0,\infty)}$ on $[0,\infty)$. This can be achieved by defining $\tau^0_{[0,\infty)}$ to be the topology generated by $\tau_{[0,\infty)}$ and the set $\{0\}$, that is, $\tau^0_{[0,\infty)}$ shall be the weakest topology containing the natural topology and the set $\{0\}$. Eventually, we choose τ_Z to be the product topology of m copies of $\tau^0_{[0,\infty)}$.

Note that the procedure of avoiding convergence to zero can also be applied to other points $a\tilde{z}_i$ if a is exactly known, which in practice is not the case (only a more or less inaccurate measurement of a might be available). Since the treated discontinuity is the only discontinuity of s, the clustering around the points $a\tilde{z}_i$ for $\tilde{z}_i > 0$ makes no problems. To make things precise we prove the following proposition.

Proposition 8.1. *The fitting functional S satisfies item (iv) of Assumption 2.1.*

Proof. Let $y \in Y$ and $z \in Z$ be such that $S(y,z) < \infty$ and take a sequence $(z_k)_{k\in\mathbb{N}}$ in Z converging to z. We have to show $S(y,z_k) \to S(y,z)$, which follows if $s(D_iy,[z_k]_i) \to s(D_iy,z_i)$ for $i = 1,\dots,m$. For i with $D_iy > 0$ this convergence is a consequence of the continuity of $s(D_iy,\bullet)$ with respect to the usual topology $\tau_{[0,\infty)}$ (cf. Lemma 7.1), because $\tau^0_{[0,\infty)}$ is stronger than $\tau_{[0,\infty)}$. If $D_iy = 0$, then $S(y,z) < \infty$ implies $z_i = 0$. Thus, by the definition of $\tau^0_{[0,\infty)}$, the convergence $[z_k]_i \to z_i$ with respect to $\tau^0_{[0,\infty)}$ implies $[z_k]_i = 0$ for sufficiently large k. Consequently $s(D_iy,[z_k]_i) = 0 = s(D_iy,z_i)$ for large k. □

Remark 8.2. Without the non-standard topology $\tau^0_{[0,\infty)}$ we could not prove continuity of the fitting functional S in the second argument. The only way to avoid the discontinuity without modifying the usual topology on $[0,\infty)^m$ is to consider only $y \in Y$ with $D_iy > 0$ for all i or to assume $z_i > 0$ for all i. But since the motivation for using such a fitting functional has been the counting of rare events, excluding zero counts is not desirable. In PET (see Chapter 6) the assumption $D_iy > 0$ would mean that the radioactive fluid disperses through the whole body.

In addition we should be aware of the fact that the discontinuity is not intrinsic to the problem but it is a consequence of an inappropriate data model. In practice convergence to zero is not possible; either there is at least one particle or there is no particle.

The example of Poisson distributed data shows that working with general topological spaces instead of restricting attention only to Banach spaces and their weak or norm topologies provides the necessary freedom for implementing complex data models.

We proceed in verifying Assumption 2.1.

Proposition 8.3. *The fitting functional S satisfies item (ii) of Assumption 2.1.*

Proof. The assertion is a direct consequence of the lower semi-continuity of s (cf. Lemma 7.1) and of the fact that $\tau^0_{[0,\infty)}$ is stronger than $\tau_{[0,\infty)}$. □

Proposition 8.4. *The fitting functional S satisfies item (iii) of Assumption 2.1.*

Proof. Let $y \in Y$, define $z \in Z$ by $z_i := D_i y$ for $i = 1, \dots, m$, and let $(z_k)_{k\in\mathbb{N}}$ be a sequence in Z with $S(y, z_k) \to 0$. We have to show $z_k \to z$, which is equivalent to $[z_k]_i \to z_i$ with respect to $\tau^0_{[0,\infty)}$ for all i. From $S(y, z_k) \to 0$ we see $s(D_i y, [z_k]_i) \to 0$ for all i. If $D_i y = 0$, this immediately gives $[z_k]_i = 0$ for sufficiently large k. If $D_i y > 0$, then the function $s(D_i y, \bullet)$ is strictly convex with a global minimum at $D_i y$ and a minimal value of zero. Thus, $s(D_i y, [z_k]_i) \to 0$ implies $[z_k]_i \to D_i y$ with respect to $\tau_{[0,\infty)}$ and therefore also with respect to $\tau^0_{[0,\infty)}$. □

The propositions of this section show that the basic theorems on existence (Theorem 3.2), stability (Theorem 3.3), and convergence (Theorem 3.4) apply to the semi-discrete setting for regularization with Poisson distributed data.

8.2. Derivation of a variational inequality

In this section we show how to obtain a variational inequality (4.3) from a source condition. Since source conditions, in contrast to variational inequalities, are based on operators mapping between normed vector spaces, we have to enrich our setting somewhat.

Assume that X is a subset of a normed vector space $\tilde{X}$ and that τ_X is the topology induced by the weak topology on $\tilde{X}$. Let $\tilde{A} : \tilde{X} \to L^1(T, \mu)$ be a bounded linear operator and assume that X is contained in the convex and τ_X-closed set $\tilde{A}^{-1}Y = \{x \in \tilde{X} : \tilde{A}x \geq 0 \text{ a.e.}\}$. Then we may define $F : X \to Y$ to be the restriction of $\tilde{A}$ to X. Further let Ω be a stabilizing functional on X which can be extended to a convex functional $\tilde{\Omega} : \tilde{X} \to (-\infty, \infty]$ on $\tilde{X}$. As error measure $E_{x^\dagger}$ we use the associated Bregman distance $B^{\tilde{\Omega}}_{\xi^\dagger}(\bullet, x^\dagger)$, where $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger) \subseteq \tilde{X}^*$.

At first we determine the distance S_Y on Y defined in Definition 2.7 and appearing in the variational inequality (4.3).

Proposition 8.5. *For all $y_1, y_2 \in Y$ the equality*

$$S_Y(y_1, y_2) = \sum_{i=1}^{m} \big(\sqrt{D_i y_1} - \sqrt{D_i y_2}\big)^2$$

is true.

Proof. At first we observe

$$S_Y(y_1, y_2) = \inf_{z\in Z} \big(S(y_1, z) + S(y_2, z)\big) = \sum_{i=1}^{m} \inf_{w\in[0,\infty)} \big(s(D_i y_1, w) + s(D_i y_2, w)\big).$$

If $D_i y_1 = 0$ or $D_i y_2 = 0$, then obviously

$$\inf_{w\in[0,\infty)} \big(s(D_i y_1, w) + s(D_i y_2, w)\big) = 0 = \big(\sqrt{D_i y_1} - \sqrt{D_i y_2}\big)^2.$$

For $D_i y_1 > 0$ and $D_i y_2 > 0$ the infimum over $[0, \infty)$ is the same as over $(0, \infty)$. The function $w \mapsto s(D_i y_1, w) + s(D_i y_2, w)$ is strictly convex and continuously differentiable on $(0, \infty)$. Thus, calculating the zeros of the derivative

$$\frac{\partial}{\partial w}\big(s(D_i y_1, w) + s(D_i y_2, w)\big) = \ln \frac{w}{D_i y_1} + \ln \frac{w}{D_i y_2}$$

shows that the infimum is attained at $w^* := \sqrt{D_i y_1 D_i y_2}$ with infimal value

$$\begin{aligned} s(D_i y_1, w^*) + s(D_i y_2, w^*) &= \sqrt{D_i y_1 D_i y_2} \ln \sqrt{\frac{D_i y_2}{D_i y_1}} + D_i y_1 - \sqrt{D_i y_1 D_i y_2} \\ &\quad + \sqrt{D_i y_1 D_i y_2} \ln \sqrt{\frac{D_i y_1}{D_i y_2}} + D_i y_2 - \sqrt{D_i y_1 D_i y_2} \\ &= \big(\sqrt{D_i y_1} - \sqrt{D_i y_2}\big)^2. \end{aligned}$$

□

As starting point for a variational inequality of the form (4.3) we use the inequality obtained in the following lemma from the source condition $\xi^\dagger \in \mathcal{R}\big((D \circ \tilde{A})^*\big)$, where $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger)$ and $x^\dagger \in X$. The mapping $D \circ \tilde{A}$ is a bounded linear operator mapping between the normed vector spaces $\tilde{X}$ and $\mathbb{R}^m$. Thus, the adjoint $(D \circ \tilde{A})^*$ is well-defined as a bounded linear operator from $\mathbb{R}^m$ into $\tilde{X}^*$. Note that we defined D only on Y, but its extension to $L^1(T, \mu)$ is canonical.

Lemma 8.6. *Let $x^\dagger \in X$ be an Ω-minimizing S-generalized solution to (1.1) such that there is a subgradient $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger) \cap \mathcal{R}\big((D \circ \tilde{A})^*\big)$. Then there is some $c > 0$ such that*

$$B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) \le \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) + c \sum_{i=1}^{m} \big| D_i \tilde{A} x - D_i \tilde{A} x^\dagger \big| \quad \text{for all } x \in \tilde{X}.$$

Proof. Let $\xi^\dagger = (D \circ \tilde{A})^* \eta^\dagger$ with $\eta^\dagger \in \mathbb{R}^m$. Then

$$\begin{aligned} -\langle \xi^\dagger, x - x^\dagger \rangle &= -\big\langle \eta^\dagger, (D \circ \tilde{A})(x - x^\dagger) \big\rangle \le \|\eta^\dagger\|_2 \|(D \circ \tilde{A})(x - x^\dagger)\|_2 \\ &\le \|\eta^\dagger\|_2 \|(D \circ \tilde{A})(x - x^\dagger)\|_1 = \|\eta^\dagger\|_2 \sum_{i=1}^{m} \big| D_i \tilde{A} x - D_i \tilde{A} x^\dagger \big| \end{aligned}$$

for all $x \in X$, where $\langle \bullet, \bullet \rangle$ denotes the duality pairing and $\|\bullet\|_p$ the p-norm on $\mathbb{R}^m$. Thus,

$$B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) = \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) - \langle \xi^\dagger, x - x^\dagger \rangle \le \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) + c \sum_{i=1}^{m} \big| D_i \tilde{A} x - D_i \tilde{A} x^\dagger \big|$$

for all $x \in \tilde{X}$ and any $c \ge \|\eta^\dagger\|_2$. □

The next lemma is an important step in constituting a connection between S_Y from Proposition 8.5 and the inequality obtained in Lemma 8.6.

Lemma 8.7. *For all $a_i, b_i \geq 0$, $i = 1, \dots, m$, the inequality*

$$\left(\sqrt{\sum_{i=1}^m b_i + \sum_{i=1}^m |a_i - b_i|} - \sqrt{\sum_{i=1}^m b_i} \right)^2 \leq \sum_{i=1}^m \bigl(\sqrt{a_i} - \sqrt{b_i}\bigr)^2$$

is true.

Proof. We start with the inequality

$$\bigl|\sqrt{a_i} - \sqrt{b_i}\bigr| \geq \sqrt{b_i + |a_i - b_i|} - \sqrt{b_i}, \tag{8.1}$$

which is obviously true for $a_i \geq b_i \geq 0$ and also for $a_i = 0$, $b_i \geq 0$. We have to verify the inequality only for $0 < a_i < b_i$. In this case the inequality is equivalent to $2\sqrt{b_i} \geq \sqrt{2b_i - a_i} + \sqrt{a_i}$. The function $f(t) := \sqrt{2b_i - t} + \sqrt{t}$ is differentiable on $(0, 2b_i)$ and monotonically increasing on $(0, b_i]$ because

$$f'(t) = \frac{-1}{2\sqrt{2b_i - t}} + \frac{1}{2\sqrt{t}} > \frac{-1}{2\sqrt{2b_i - b_i}} + \frac{1}{2\sqrt{b_i}} = 0 \quad \text{for all } t \in (0, b_i].$$

Thus, $f(t) \leq f(b_i) = 2\sqrt{b_i}$ and therefore $2\sqrt{b_i} \geq \sqrt{2b_i - a_i} + \sqrt{a_i}$ if $0 < a_i < b_i$.

From inequality (8.1) we now obtain

$$\begin{aligned} \sum_{i=1}^m \bigl(\sqrt{a_i} - \sqrt{b_i}\bigr)^2 &\geq \sum_{i=1}^m \bigl(\sqrt{b_i + |a_i - b_i|} - \sqrt{b_i}\bigr)^2 \\ &= \sum_{i=1}^m |a_i - b_i| + 2\sum_{i=1}^m b_i - 2\sum_{i=1}^m \sqrt{b_i + |a_i - b_i|}\sqrt{b_i}. \end{aligned}$$

Interpreting the last of the three sums as an inner product in $\mathbb{R}^m$ we can apply the Cauchy–Schwarz inequality. This yields

$$\begin{aligned} \sum_{i=1}^m \bigl(\sqrt{a_i} - \sqrt{b_i}\bigr)^2 &\geq \sum_{i=1}^m |a_i - b_i| + 2\sum_{i=1}^m b_i - 2\sqrt{\sum_{i=1}^m \bigl(b_i + |a_i - b_i|\bigr)}\sqrt{\sum_{i=1}^m b_i} \\ &= \left(\sqrt{\sum_{i=1}^m b_i + \sum_{i=1}^m |a_i - b_i|} - \sqrt{\sum_{i=1}^m b_i} \right)^2, \end{aligned}$$

which proves the assertion of the lemma. □

Now we are in the position to establish a variational inequality.

Theorem 8.8. *Let $x^\dagger \in X$ be an Ω-minimizing S-generalized solution to (1.1) for which there is a subgradient $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger) \cap \mathcal{R}\bigl((D \circ \tilde{A})^*\bigr)$ and assume that there are constants $\tilde{\beta} \in (0, 1]$ and $\tilde{c} > 0$ such that*

$$\tilde{\beta} B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) - \bigl(\tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger)\bigr) \leq \tilde{c} \quad \text{for all } x \in \tilde{X}. \tag{8.2}$$

Then

$$\tilde{\beta} B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \bar{c}\sqrt{S_Y\bigl(F(x), F(x^\dagger)\bigr)} \quad \text{for all } x \in X \tag{8.3}$$

with $\bar{c} > 0$, that is, $x^\dagger$ satisfies Assumption 4.7 with $\varphi(t) = \bar{c}\sqrt{t}$, $\beta = \tilde{\beta}$, and $M = X$.

Proof. By $\|\bullet\|_1$ we denote the 1-norm on $\mathbb{R}^m$. From Lemma 8.6 we know

$$B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) \leq \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) + c\|D\tilde{A}x - D\tilde{A}x^\dagger\|_1 \quad \text{for all } x \in \tilde{X}$$

with $c > 0$ and Proposition 12.14 with φ replaced by the concave and monotonically increasing function $\tilde{\varphi}(t) := \sqrt{\|D\tilde{A}x^\dagger\|_1 + t} - \sqrt{\|D\tilde{A}x^\dagger\|_1}$, $t \in [0, \infty)$, yields

$$\tilde{\beta} B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) \leq \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) + \bar{c}\tilde{\varphi}\big(\|D\tilde{A}x - D\tilde{A}x^\dagger\|_1\big) \quad \text{for all } x \in \tilde{X}$$

with $\bar{c} > 0$. For $x \in X$ we have $\tilde{\varphi}\big(\|D\tilde{A}x - D\tilde{A}x^\dagger\|_1\big) = \tilde{\varphi}\big(\|DF(x) - DF(x^\dagger)\|_1\big)$ and Lemma 8.7, in combination with Proposition 8.5, provides $\tilde{\varphi}\big(\|DF(x) - DF(x^\dagger)\|_1\big) \leq \sqrt{S_Y(F(x), F(x^\dagger))}$. Taking into account $\tilde{\Omega} = \Omega$ on X we arrive at

$$\tilde{\beta} B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \bar{c}\sqrt{S_Y\big(F(x), F(x^\dagger)\big)} \quad \text{for all } x \in X.$$

□

The assumption (8.2) is discussed in Remark 12.15.

Having the variational inequality (8.3) at hand we can apply the convergence rates theorems of Part I (see Theorem 4.11, Theorem 4.24, Theorem 5.4) to the semi-discrete setting for regularization with Poisson distributed data.

Note that $\mathcal{R}\big((D \circ \tilde{A})^*\big)$ is a finite-dimensional subspace of X^* and thus the source condition $\xi^\dagger \in \mathcal{R}\big((D \circ \tilde{A})^*\big)$ is a very strong assumption. Such a strong assumption for obtaining convergence in terms of Bregman distances is quite natural because we only have finite-dimensional data and thus cannot expect to obtain convergence rates for a wide class of elements $x^\dagger$. To avoid difficulties arising from small data spaces we introduced the notion of S-generalized solutions (see Section 3.1) and studied convergence to such ones. But the Bregman distance does not follow this concept.

Eventually, we emphasize that the constant c in (8.3) does not depend on the dimension m of the data space. Thus, also the error estimates obtained in Theorem 4.11 and Theorem 4.24 from such a variational inequality do not depend on m.

9. The continuous setting

Next to the semi-discrete setting analyzed in the previous chapter we also suggested a continuous model for regularization with Poisson distributed data in Chapter 7. The solution space (X, τ_X) is an arbitrary topological space and the space (Y, τ_Y) of right-hand sides is $Y := \{y \in L^1(T, \mu) : y \geq 0 \text{ a.e.}\}$ equipped with the topology τ_Y induced by the weak $L^1(T, \mu)$-topology. Data comes from the same set as the right-hand sides, that is, $Z := Y$. But the topology τ_Z on the data space Z shall be induced by the fitting functional S (cf. Proposition 2.10). In the continuous setting this fitting functional is given by

$$S(y_1, y_2) = \int_T s(y_1(t), y_2(t)) \, \mathrm{d}\mu(t) \quad \text{for } y_1, y_2 \in Y$$

with s from (7.2). It was shown in Proposition 7.2 that the integral is well-defined. Note that we use the letter y also for the second argument of S because $Y = Z$ and because we did so in Proposition 2.10.

In the present chapter we show that the fitting functional S (almost) satisfies the assumptions of Proposition 2.10 and thus also the basic Assumption 2.1. A second step consists in deriving variational inequalities (4.3) as a prerequisite for proving convergence rates.

9.1. Fundamental properties of the fitting functional

In analyzing the fitting functional S we have to be very careful because the definition of the integrand s depends on the zeros of the arguments y_1, y_2. To simplify expositions, for $y \in Y$ we denote by $N_y \subseteq T$ a measurable set such that $y = 0$ almost everywhere on N_y and $y > 0$ almost everywhere on $T \setminus N_y$. Typically there are many sets N_y with this property, but the difference between two such sets is always a set of measure zero. Thus, the integral of a function over a set N_y does not depend on the concrete choice of N_y.

With this notation at hand the fitting functional can be written as follows.

Lemma 9.1. *For all $y_1, y_2 \in Y$ the equality*

$$S(y_1, y_2) = \begin{cases} \displaystyle\int_{N_{y_2} \setminus N_{y_1}} y_1 \, \mathrm{d}\mu + \int_{T \setminus N_{y_2}} y_2 \ln \frac{y_2}{y_1} + y_1 - y_2 \, \mathrm{d}\mu, & \text{if } \mu(N_{y_1} \setminus N_{y_2}) = 0, \\ \infty, & \text{if } \mu(N_{y_1} \setminus N_{y_2}) > 0 \end{cases}$$

is satisfied.

Proof. We write T as the union of mutually disjoint sets:

$$T = (N_{y_1} \cap N_{y_2}) \cup (N_{y_1} \setminus N_{y_2}) \cup (N_{y_2} \setminus N_{y_1}) \cup \big(T \setminus (N_{y_1} \cup N_{y_2})\big).$$

If $\mu(N_{y_1} \setminus N_{y_2}) > 0$, then $\int_{N_{y_1} \setminus N_{y_2}} s(y_1, y_2)\,\mathrm{d}\mu = \infty$ and thus $S(y_1, y_2) = \infty$.

Now assume $\mu(N_{y_1} \setminus N_{y_2}) = 0$. Then

$$\int\limits_{N_{y_1} \cap N_{y_2}} s(y_1, y_2)\,\mathrm{d}\mu = 0 \quad \text{and} \quad \int\limits_{N_{y_1} \setminus N_{y_2}} s(y_1, y_2)\,\mathrm{d}\mu = 0.$$

Further

$$\int\limits_{N_{y_2} \setminus N_{y_1}} s(y_1, y_2)\,\mathrm{d}\mu = \int\limits_{N_{y_2} \setminus N_{y_1}} y_1\,\mathrm{d}\mu$$

and

$$\int\limits_{T \setminus (N_{y_1} \cup N_{y_2})} s(y_1, y_2)\,\mathrm{d}\mu = \int\limits_{T \setminus (N_{y_1} \cup N_{y_2})} y_2 \ln \frac{y_2}{y_1} + y_1 - y_2\,\mathrm{d}\mu.$$

Writing the set under the last integral sign as

$$T \setminus (N_{y_1} \cup N_{y_2}) = (T \setminus N_{y_2}) \setminus (N_{y_1} \setminus N_{y_2})$$

we see that replacing it by $T \setminus N_{y_2}$ does not change the integral (because $\mu(N_{y_1} \setminus N_{y_2}) = 0$). Summing up the four integrals the assertion of the lemma follows. □

We now start to verify the assumptions of Proposition 2.10.

Proposition 9.2. *The fitting functional S satisfies item (i) in Proposition 2.10.*

Proof. Let $y_1, y_2 \in Y$. If $y_1 = y_2$ then $\mu(N_{y_1} \setminus N_{y_2}) = 0$ and $\mu(N_{y_2} \setminus N_{y_1}) = 0$. Thus, Lemma 9.1 provides

$$S(y_1, y_2) = \int\limits_{T \setminus N_{y_2}} y_2 \ln \frac{y_2}{y_1} + y_1 - y_2\,\mathrm{d}\mu.$$

The integrand is zero a.e. on $T \setminus N_{y_2}$ and therefore $S(y_1, y_2) = 0$.

For the reverse direction assume $S(y_1, y_2) = 0$. Lemma 9.1 gives $\mu(N_{y_1} \setminus N_{y_2}) = 0$ as well as

$$\int\limits_{N_{y_2} \setminus N_{y_1}} y_1\,\mathrm{d}\mu + \int\limits_{T \setminus N_{y_2}} y_2 \ln \frac{y_2}{y_1} + y_1 - y_2\,\mathrm{d}\mu = 0.$$

Since both integrands are nonnegative both integrals vanish. From the first summand we thus see $y_1 = 0$ a.e. on $N_{y_2} \setminus N_{y_1}$, that is, $y_1 = 0$ a.e. on N_{y_2}. From the second summand we obtain $y_1 = y_2$ a.e. on $T \setminus N_{y_2}$. This proves the assertion. □

Next we prove the lower semi-continuity of S.

Proposition 9.3. *The fitting functional S satisfies item (ii) in Proposition 2.10.*

Proof. The proof is an adaption of the corresponding proofs in [Pös08, Theorem 2.19] and [Gun06, Lemma 1.1.3]. We have to show that the sublevel sets $M_S(c) := \{(y_1, y_2) \in Y \times Y : S(y_1, y_2) \leq c\} \subseteq (L^1(T,\mu))^2$ are closed with respect to the weak $(L^1(T,\mu))^2$-topology for all $c \geq 0$ (for $c < 0$ these sets are empty).

Since the $M_S(c)$ are convex (S is convex, cf. Lemma 7.1) and convex sets in Banach spaces are weakly closed if and only if they are closed, it suffices to show the closedness of $M_S(c)$ with respect to the $(L^1(T,\mu))^2$-norm. So let $(y_1^k, y_2^k)_{k\in\mathbb{N}}$ be a sequence in $M_S(c)$ converging to some element $(y_1, y_2) \in (L^1(T,\mu))^2$ with respect to the $(L^1(T,\mu))^2$-norm. Then there exists a subsequence $(y_1^{k_n}, y_2^{k_n})_{n\in\mathbb{N}}$ such that $(y_1^{k_n})_{n\in\mathbb{N}}$ and $(y_2^{k_n})_{n\in\mathbb{N}}$ converge almost everywhere pointwise to y_1 and y_2, respectively (cf. [Kan03, Lemma 1.30]). By the lower semi-continuity of s and Fatou's lemma we now get

$$\begin{aligned} S(y_1, y_2) &= \int_T s(y_1, y_2)\,\mathrm{d}\mu \leq \int_T \liminf_{n\to\infty} s(y_1^{k_n}, y_2^{k_n})\,\mathrm{d}\mu \\ &\leq \liminf_{n\to\infty} \int_T s(y_1^{k_n}, y_2^{k_n})\,\mathrm{d}\mu = \liminf_{n\to\infty} S(y_1^{k_n}, y_2^{k_n}) \leq c, \end{aligned}$$

that is, $(y_1, y_2) \in M_S(c)$. □

Before we go on in verifying the assumptions of Proposition 2.10 we prove the following lemma.

Lemma 9.4. *Let $c > 0$. Then for all $y_1, y_2 \in Y$ with $y_1 \leq c$ a.e. and $y_2 \leq c$ a.e. the inequality*

$$\|y_1 - y_2\|_{L^1(T,\mu)}^2 \leq (1 + 2\tilde{c})c\mu(T)S(y_1, y_2)$$

is true, where

$$\tilde{c} := \sup_{u\in(0,\infty)\setminus\{1\}} \frac{(1-u)^2}{(1+u)(u\ln u + 1 - u)} < \infty.$$

Proof. If $S(y_1, y_2) = \infty$, then the assertion is trivially true. If $S(y_1, y_2) < \infty$, then $\mu(N_{y_1} \setminus N_{y_2}) = 0$ and

$$S(y_1, y_2) = \int_{N_{y_2}\setminus N_{y_1}} y_1\,\mathrm{d}\mu + \int_{T\setminus N_{y_2}} y_2 \ln\frac{y_2}{y_1} + y_1 - y_2\,\mathrm{d}\mu$$

by Lemma 9.1. We write the $L^1(T,\mu)$-norm as

$$\|y_1 - y_2\|_{L^1(T,\mu)} = \int_T |y_1 - y_2|\,\mathrm{d}\mu = \int_{N_{y_2}\setminus N_{y_1}} y_1\,\mathrm{d}\mu + \int_{T\setminus N_{y_2}} |y_1 - y_2|\,\mathrm{d}\mu.$$

The first summand can be estimated by

$$\int_{N_{y_2}\setminus N_{y_1}} y_1\,\mathrm{d}\mu = \sqrt{\int_{N_{y_2}\setminus N_{y_1}} y_1\,\mathrm{d}\mu}\sqrt{\int_{N_{y_2}\setminus N_{y_1}} y_1\,\mathrm{d}\mu} \leq \sqrt{c\mu(N_{y_2}\setminus N_{y_1})}\sqrt{\int_{N_{y_2}\setminus N_{y_1}} y_1\,\mathrm{d}\mu}.$$

To bound the second summand we introduce a measurable set $\tilde{T} \subseteq T$ such that $y_1 \neq y_2$ a.e. on $\tilde{T}$ and $y_1 = y_2$ a.e. on $T \setminus \tilde{T}$. Then, following an idea in [Gil10, proof of Theorem 3],

$$\begin{aligned}\int_{T\setminus N_{y_2}} |y_1 - y_2| \,\mathrm{d}\mu &= \int_{\tilde{T}\setminus N_{y_2}} \frac{|y_1 - y_2|}{\sqrt{y_2 \ln \frac{y_2}{y_1} + y_1 - y_2}} \sqrt{y_2 \ln \frac{y_2}{y_1} + y_1 - y_2} \,\mathrm{d}\mu \\ &\leq \sqrt{\int_{\tilde{T}\setminus N_{y_2}} \frac{(y_1 - y_2)^2}{y_2 \ln \frac{y_2}{y_1} + y_1 - y_2} \,\mathrm{d}\mu} \sqrt{\int_{\tilde{T}\setminus N_{y_2}} y_2 \ln \frac{y_2}{y_1} + y_1 - y_2 \,\mathrm{d}\mu}.\end{aligned}$$

Here we applied the Cauchy–Schwarz inequality. This is allowed because $S(y_1, y_2) < \infty$ implies that the second factor is finite and

$$\begin{aligned}\int_{\tilde{T}\setminus N_{y_2}} \frac{(y_1 - y_2)^2}{y_2 \ln \frac{y_2}{y_1} + y_1 - y_2} \,\mathrm{d}\mu &= \int_{\tilde{T}\setminus N_{y_2}} \frac{\left(1 - \frac{y_2}{y_1}\right)^2}{\left(1 + \frac{y_2}{y_1}\right)\left(\frac{y_2}{y_1} \ln \frac{y_2}{y_1} + 1 - \frac{y_2}{y_1}\right)} (y_1 + y_2) \,\mathrm{d}\mu \\ &\leq 2c\mu(\tilde{T} \setminus N_{y_2}) \sup_{u\in(0,\infty)\setminus\{1\}} \frac{(1-u)^2}{(1+u)(u \ln u + 1 - u)}\end{aligned}$$

shows that also the second factor is finite if the supremum is. We postpone the discussion of the supremum to the end of this proof. Putting all estimates together we obtain

$$\|y_1 - y_2\|_{L^1(T,\mu)} \leq \sqrt{c\mu(T)} \sqrt{\int_{N_{y_2}\setminus N_{y_1}} y_1 \,\mathrm{d}\mu} + \sqrt{2c\tilde{c}\mu(T)} \sqrt{\int_{T\setminus N_{y_2}} y_2 \ln \frac{y_2}{y_1} + y_1 - y_2 \,\mathrm{d}\mu}$$

and applying the Cauchy–Schwarz inequality of $\mathbb{R}^2$ gives

$$\|y_1 - y_2\|^2_{L^1(T,\mu)} \leq (1 + 2\tilde{c})c\mu(T)S(y_1, y_2).$$

It remains to show that $h(u) := \frac{(1-u)^2}{(1+u)(u \ln u + 1 - u)}$ is bounded on $(0, \infty)\setminus\{1\}$. Obviously $h(u) \geq 0$ for $u \in (0, \infty) \setminus \{1\}$, $\lim_{u\to 0} h(u) = 1$, and $\lim_{u\to\infty} h(u) = 0$. Applying l'Hôpital's rule we see $\lim_{u\to 1\pm 0} h(u) = 1$. Since h is continuous on $(0, \infty) \setminus \{1\}$ these observations show that h is bounded. □

Remark 9.5. From numerical maximization one sees that $\tilde{c} \approx 1.12$ in Lemma 9.4.

Instead of item (iii) in Proposition 2.10 we prove a weaker assertion. We require the additional assumption that all involved elements from Y have a common upper bound $c > 0$. This restriction is not very serious because in practical imaging problems the range of the images is usually a finite interval.

Proposition 9.6. *Let $y \in Y$ and let $(y_k)_{k\in\mathbb{N}}$ be a sequence in Y such that $S(y, y_k) \to 0$. If there is some $c > 0$ such that $y \leq c$ a.e. and $y_k \leq c$ a.e. for sufficiently large k, then $y_k \to y$.*

Proof. The assertion is a direct consequence of Lemma 9.4 because τ_Y is weaker than the norm topology on $L^1(T,\mu)$. □

Finally we show a weakened version of item (iv) in Proposition 2.10. The influence of the modification is discussed in Remark 9.8 below.

Proposition 9.7. *Let $y, \tilde{y} \in Y$ such that $S(y,\tilde{y}) < \infty$ and*

$$\operatorname*{ess\,sup}_{T\setminus(N_{\tilde{y}}\cup N_y)} \left|\ln\frac{\tilde{y}}{y}\right| < \infty,$$

and let $(y_k)_{k\in\mathbb{N}}$ be a sequence in Y with $S(\tilde{y}, y_k) \to 0$. If there is some $c > 0$ such that $y \le c$ a.e. and $y_k \le c$ a.e. for sufficiently large k, then $S(y, y_k) \to S(y,\tilde{y})$.

Proof. From $S(y,\tilde{y}) < \infty$ we know $\mu(N_y \setminus N_{\tilde{y}}) = 0$ and $S(\tilde{y}, y_k) \to 0$ gives $\mu(N_{\tilde{y}} \setminus N_{y_k}) = 0$ for sufficiently large k (cf. Lemma 9.1). Thus

$$0 \le \mu(N_y \setminus N_{y_k}) \le \mu\big((N_y \setminus N_{\tilde{y}}) \cup (N_{\tilde{y}} \setminus N_{y_k})\big) \le \mu(N_y \setminus N_{\tilde{y}}) + \mu(N_{\tilde{y}} \setminus N_{y_k}) = 0,$$

that is,

$$S(y, y_k) = \int\limits_{N_{y_k}\setminus N_y} y \,\mathrm{d}\mu + \int\limits_{T\setminus N_{y_k}} y_k \ln\frac{y_k}{y} + y - y_k \,\mathrm{d}\mu$$

for large k.

The assertion is proven if we can show

$$|S(y, y_k) - S(y,\tilde{y})| \le c_1 S(\tilde{y}, y_k) + c_2 \|y_k - \tilde{y}\|_{L^1(T,\mu)}$$

for some $c_1, c_2 \ge 0$ (cf. Lemma 9.4). We start with

$$\begin{aligned}\int\limits_{N_{y_k}\setminus N_y} y \,\mathrm{d}\mu - \int\limits_{N_{\tilde{y}}\setminus N_y} y \,\mathrm{d}\mu &= \int\limits_{N_{y_k}} y \,\mathrm{d}\mu - \int\limits_{N_{\tilde{y}}} y \,\mathrm{d}\mu \\ &= \int\limits_{N_{y_k}\setminus N_{\tilde{y}}} y \,\mathrm{d}\mu + \int\limits_{N_{y_k}\cap N_{\tilde{y}}} y \,\mathrm{d}\mu - \int\limits_{N_{\tilde{y}}} y \,\mathrm{d}\mu = \int\limits_{N_{y_k}\setminus N_{\tilde{y}}} y \,\mathrm{d}\mu.\end{aligned}$$

The second equality follows from $\mu\big(N_{\tilde{y}} \setminus (N_{y_k} \cap N_{\tilde{y}})\big) = \mu(N_{\tilde{y}} \setminus N_{y_k}) = 0$. As the second step we write the difference

$$\int\limits_{T\setminus N_{y_k}} y_k \ln\frac{y_k}{y} + y - y_k \,\mathrm{d}\mu - \int\limits_{T\setminus N_{\tilde{y}}} \tilde{y} \ln\frac{\tilde{y}}{y} + y - \tilde{y} \,\mathrm{d}\mu \tag{9.1}$$

as a sum of three integrals over the mutually disjoint sets $(T \setminus N_{y_k}) \setminus (T \setminus N_{\tilde{y}})$, $(T \setminus N_{y_k}) \cap (T \setminus N_{\tilde{y}})$, and $(T \setminus N_{\tilde{y}}) \setminus (T \setminus N_{y_k})$. The integral over the first set is zero because $\mu\big((T \setminus N_{y_k}) \setminus (T \setminus N_{\tilde{y}})\big) = \mu(N_{\tilde{y}} \setminus N_{y_k}) = 0$. The second set can be replaced by $T \setminus N_{y_k}$ because $\mu\big((T \setminus N_{y_k}) \cap (T \setminus N_{\tilde{y}})\big) = \mu\big((T \setminus N_{y_k}) \setminus (N_{\tilde{y}} \setminus N_{y_k})\big) = \mu(T \setminus N_{y_k})$. The third set equals $N_{y_k} \setminus N_{\tilde{y}}$. Thus, the difference (9.1) is

$$\int\limits_{T\setminus N_{y_k}} y_k \ln\frac{y_k}{y} + y - y_k - \tilde{y} \ln\frac{\tilde{y}}{y} - y + \tilde{y} \,\mathrm{d}\mu - \int\limits_{N_{y_k}\setminus N_{\tilde{y}}} \tilde{y} \ln\frac{\tilde{y}}{y} + y - \tilde{y} \,\mathrm{d}\mu.$$

Combining both steps we obtain

$$S(y,y_k) - S(y,\tilde{y}) = \int\limits_{T\setminus N_{y_k}} y_k \ln\frac{y_k}{y} - y_k - \tilde{y}\ln\frac{\tilde{y}}{y} + \tilde{y}\,\mathrm{d}\mu + \int\limits_{N_{y_k}\setminus N_{\tilde{y}}} \tilde{y} - \tilde{y}\ln\frac{\tilde{y}}{y}\,\mathrm{d}\mu \quad (9.2)$$

The second integral can be bounded by

$$\int\limits_{N_{y_k}\setminus N_{\tilde{y}}} \tilde{y} - \tilde{y}\ln\frac{\tilde{y}}{y}\,\mathrm{d}\mu \le \left(\operatorname*{ess\,sup}_{N_{y_k}\setminus N_{\tilde{y}}} \left|1 - \ln\frac{\tilde{y}}{y}\right|\right) \int\limits_{N_{y_k}\setminus N_{\tilde{y}}} \tilde{y}\,\mathrm{d}\mu.$$

For the first integral we obtain the bound

$$\begin{aligned}
&\int\limits_{T\setminus N_{y_k}} y_k\ln\frac{y_k}{y} - y_k - \tilde{y}\ln\frac{\tilde{y}}{y} + \tilde{y}\,\mathrm{d}\mu \\
&\qquad = \int\limits_{T\setminus N_{y_k}} y_k\ln\frac{y_k}{\tilde{y}} + \tilde{y} - y_k + (y_k - \tilde{y})\ln\frac{\tilde{y}}{y}\,\mathrm{d}\mu \\
&\qquad \le \int\limits_{T\setminus N_{y_k}} y_k\ln\frac{y_k}{\tilde{y}} + \tilde{y} - y_k\,\mathrm{d}\mu + \operatorname*{ess\,sup}_{T\setminus N_{y_k}} \left|\ln\frac{\tilde{y}}{y}\right| \|y_k - \tilde{y}\|_{L^1(T,\mu)}.
\end{aligned}$$

Thus,

$$S(y,y_k) - S(y,\tilde{y}) \le c_1 S(\tilde{y},y_k) + c_2\|y_k - \tilde{y}\|_{L^1(T,\mu)}$$

with

$$c_1 := \max\left\{1, \operatorname*{ess\,sup}_{T\setminus(N_{\tilde{y}}\cup N_y)} \left|1 - \ln\frac{\tilde{y}}{y}\right|\right\} \quad \text{and} \quad c_2 := \operatorname*{ess\,sup}_{T\setminus(N_{\tilde{y}}\cup N_y)} \left|\ln\frac{\tilde{y}}{y}\right|.$$

Using the same arguments one shows

$$-\big(S(y,y_k) - S(y,\tilde{y})\big) \le c_1 S(\tilde{y},y_k) + c_2\|y_k - \tilde{y}\|_{L^1(T,\mu)}.$$

Therefore, the assertion of the proposition is true. □

Remark 9.8. The additional assumption in Proposition 9.7 that there exists a common bound $c > 0$ is not very restrictive as noted before Proposition 9.6. Also requiring that $\ln\frac{\tilde{y}}{y}$ is essentially bounded is not too strong: Careful inspection of the proofs in Chapter 3 shows that this assumption is only of importance in the proof of Theorem 3.3 (stability). There $\tilde{y} := F(x_\alpha^y)$. Thus, the additional assumption reduces to a similarity condition between a data element y and the images of corresponding minimizers x_α^y of the Tikhonov-type functional T_α^y.

9.2. Derivation of a variational inequality

The aim of this section is to obtain a variational inequality (4.3) from a source condition. As in Section 8.2 we have to enrich the setting slightly to allow the formulation of source

conditions. This enrichment is exactly the same as for the semi-discrete setting. But to improve readability we repeat it here.

Assume that X is a subset of a normed vector space $\tilde{X}$ and that τ_X is the topology induced by the weak topology on $\tilde{X}$. Let $\tilde{A} : \tilde{X} \to L^1(T, \mu)$ be a bounded linear operator and assume that X is contained in the convex and τ_X-closed set $\tilde{A}^{-1}Y = \{x \in \tilde{X} : \tilde{A}x \geq 0 \text{ a.e.}\}$. Then we may define $F : X \to Y$ to be the restriction of $\tilde{A}$ to X. Further let Ω be a stabilizing functional on X which can be extended to a convex functional $\tilde{\Omega} : \tilde{X} \to (-\infty, \infty]$ on $\tilde{X}$. As error measure $E_{x^\dagger}$ we use the associated Bregman distance $B^{\tilde{\Omega}}_{\xi^\dagger}(\bullet, x^\dagger)$, where $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger) \subseteq \tilde{X}^*$.

At first we determine the distance S_Y on Y defined in Definition 2.7 and appearing in the variational inequality (4.3).

Proposition 9.9. *For all $y_1, y_2 \in Y$ the equality*

$$S_Y(y_1, y_2) = \int_T \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 \mathrm{d}\mu$$

is true.

Proof. At first we show

$$S_Y(y_1, y_2) = \inf_{y \in Y} \big(S(y_1, y) + S(y_2, y)\big) \leq \int_T \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 \mathrm{d}\mu.$$

Take $y = \sqrt{y_1 y_2}$. Then $\mu(N_{y_1} \setminus N_{\sqrt{y_1 y_2}}) = 0$ and $\mu(N_{y_2} \setminus N_{\sqrt{y_1 y_2}}) = 0$. Thus,

$$\begin{aligned}
&S(y_1, \sqrt{y_1 y_2}) + S(y_2, \sqrt{y_1 y_2}) \\
&\quad = \int_{N_{\sqrt{y_1 y_2}} \setminus N_{y_1}} y_1 \,\mathrm{d}\mu + \int_{N_{\sqrt{y_1 y_2}} \setminus N_{y_2}} y_2 \,\mathrm{d}\mu \\
&\qquad + \int_{T \setminus N_{\sqrt{y_1 y_2}}} \sqrt{y_1 y_2} \ln \frac{\sqrt{y_2}}{\sqrt{y_1}} + y_1 - \sqrt{y_1 y_2} + \sqrt{y_1 y_2} \ln \frac{\sqrt{y_1}}{\sqrt{y_2}} + y_2 - \sqrt{y_1 y_2} \,\mathrm{d}\mu \\
&\quad = \int_T \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 \mathrm{d}\mu,
\end{aligned}$$

where we used the fact that the difference between $N_{\sqrt{y_1 y_2}}$ and $N_{y_1} \cup N_{y_2}$ is a set of measure zero.

It remains to show

$$S(y_1, \sqrt{y_1 y_2}) + S(y_2, \sqrt{y_1 y_2}) \leq S(y_1, y) + S(y_2, y) \quad \text{for all } y \in Y.$$

If the right-hand side of this inequality is finite (only this case is of interest), then

$$\begin{aligned}
&S(y_1, \sqrt{y_1 y_2}) + S(y_2, \sqrt{y_1 y_2}) - S(y_1, y) - S(y_2, y) \\
&\quad = \int_{N_y} \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 - y_1 - y_2 \,\mathrm{d}\mu \\
&\qquad + \int_{T \setminus N_y} \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 - y \ln \frac{y}{y_1} - y_1 + y - y \ln \frac{y}{y_2} - y_2 + y \,\mathrm{d}\mu
\end{aligned}$$

and simple rearrangements yield

$$\begin{aligned} &S(y_1, \sqrt{y_1 y_2}) + S(y_2, \sqrt{y_1 y_2}) - S(y_1, y) - S(y_2, y) \\ &\quad = \int_{N_y} -2\sqrt{y_1 y_2}\,\mathrm{d}\mu + \int_{T \setminus N_y} -2\sqrt{y_1 y_2}\left(1 - \frac{y}{\sqrt{y_1 y_2}} + \frac{y}{\sqrt{y_1 y_2}} \ln \frac{y}{\sqrt{y_1 y_2}}\right) \mathrm{d}\mu \leq 0. \end{aligned}$$

Thus, the proof is complete. □

The starting point for obtaining a variational inequality is the following lemma.

Lemma 9.10. *Let $x^\dagger \in X$ be an Ω-minimizing S-generalized solution to (1.1) such that there is a subgradient $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger) \cap \mathcal{R}(\tilde{A}^*)$. Then there is some $c > 0$ such that*

$$B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) \leq \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) + c\|\tilde{A}x - \tilde{A}x^\dagger\|_{L^1(T,\mu)} \quad \text{for all } x \in \tilde{X}.$$

Proof. Let $\xi^\dagger = \tilde{A}^*\eta^\dagger$ with $\eta^\dagger \in L^1(T,\mu)^*$. Then

$$-\langle \xi^\dagger, x - x^\dagger\rangle = -\langle \eta^\dagger, \tilde{A}(x - x^\dagger)\rangle \leq \|\eta^\dagger\|_{L^1(T,\mu)^*}\|\tilde{A}(x - x^\dagger)\|_{L^1(T,\mu)}$$

for all $x \in \tilde{X}$, where $\langle \bullet, \bullet \rangle$ denotes the duality pairing. Thus,

$$B_{\xi^\dagger}^{\tilde{\Omega}}(x, x^\dagger) = \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) - \langle \xi^\dagger, x - x^\dagger\rangle \leq \tilde{\Omega}(x) - \tilde{\Omega}(x^\dagger) + c\|\tilde{A}x - \tilde{A}x^\dagger\|_{L^1(T,\mu)}$$

for all $x \in \tilde{X}$ and any $c \geq \|\eta^\dagger\|_{L^1(T,\mu)^*}$. □

The next lemma is an important step in constituting a connection between S_Y from Proposition 9.9 and the inequality obtained in Lemma 9.10.

Lemma 9.11. *For all $y_1, y_2 \in Y$ the inequality*

$$\left(\sqrt{\|y_2\|_{L^1(T,\mu)} + \|y_1 - y_2\|_{L^1(T,\mu)}} - \sqrt{\|y_2\|_{L^1(T,\mu)}}\right)^2 \leq \int_T \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 \mathrm{d}\mu$$

is true.

Proof. In the proof of Lemma 8.7 we verified the inequality

$$\left|\sqrt{a} - \sqrt{b}\right| \geq \sqrt{b + |a - b|} - \sqrt{b} \quad \text{for all } a, b \geq 0.$$

From this inequality we obtain

$$\begin{aligned} \int_T \left(\sqrt{y_1} - \sqrt{y_2}\right)^2 \mathrm{d}\mu &\geq \int_T \left(\sqrt{y_2 + |y_1 - y_2|} - \sqrt{y_2}\right)^2 \mathrm{d}\mu \\ &= \int_T |y_1 - y_2|\,\mathrm{d}\mu + 2\int_T y_2\,\mathrm{d}\mu - 2\int_T \sqrt{y_2 + |y_1 - y_2|}\sqrt{y_2}\,\mathrm{d}\mu. \end{aligned}$$

Interpreting the last of the three integrals as an inner product in $L^2(T,\mu)$ we can apply the Cauchy–Schwarz inequality. This yields

$$\begin{aligned}\int_T \bigl(\sqrt{y_1}-\sqrt{y_2}\bigr)^2\,\mathrm{d}\mu &\geq \int_T |y_1-y_2|\,\mathrm{d}\mu + 2\int_T y_2\,\mathrm{d}\mu - 2\sqrt{\int_T y_2+|y_1-y_2|\,\mathrm{d}\mu}\sqrt{\int_T y_2\,\mathrm{d}\mu}\\ &= \left(\sqrt{\int_T y_2\,\mathrm{d}\mu+\int_T |y_1-y_2|\,\mathrm{d}\mu}-\sqrt{\int_T y_2\,\mathrm{d}\mu}\right)^2,\end{aligned}$$

which proves the assertion of the lemma. □

Now we are in the position to establish a variational inequality.

Theorem 9.12. *Let $x^\dagger \in X$ be an Ω-minimizing S-generalized solution to (1.1) for which there is a subgradient $\xi^\dagger \in \partial\tilde{\Omega}(x^\dagger)\cap\mathcal{R}(\tilde{A}^*)$ and assume that there are constants $\tilde{\beta}\in(0,1]$ and $\tilde{c}>0$ such that*

$$\tilde{\beta}B_{\xi^\dagger}^{\tilde{\Omega}}(x,x^\dagger)-\bigl(\tilde{\Omega}(x)-\tilde{\Omega}(x^\dagger)\bigr)\leq\tilde{c}\quad \textit{for all } x\in\tilde{X}. \tag{9.3}$$

Then

$$\tilde{\beta}B_{\xi^\dagger}^{\tilde{\Omega}}(x,x^\dagger)\leq\Omega(x)-\Omega(x^\dagger)+\bar{c}\sqrt{S_Y\bigl(F(x),F(x^\dagger)\bigr)}\quad \textit{for all } x\in X \tag{9.4}$$

with $\bar{c}>0$, that is, $x^\dagger$ satisfies Assumption 4.7 with $\varphi(t)=\bar{c}\sqrt{t}$, $\beta=\tilde{\beta}$, and $M=X$.

Proof. From Lemma 9.10 we know

$$B_{\xi^\dagger}^{\tilde{\Omega}}(x,x^\dagger)\leq\tilde{\Omega}(x)-\tilde{\Omega}(x^\dagger)+c\|\tilde{A}x-\tilde{A}x^\dagger\|_{L^1(T,\mu)}\quad\text{for all } x\in\tilde{X}$$

with $c>0$ and Proposition 12.14 with φ replaced by the concave and monotonically increasing function $\tilde{\varphi}(t):=\sqrt{\|\tilde{A}x^\dagger\|_{L^1(T,\mu)}+t}-\sqrt{\|\tilde{A}x^\dagger\|_{L^1(T,\mu)}}$, $t\in[0,\infty)$, yields

$$\tilde{\beta}B_{\xi^\dagger}^{\tilde{\Omega}}(x,x^\dagger)\leq\tilde{\Omega}(x)-\tilde{\Omega}(x^\dagger)+\bar{c}\tilde{\varphi}\bigl(\|\tilde{A}x-\tilde{A}x^\dagger\|_{L^1(T,\mu)}\bigr)\quad\text{for all } x\in\tilde{X}$$

with $\bar{c}>0$. For $x\in X$ we have $\tilde{\varphi}\bigl(\|\tilde{A}x-\tilde{A}x^\dagger\|_{L^1(T,\mu)}\bigr)=\tilde{\varphi}\bigl(\|F(x)-F(x^\dagger)\|_{L^1(T,\mu)}\bigr)$ and Lemma 9.11, in combination with Proposition 9.9, provides $\tilde{\varphi}\bigl(\|F(x)-F(x^\dagger)\|_{L^1(T,\mu)}\bigr)\leq\sqrt{S_Y(F(x),F(x^\dagger)}$. Taking into account $\tilde{\Omega}=\Omega$ on X we arrive at

$$\tilde{\beta}B_{\xi^\dagger}^{\tilde{\Omega}}(x,x^\dagger)\leq\Omega(x)-\Omega(x^\dagger)+\bar{c}\sqrt{S_Y\bigl(F(x),F(x^\dagger)\bigr)}\quad\text{for all } x\in X.$$

□

The assumption (9.3) is discussed in Remark 12.15.

Having the variational inequality (9.4) at hand we can apply the convergence rates theorems of Part I (see Theorem 4.11, Theorem 4.24, Theorem 5.4) to the continuous setting for regularization with Poisson distributed data.

10. Numerical example

The present chapter shall provide a glimpse on the influence of the fitting functional on regularized solutions. We consider the semi-discrete setting for regularization with Poisson distributed data described in Chapter 8. The aim is not to present an exhaustive numerical study but to justify the effort we have undertaken to analyze non-metric fitting functionals in Part I. A detailed numerical investigation of different noise adapted fitting functionals would be very interesting and should be realized in future. At the moment only few results on this question can be found in the literature (see, e.g., [Bar08]).

10.1. Specification of the test case

Since Poisson distributed data mainly occur in imaging we aim to solve a typical deblurring problem. We restrict ourselves to a one-dimensional setting. This saves computation time and allows for a detailed illustration of the results. Next to the fitting functional and the operator a major decision is the choice of a suitable stabilizing functional Ω. Common variants in the field of imaging are total variation penalties and sparsity constraints. For our experiments we use a sparsity constraint with respect to the Haar base. In the sequel of this section we make things more precise.

10.1.1. Haar wavelets

We consider square-integrable functions over the interval $(0,1)$. Each such function $u \in L^2(0,1)$ can be decomposed with respect to the *Haar base* $\{\phi_{0,0}\} \cup \{\psi_{k,l} : l \in \mathbb{N}_0,\ k = 0,\dots,2^l-1\} \subseteq L^2(0,1)$, where we define $\phi, \phi_{l,k}, \psi, \psi_{l,k} \in L^2(0,1)$ by

$$\phi :\equiv 1, \quad \psi(t) := \begin{cases} 1, & t \in \left(0, \frac{1}{2}\right], \\ -1, & t \in \left(\frac{1}{2}, 1\right), \end{cases} \quad \phi_{l,k} := 2^{\frac{l}{2}} \phi(2^l \bullet - k), \quad \psi_{l,k} := 2^{\frac{l}{2}} \psi(2^l \bullet - k)$$

for $l \in \mathbb{N}_0$ and $k = 0,\dots,2^l-1$. The decomposition coefficients are given by

$$c_{0,0} := \langle u, \phi_{0,0} \rangle, \quad d_{l,k} := \langle u, \psi_{l,k} \rangle \quad \text{for } l \in \mathbb{N}_0 \text{ and } k = 0,\dots,2^l-1.$$

These coefficients can be arranged as a sequence $x := (x_j)_{j \in \mathbb{N}} \in l^2(\mathbb{N})$ by

$$x_1 := c_{0,0}, \quad x_{1+2^l+k} := d_{l,k} \quad \text{for } l \in \mathbb{N}_0 \text{ and } k = 0,\dots,2^l-1. \tag{10.1}$$

The decomposition process is a linear mapping $W : L^2(0,1) \to l^2(\mathbb{N})$, $u \mapsto x$ and the synthesis of u from x is given by

$$V : l^2(\mathbb{N}_0) \to L^2(0,1), \quad x \mapsto c_{0,0}\phi_{0,0} + \sum_{l=0}^{\infty} \sum_{k=0}^{2^l-1} d_{l,k} \psi_{l,k}.$$

A fast method for obtaining the corresponding coefficients is the *fast wavelet transform*: Given a piecewise constant function $u = \sum_{j=1}^{n} u_j \chi_{(\frac{j-1}{n}, \frac{j}{n})}$ with $n := 2^{\bar{l}+1}$, $\bar{l} \in \mathbb{N}_0$, $u_j \in \mathbb{R}$, and $\chi_{(a,b)}$ being one on (a, b) and zero outside this interval we first define

$$c_{\bar{l}+1,k} := \langle u, \phi_{\bar{l}+1,k} \rangle = 2^{-\frac{\bar{l}+1}{2}} u_{k+1} \quad \text{for } k = 0, \dots, 2^{\bar{l}+1} - 1.$$

Then the numbers

$$c_{l,k} := \langle u, \phi_{l,k} \rangle = \frac{1}{\sqrt{2}}(c_{l+1,2k} + c_{l+1,2k+1}), \quad \text{for } k = 0, \dots, 2^l - 1$$

and

$$d_{l,k} := \langle u, \psi_{l,k} \rangle = \frac{1}{\sqrt{2}}(c_{l+1,2k} - c_{l+1,2k+1}) \quad \text{for } k = 0, \dots, 2^l - 1$$

have to be computed gradually for $l = \bar{l}, \dots, 0$. Now the function u can be written as

$$u = c_{0,0}\phi_{0,0} + \sum_{l=0}^{\bar{l}} \sum_{k=0}^{2^l-1} d_{l,k}\psi_{l,k}.$$

There is also a fast algorithm for the synthesis of u from its Haar coefficients: Given $c_{0,0}$ and $d_{l,k}$ for $l = 0, \dots, \bar{l}$ and $k = 0, \dots, 2^l - 1$ we gradually compute

$$c_{l+1,k} := \begin{cases} \frac{1}{\sqrt{2}}\big(c_{l,\frac{k}{2}} + d_{l,\frac{k}{2}}\big), & k \text{ is even}, \\ \frac{1}{\sqrt{2}}\big(c_{l,\frac{k-1}{2}} - d_{l,\frac{k-1}{2}}\big), & k \text{ is odd}, \end{cases} \qquad k = 0, \dots, 2^{l+1} - 1$$

for $l = 0, \dots, \bar{l}$. Then u is obtained from

$$u = \sum_{j=1}^{n} u_j \chi_{(\frac{j-1}{n}, \frac{j}{n})} \quad \text{with} \quad u_j = 2^{\frac{\bar{l}+1}{2}} c_{\bar{l}+1,j-1}.$$

Arranging the Haar coefficients as a sequence $x = (x_j)_{j \in \mathbb{N}} \in l^2(\mathbb{N})$ (see (10.1)), only the first $n = 2^{\bar{l}+1}$ elements of x are non-zero in case of the piecewise constant function u from above. That is, the vector $\underline{x} := [x_1, \dots, x_n]^{\mathrm{T}}$ contains the same information as the vector $\underline{u} := [u_1, \dots, u_n]^{\mathrm{T}}$. From this viewpoint the fast wavelet transform (as a linear mapping) can be expressed by a matrix $\underline{W} \in \mathbb{R}^{n \times n}$ and the inverse transform (synthesis) by a matrix $\underline{V} \in \mathbb{R}^{n \times n}$. Note, that $\frac{1}{\sqrt{n}}\underline{W}$ and $\frac{1}{\sqrt{n}}\underline{V}$ are orthonormal matrices, that is, $\underline{W}^{\mathrm{T}}\underline{W} = n\underline{I}$ and $\underline{V}^{\mathrm{T}}\underline{V} = n\underline{I}$. The corresponding inverse matrices satisfy

$$\underline{V} = \underline{W}^{-1} = \frac{1}{n}\underline{W}^{\mathrm{T}} \quad \text{and} \quad \underline{W} = \underline{V}^{-1} = \frac{1}{n}\underline{V}^{\mathrm{T}}.$$

In the sequel we only use the matrix $\underline{V}$ and the relations

$$\underline{x} = \frac{1}{n}\underline{V}^{\mathrm{T}}\underline{u} \quad \text{and} \quad \underline{u} = \underline{V}\underline{x}.$$

Further details on the Haar system are given in all books on wavelets (e.g. [Dau92, GC99]).

10.1.2. Operator and stabilizing functional

In our numerical tests we apply a linear blurring operator $B : U \to Y$, where $U := \{u \in L^2(0,1) : u \geq 0 \text{ a.e.}\}$ and Y is as in Chapter 8 with $T := (0,1)$, $T_i := (\frac{i-1}{n}, \frac{i}{n})$ for $i = 1, \dots, n$, and μ being the Lebesgue measure on T (see Chapter 8 for details on the semi-discrete setting under consideration). For simplicity we use a constant kernel $\frac{1}{b}\chi_{[-\frac{b}{2},\frac{b}{2}]}$ of width $b \in (0,1)$ and extend the functions $u \in U$ by reflection to $\mathbb{R}$: for $k \in \mathbb{Z}$ and $t \in (0,1)$ define

$$\tilde{u}(k+t) := \begin{cases} u(t), & k \text{ is even}, \\ u(1-t), & k \text{ is odd}. \end{cases}$$

Then the operator B is given by

$$\begin{aligned}(Bu)(t) &= \int_{-\infty}^{\infty} \frac{1}{b}\chi_{[-\frac{b}{2},\frac{b}{2}]}(t-s)\tilde{u}(s)\,\mathrm{d}s = \frac{1}{b}\int_{-\frac{b}{2}}^{\frac{b}{2}} \tilde{u}(t-s)\,\mathrm{d}s \\ &= \begin{cases} \frac{1}{b}\int_{-\frac{b}{2}}^{t} u(t-s)\,\mathrm{d}s + \frac{1}{b}\int_{t}^{\frac{b}{2}} u(s-t)\,\mathrm{d}s, & 0 < t < \frac{b}{2}, \\ \frac{1}{b}\int_{-\frac{b}{2}}^{\frac{b}{2}} u(t-s)\,\mathrm{d}s, & \frac{b}{2} \leq t \leq 1-\frac{b}{2}, \\ \frac{1}{b}\int_{-\frac{b}{2}}^{t-1} u(2+s-t)\,\mathrm{d}s + \frac{1}{b}\int_{t-1}^{\frac{b}{2}} u(t-s)\,\mathrm{d}s, & 1-\frac{b}{2} < t < 1. \end{cases}\end{aligned}$$

Instead of a function $u \in U$ we would like to work with its Haar coefficients, that is, the final operator $F : X \to Y$ in equation (1.1) reads as $F := B \circ V$ (restricted to X) with V from Subsection 10.1.1 and $X := \{x \in l^2(\mathbb{N}) : Vx \geq 0 \text{ a.e.}\}$. For the analysis carried out in Part I the topology τ_X shall be induced by the weak topology on $l^2(\mathbb{N})$; but for the numerical experiments this choice does not matter.

The reason for using the Haar decomposition is the choice of the stabilizing functional $\Omega : X \to [0,\infty]$. It shall penalize the values and the number of the Haar coefficients. This can be achieved by

$$\Omega(x) := \sum_{j=1}^{\infty} w_j |x_j|,$$

where $(w_j)_{j\in\mathbb{N}}$ is a bounded sequence of weights satisfying $w_j > 0$ for all $j \in \mathbb{N}$. Such sparsity constraints have been studied in detail during the last decade in the context of inverse problems and turned out to be well suited for stabilizing imaging problems (see, e.g., [DDDM04, FSBM10]).

Proposition 10.1. *The operator $F : X \to Y$ is continuous with respect to τ_X and τ_Y. The functional $\Omega : X \to [0,\infty]$ has τ_X-compact sublevel sets.*

Proof. To prove continuity of F it suffices to show that $B : L^2(0,1) \to L^1(0,1)$ is continuous with respect to the norm topologies. This is the case because for each $u \in L^2(0,1)$ we have

$$\begin{aligned}\|Bu\|_{L^1(0,1)} &\le \frac{1}{b} \int_0^1 \int_{-\frac{b}{2}}^{\frac{b}{2}} |\tilde{u}(t-s)| \,\mathrm{d}s \,\mathrm{d}t \le \frac{1}{b} \int_0^1 \int_{-\frac{1}{2}}^{\frac{3}{2}} |\tilde{u}(s)| \,\mathrm{d}s \,\mathrm{d}t \\ &\le \frac{2}{b} \int_0^1 \int_0^1 |u(s)| \,\mathrm{d}s \,\mathrm{d}t \le \frac{2}{b} \|u\|_{L^2(0,1)}.\end{aligned}$$

We now show the weak compactness of the sublevel sets $M_{\tilde{\Omega}}(c) := \{x \in l^2(\mathbb{N}) : \tilde{\Omega}(x) \le c\}$ of the functional $\tilde{\Omega} : l^2(\mathbb{N}) \to [0,\infty]$ defined by

$$\tilde{\Omega}(x) := \sum_{j=1}^{\infty} w_j |x_j|.$$

This functional is known to be weakly lower semi-continuous (that is, the sets $M_{\tilde{\Omega}}(c)$ are weakly closed) and coercive (see, e.g., [Gra10c]). Coercivity means that $\tilde{\Omega}(x) \to \infty$ if $\|x\|_{l^2(\mathbb{N})} \to \infty$; therefore each sequence in $M_{\tilde{\Omega}}(c)$ is bounded. Since bounded sequences in a separable Hilbert space contain weakly convergent subsequences, the sets $M_{\tilde{\Omega}}(c)$ are relatively weakly compact.

The sublevel sets $M_\Omega(c) := \{x \in X : \Omega(x) \le c\}$ of Ω can be written as $M_\Omega(c) = M_{\tilde{\Omega}}(c) \cap X$. The set X as a subset of $l^2(\mathbb{N})$ is closed and convex, and hence also weakly closed. Since the intersection of a weakly compact and a weakly closed set remains weakly compact, the assertion is true. □

The proposition and the results of Section 8.1 show that the basic Assumption 2.1 is satisfied and thus the theory developed in Part I applies to the present example.

10.1.3. Discretization

For given data $\underline{z} = [z_1, \dots, z_m]^{\mathrm{T}} \in [0,\infty)^m$ we aim to approximate a minimizer of the Tikhonov-type functional

$$T_\alpha^{\underline{z}}(x) = \sum_{i=1}^{m} s(D_i F(x), z_i) + \alpha\Omega(x), \quad x \in X,$$

numerically (note that we write $\underline{z}$ instead of z in this chapter to emphasize that it is a finite-dimensional vector). The function s is defined in (7.2) and the operator F and the stabilizing functional Ω have been introduced in Subsection 10.1.2. Remember the definition of the operators D_i:

$$D_i y = \int_{T_i} y \,\mathrm{d}\mu = \int_{\frac{i-1}{m}}^{\frac{i}{m}} y(t) \,\mathrm{d}t, \quad i = 1, \dots, m.$$

We discretize the minimization problem by cutting off the Haar coefficients of fine levels l, which results in restricting the minimization to a finite-dimensional subspace of X. For $\bar{l} \in \mathbb{N}$ define

$$X_{\bar{l}} := \big\{x \in X : x_j = 0 \text{ for } j > 2^{\bar{l}+1}\big\};$$

the dimension of this subspace is $n := 2^{\bar{l}+1}$ (here 'dimension' means the dimension of the affine hull of $X_{\bar{l}}$ in $l^2(\mathbb{N})$).

For $x \in X_{\bar{l}}$ the corresponding function $Vx \in L^2(0,1)$ (see Subsection 10.1.1 for the definition of V) is piecewise constant and we may write it as

$$Vx = \sum_{j=1}^{n} u_j \chi_{(\frac{j-1}{n}, \frac{j}{n})}$$

with $\underline{u} = [u_1, \dots, u_n]^{\mathrm{T}} \in [0,\infty)^n$. Identifying $x = (x_1, \dots, x_n, 0, \dots) \in X_{\bar{l}}$ with $\underline{x} = [x_1, \dots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$ yields $\underline{u} = \underline{V}\underline{x}$ with $\underline{V}$ from Subsection 10.1.1 and therefore

$$X_{\bar{l}} = \big\{x = (x_1, \dots, x_n, 0, \dots) \in l^2(\mathbb{N}) : \underline{V}\underline{x} \geq \underline{0}\big\}.$$

Since $DBVx$ is in $[0,\infty)^m$, the restriction of the operator $D \circ B$ to piecewise constant functions Vx, $x \in X_{\bar{l}}$, can be expressed by a matrix $\underline{B} \in \mathbb{R}^{m \times n}$. Then $DF(x) = DBVx = \underline{B}\underline{V}\underline{x}$. For later use we set $\underline{A} := \underline{B}\underline{V}$. The matrix $\underline{B}$ is given below for the special case $m = n$. On $X_{\bar{l}}$ the stabilizing functional Ω reduces to $\Omega(x) = \sum_{j=1}^{n} w_j |x_j|$.

Next we verify that the chosen discretization yields arbitrarily exact approximations of minimizers of T_α^z over X if $\bar{l}$ is large enough.

Proposition 10.2. *The sequence $(X_{\bar{l}})_{\bar{l} \in \mathbb{N}}$ of τ_X-closed subspaces $X_1 \subseteq X_2 \subseteq \cdots \subseteq X$ satisfies Assumption 3.6 (with n replaced by $\bar{l}$ there).*

Proof. For $x \in X$ choose the sequence $(x_{\bar{l}})_{\bar{l} \in \mathbb{N}}$ with $x_{\bar{l}} := (x_1, \dots, x_{2^{\bar{l}+1}}, 0, \dots) \in X_{\bar{l}}$ (by the construction of the Haar system we have $Vx_{\bar{l}} \geq 0$ a.e. if $Vx \geq 0$ a.e.). Then $\|x_{\bar{l}} - x\|_{l^2(\mathbb{N})} \to 0$ if $\bar{l} \to \infty$.

From the proof of Proposition 10.1 we know that B is continuous with respect to the norm topologies on $L^2(0,1)$ and $L^1(0,1)$. Therefore $D \circ B \circ V$ is continuous with respect to the norm topologies on $l^2(\mathbb{N})$ and $\mathbb{R}^m$, that is, $\|DF(x_{\bar{l}}) - DF(x)\|_{\mathbb{R}^m} \to 0$ if $\bar{l} \to \infty$. Since $s(\bullet, w) : [0,\infty) \to [0,\infty]$ is continuous for each $w \in [0,\infty)$, we obtain $\sum_{i=1}^{m} s(D_i F(x_{\bar{l}}), z_i) \to \sum_{i=1}^{m} s(D_i F(x), z_i)$ if $\bar{l} \to \infty$ for all $\underline{z} \in [0,\infty)^m$.

For the stabilizing functional we obviously have

$$\Omega(x_{\bar{l}}) = \sum_{j=1}^{2^{\bar{l}+1}} w_j |x_j| \to \sum_{j=1}^{\infty} w_j |x_j| = \Omega(x)$$

if $\bar{l} \to \infty$. □

The proposition shows that Corollary 3.10 in Section 3.5 on the discretization of Tikhonov-type minimization problems applies to the present example.

For simplicity we assume $m = n$ in the sequel, that is, the solution space $X_{\bar{l}}$ shall have the same dimension as the data space $Z = [0,\infty)^m$. In this case we can explicitly

calculate the matrix $\underline{B} \in \mathbb{R}^{n\times n}$. If the kernel width b of the blurring operator B is of the form $b = \frac{2q}{n}$ with $q \in \{1, \dots, \frac{n}{2} - 1\}$, then simple but lengthy calculations show that the elements b_{ij} of the matrix $\underline{B}$ are given by

$$b_{ij} = \begin{cases} \frac{4}{4qn} & \text{for } i = 1, \dots, q \text{ and } j = 1, \dots, q - i, \\ \frac{3}{4qn} & \text{for } i = 1, \dots, q \text{ and } j = q - i + 1, \\ \frac{2}{4qn} & \text{for } i = 1, \dots, q \text{ and } j = q - i + 2, \dots, q + i - 1, \\ \frac{1}{4qn} & \text{for } i = 1, \dots, q \text{ and } j = q + i, \\ \frac{1}{4qn} & \text{for } i = q + 1, \dots, n - q \text{ and } j = i - q, \\ \frac{2}{4qn} & \text{for } i = q + 1, \dots, n - q \text{ and } j = i - q + 1, \dots, i + q - 1, \\ \frac{1}{4qn} & \text{for } i = q + 1, \dots, n - q \text{ and } j = i + q, \\ \frac{1}{4qn} & \text{for } i = n - q + 1, \dots, n \text{ and } j = i - q, \\ \frac{2}{4qn} & \text{for } i = n - q + 1, \dots, n \text{ and } j = i - q + 1, \dots, 2n - i - q, \\ \frac{3}{4qn} & \text{for } i = n - q + 1, \dots, n \text{ and } j = 2n - q - i + 1, \\ \frac{4}{4qn} & \text{for } i = n - q + 1, \dots, n \text{ and } j = 2n - q - i + 2, \dots, n, \\ 0 & \text{else.} \end{cases}$$

Eventually, the discretized minimization problem reads as

$$\sum_{i=1}^{n} s([\underline{Ax}]_i, z_i) + \alpha \sum_{j=1}^{n} w_j |x_j| \to \min_{\underline{x} \in \mathbb{R}^n : \underline{Vx} \geq \underline{0}}.$$

The nonnegativity constraint can also be included in the stabilizing functional by setting

$$\underline{\Omega}(\underline{x}) := \begin{cases} \sum_{j=1}^{n} w_j |x_j|, & \underline{Vx} \geq \underline{0}, \\ \infty, & \text{else} \end{cases}$$

for $\underline{x} \in \mathbb{R}^n$. Then the minimization problem to be solved becomes

$$\sum_{i=1}^{n} s([\underline{Ax}]_i, z_i) + \alpha \underline{\Omega}(\underline{x}) \to \min_{\underline{x} \in \mathbb{R}^n}. \tag{10.2}$$

At points $\underline{x}$ where $\sum_{i=1}^{n} s([\underline{Ax}]_i, z_i)$ is not defined the stabilizing functional is infinite and thus the whole objective function can be assumed to be infinite at such points.

In the sequel we only consider the discretized minimization problem (10.2). But the analytic results and also the minimization algorithm are expected to work in infinite dimensions, too. Only few modifications would be required.

10.2. An optimality condition

Before we provide an algorithm for solving the minimization problem (10.2) we have to think about a stopping criterion. If the objective function would be differentiable,

then we would search for vectors $\underline{x} \in \mathbb{R}^n$ where the gradient norm of the objective function becomes zero (or at least very small). The objective function in (10.2) is not differentiable. Thus, we have to apply techniques from non-smooth convex analysis.

At first we calculate the gradient of the mapping

$$S^{\underline{z}} : X_{\bar{l}} \to [0, \infty], \quad S^{\underline{z}}(\underline{x}) := \sum_{i=1}^{n} s([\underline{Ax}]_i, z_i). \tag{10.3}$$

For some vector $\underline{v} \in \mathbb{R}^n$ we define sets

$$I_0(\underline{v}) := \{i \in \{1, \dots, n\} : v_i = 0\} \quad \text{and} \quad I_*(\underline{v}) := \{1, \dots, n\} \setminus I_0(v).$$

Then $S^{\underline{z}}(\underline{x}) = \infty$ if $I_0(\underline{Ax}) \cap I_*(\underline{z}) \neq \emptyset$ and

$$S^{\underline{z}}(\underline{x}) = \sum_{i \in I_0(\underline{z})} [\underline{Ax}]_i + \sum_{i \in I_*(\underline{z})} \left(z_i \ln \frac{z_i}{[\underline{Ax}]_i} + [\underline{Ax}]_i - z_i \right) < \infty \tag{10.4}$$

if $I_0(\underline{Ax}) \cap I_*(\underline{z}) = \emptyset$. In the last case we used $I_*(\underline{Ax}) \cap I_*(\underline{z}) = I_*(\underline{z})$. The essential domain of $S^{\underline{z}}$ is thus given by

$$D(S^{\underline{z}}) := \{\underline{x} \in X_{\bar{l}} : I_0(\underline{Ax}) \cap I_*(\underline{z}) = \emptyset\}.$$

Lemma 10.3. *For all $\underline{x} \in D(S^{\underline{z}})$ the mapping $S^{\underline{z}}$ has partial derivatives $\frac{\partial}{\partial x_j} S^{\underline{z}}(\underline{x})$, $j = 1, \dots, n$. If $I_0(\underline{Ax}) \cap I_0(\underline{z}) \neq \emptyset$ for some $\underline{x} \in D(S^{\underline{z}})$, then the corresponding partial derivatives have to be understood as one-sided derivatives. The gradient $\nabla S^{\underline{z}}$ is given by*

$$\nabla S^{\underline{z}}(\underline{x}) = \underline{A}^{\mathrm{T}} h(\underline{Ax}), \quad \underline{x} \in D(S^{\underline{z}}),$$

with

$$h : [0, \infty)^n \to [0, \infty), \quad [h(\underline{y})]_i := \begin{cases} 1, & i \in I_0(\underline{z}), \\ 1 - \frac{z_i}{y_i}, & i \in I_*(\underline{z}). \end{cases}$$

Proof. The assertion follows from (10.4) and from the chain rule. □

Obviously a minimizer $\underline{x}^* \in \mathbb{R}^n$ of (10.2) has to satisfy $\underline{x}^* \in D(S^{\underline{z}})$ and $\underline{Vx}^* \geq \underline{0}$. We state a first optimality criterion in terms of subgradients.

Lemma 10.4. *A vector $\underline{x}^* \in D(S^{\underline{z}})$ with $\underline{Vx}^* \geq \underline{0}$ minimizes (10.2) if and only if*

$$-\frac{1}{\alpha} \nabla S^{\underline{z}}(\underline{x}^*) \in \partial \underline{\Omega}(\underline{x}^*).$$

Proof. This is a standard result in convex analysis. See, e.g., [ABM06, Proposition 9.5.3] in combination with [ABM06, Theorem 9.5.4]. □

Lemma 10.5. *Let $\underline{x} \in \mathbb{R}^n$ satisfy $\underline{Vx} \geq \underline{0}$. A vector $\underline{\xi} \in \mathbb{R}^n$ belongs to the subdifferential $\partial \underline{\Omega}(\underline{x})$ if and only if it has the form $\underline{\xi} = \underline{\eta} + \underline{V}^{\mathrm{T}} \underline{\zeta}$ with vectors $\underline{\eta}, \underline{\zeta} \in \mathbb{R}^n$ satisfying*

$$\eta_j = (\operatorname{sgn} x_j) w_j \quad \text{if } x_j \neq 0, \qquad \eta_j \in [-w_j, w_j] \quad \text{if } x_j = 0$$

and

$$\zeta_i = 0 \quad \text{if } [\underline{Vx}]_i > 0, \qquad \zeta_i \in (-\infty, 0] \quad \text{if } [\underline{Vx}]_i = 0.$$

Proof. We write $\underline{\Omega} = \Gamma + \Lambda$ with

$$\Gamma(\underline{x}) := \sum_{j=1}^{n} w_j |x_j| \quad \text{and} \quad \Lambda(\underline{x}) := \begin{cases} 0, & \underline{Vx} \geq \underline{0}, \\ \infty, & \text{else} \end{cases}$$

for $\underline{x} \in \mathbb{R}^n$. Then $\partial\underline{\Omega}(\underline{x}) = \partial\Gamma(\underline{x}) + \partial\Lambda(\underline{x})$ for $\underline{x}$ with $\underline{Vx} \geq \underline{0}$ (cf. [ABM06, Theorem 9.5.4]). Thus, it suffices to show that $\partial\Gamma(\underline{x})$ consists exactly of the vectors $\underline{\eta}$ described in the lemma and that $\partial\Lambda(\underline{x})$ consists exactly of the vectors $\underline{V}^{\mathrm{T}}\underline{\zeta}$ with $\underline{\zeta}$ as in the lemma.

So let $\underline{\eta} \in \partial\Gamma(\underline{x})$. Then the subgradient inequality reads as

$$\sum_{j=1}^{n} w_j(|\tilde{x}_j| - |x_j|) \geq \sum_{j=1}^{n} \eta_j(\tilde{x}_j - x_j) \quad \text{for all } \underline{\tilde{x}} \in \mathbb{R}^n. \tag{10.5}$$

Fixing j and setting all but the j-th component of $\underline{\tilde{x}}$ to the values of the corresponding components of $\underline{x}$ the inequality reduces to

$$w_j(|\tilde{x}_j| - |x_j|) \geq \eta_j(\tilde{x}_j - x_j) \quad \text{for all } \tilde{x}_j \in \mathbb{R}.$$

If $x_j > 0$, then $\tilde{x}_j := 1 + x_j$ and $\tilde{x}_j := 0$ lead to $\eta_j = w_j$. For $x_j < 0$ we set $\tilde{x}_j := -1 + x_j$ and $\tilde{x}_j := 0$, which gives $\eta_j = -w_j$. And in the case $x_j = 0$ the bounds $-w_j \leq \eta_j \leq w_j$ follow from $\tilde{x}_j := \pm 1$.

Now let $\underline{\eta}$ be as in the lemma. Then the subgradient inequality (10.5) follows immediately from the three implications

$$\begin{aligned} x_j > 0 \quad &\Rightarrow \quad \eta_j(\tilde{x}_j - x_j) = w_j(\tilde{x}_j - |x_j|) \leq w_j(|\tilde{x}_j| - |x_j|), \\ x_j < 0 \quad &\Rightarrow \quad \eta_j(\tilde{x}_j - x_j) = w_j(-\tilde{x}_j - |x_j|) \leq w_j(|\tilde{x}_j| - |x_j|), \\ x_j = 0 \quad &\Rightarrow \quad \eta_j(\tilde{x}_j - x_j) = \eta_j \tilde{x}_j \leq |\eta_j||\tilde{x}_j| \leq w_j|\tilde{x}_j|. \end{aligned}$$

Let $\underline{\tilde{\zeta}} \in \partial\Lambda(\underline{x})$. Since $\underline{V}$ is invertible the vector $\underline{V}^{-T}\underline{\tilde{\zeta}}$ is well-defined and if we could show that $\underline{V}^{-T}\underline{\tilde{\zeta}}$ is of the same form as $\underline{\zeta}$ in the lemma, then $\underline{\tilde{\zeta}} = V^T\big(\underline{V}^{-T}\underline{\tilde{\zeta}}\big)$ satisfies the desired representation.

The subgradient inequality is equivalent to

$$0 \geq \big(\underline{V}^{-T}\underline{\tilde{\zeta}}\big)^{\mathrm{T}}(\underline{V\tilde{x}} - \underline{Vx}) \quad \text{for all } \underline{\tilde{x}} \in \mathbb{R}^n \text{ with } \underline{V\tilde{x}} \geq \underline{0}. \tag{10.6}$$

Thus, setting $\underline{\tilde{x}} := \underline{x} + \underline{V}^{-1}\underline{e}_i$ the inequality implies $\big[\underline{V}^{-T}\underline{\tilde{\zeta}}\big]_i \leq 0$ (here $\underline{e}_i$ has a one in the i-th component and zeros in all other components). And if $[\underline{Vx}]_i > 0$, then $\underline{\tilde{x}} := \underline{x} - [\underline{Vx}]_i\underline{V}^{-1}\underline{e}_i$ yields $\big[\underline{V}^{-T}\underline{\tilde{\zeta}}\big]_i \geq 0$.

Finally, we have to show that $\underline{\tilde{\zeta}} := \underline{V}^{\mathrm{T}}\underline{\zeta}$ with $\underline{\zeta}$ as in the lemma belongs to $\partial\Lambda(\underline{x})$. Therefore, we observe

$$\underline{\tilde{\zeta}}^{\mathrm{T}}(\underline{\tilde{x}} - \underline{x}) = \underline{\zeta}^{\mathrm{T}}(V\underline{\tilde{x}} - V\underline{x}) \leq 0 \quad \text{for all } \underline{\tilde{x}} \in \mathbb{R}^n \text{ with } \underline{V\tilde{x}} \geq \underline{0}.$$

This is equivalent to (10.6) and (10.6) is equivalent to the subgradient inequality. □

Combining the two lemmas we have the following optimality criterion: A vector $\underline{x}^*$ is a minimizer of the discretized Tikhonov functional (10.2) if and only if

$$-\frac{1}{\alpha}\nabla S^{\underline{z}}(\underline{x}^*) = \underline{\eta} + \underline{V}^{\mathrm{T}}\underline{\zeta}$$

with $\underline{\eta}$ and $\underline{\zeta}$ from the previous lemma. There is no chance to check this condition numerically. Thus, we cannot use it for stopping a minimization algorithm.

The following proposition provides a more useful characterization of the subgradients of $\underline{\Omega}$.

Proposition 10.6. *Let $\underline{x} \in \mathbb{R}^n$ satisfy $\underline{Vx} \geq \underline{0}$ and denote the elements of $\underline{V}$ by v_{ij}, $i, j = 1, \ldots, n$. A vector $\underline{\xi} \in \mathbb{R}^n$ belongs to the subdifferential $\partial\underline{\Omega}(\underline{x})$ if and only if*

$$\sum_{j=1}^{n} \big(\xi_j - (\operatorname{sgn} x_j) w_j\big) x_j = 0 \quad \textit{and}$$

$$\sum_{j=1}^{n} \big(\xi_j - (\operatorname{sgn} x_j) w_j\big) v_{ij} \leq \sum_{j:x_j=0} w_j |v_{ij}| \quad \textit{for } i = 1, \ldots, n.$$

Proof. From Lemma 10.5 we know that $\underline{\xi} \in \partial\underline{\Omega}(\underline{x})$ if and only if there are $\underline{\eta}, \underline{\zeta} \in \mathbb{R}^n$ as in the lemma such that $\underline{\xi} - \underline{\eta} = \underline{V}^{\mathrm{T}}\underline{\zeta}$ or, equivalently, $\underline{V}^{-\mathrm{T}}(\underline{\xi} - \underline{\eta}) = \underline{\zeta}$. We reformulate the last equality as

$$(\underline{\xi} - \underline{\eta})^{\mathrm{T}}\underline{x} = 0, \quad \underline{V}^{-\mathrm{T}}(\underline{\xi} - \underline{\eta}) \leq \underline{0}, \tag{10.7}$$

that is, we suppress $\underline{\zeta}$. Necessity follows from $(\underline{\xi}-\underline{\eta})^{\mathrm{T}}\underline{x} = \big(\underline{V}^{-T}(\underline{\xi}-\underline{\eta})\big)^{\mathrm{T}}\underline{Vx} = \underline{\zeta}^{\mathrm{T}}\underline{Vx} = 0$ and $\underline{V}^{-\mathrm{T}}(\underline{\xi} - \underline{\eta}) = \underline{\zeta} \leq \underline{0}$. To show sufficiency we set $\underline{\zeta} := \underline{V}^{-\mathrm{T}}(\underline{\xi} - \underline{\eta})$ and verify that $\underline{\zeta}$ is as in Lemma 10.5. Obviously $\underline{\zeta} \leq \underline{0}$. If there would by an index i with $[\underline{Vx}]_i > 0$ and $\underline{\zeta}_i < 0$, then we would obtain the contradiction $0 = (\underline{\xi} - \underline{\eta})^{\mathrm{T}}\underline{x} = \underline{\zeta}^{\mathrm{T}}\underline{Vx} \leq \underline{\zeta}_i[\underline{Vx}]_i < 0$.

Next, we write (10.7) as

$$\sum_{j=1}^{n} \big(\xi_j - (\operatorname{sgn} x_j) w_j\big) x_j = 0, \quad \underline{V}(\underline{\xi} - \underline{\eta}) \leq \underline{0}$$

(remember $\underline{V}^{-\mathrm{T}} = \frac{1}{n}\underline{V}$, see Subsection 10.1.1). Thus, it remains to show

$$\big[\underline{V}(\underline{\xi} - \underline{\eta})\big]_i \leq 0 \quad \Leftrightarrow \quad \sum_{j=1}^{n} \big(\xi_j - (\operatorname{sgn} x_j) w_j\big) v_{ij} - \sum_{j:x_j=0} w_j |v_{ij}| \leq 0$$

for each $i \in \{1, \ldots, n\}$. To verify the '$\Rightarrow$' direction, first note that

$$\sum_{j=1}^{n} \big(\xi_j - (\operatorname{sgn} x_j) w_j\big) v_{ij} - \sum_{j:x_j=0} w_j |v_{ij}| = \sum_{j:x_j\neq 0} (\xi_j - \eta_j) v_{ij} + \sum_{j:x_j=0} (\xi_j v_{ij} - w_j |v_{ij}|).$$

For j with $x_j = 0$ we have $w_j \geq (\operatorname{sgn} v_{ij})\eta_j$, which implies

$$\sum_{j:x_j\neq 0} (\xi_j - \eta_j) v_{ij} + \sum_{j:x_j=0} (\xi_j v_{ij} - w_j |v_{ij}|) \leq \sum_{j=1}^{n} (\xi_j - \eta_j) v_{ij} = \big[\underline{V}(\underline{\xi} - \underline{\eta})\big]_i \leq 0.$$

The '$\Leftarrow$' direction follows if we set $\eta_j := (\operatorname{sgn} v_{ij}) w_j$ for j with $x_j = 0$. □

The proposition provides a numerical measure of optimality. That is, we can check the optimality criterion in Lemma 10.4 numerically and use it as a stopping criterion for minimization algorithms, or at least to assess the quality of an algorithm's output. We should note, that this optimality measure is not invariant with respect to scaling in $\underline{x}$ and $\underline{\xi} = -\frac{1}{\alpha}\nabla S^{\underline{z}}(\underline{x})$. Dividing the equality in the proposition by $\|\underline{x}\|_2$ might be a good idea, but does not solve the scaling problem completely.

10.3. The minimization algorithm

In this section we describe an algorithm for solving the minimization problem (10.2). We combine the *generalized gradient projection method* investigated in [BL08] with a framework for gradient projection methods applied in [BZZ09]. Under suitable assumptions the authors of [BL08] show convergence of generalized gradient projection methods in infinite-dimensional spaces. Thus, we expect that our concrete algorithm is not too sensitive with respect to the discretization level n. The method works as follows:

Algorithm 10.7.

0. Set starting point
$$\underline{x}_0 := \frac{1}{n}\left(\sum_{i=1}^{n} z_i\right)\underline{V}^{\mathrm{T}}\underline{e},$$
where $\underline{e} := [1,\ldots,1]^{\mathrm{T}}$. One can show $[\underline{A}\underline{x}_0]_\iota = \frac{1}{n}\sum_{i=1}^{n} z_i$ for $\iota = 1,\ldots,n$ and $S^{\underline{z}}(\underline{x}_0) + \alpha\underline{\Omega}(\underline{x}_0) < \infty$ (see (10.3) for the definition of $S^{\underline{z}}$).

1. Iterate for $k = 0, 1, 2, \ldots$:

   2. Calculate the gradient $\nabla S^{\underline{z}}(\underline{x}_k)$ (see Lemma 10.3) and choose a step length $s_k > 0$ (see Subsection 10.3.1).

   3. Set $\underline{x}_{k+\frac{1}{2}}$ to the generalized projection of $\underline{x}_k - s_k\nabla S^{\underline{z}}(\underline{x}_k)$ (see Subsection 10.3.2).

   4. Do a line search in direction of the projected step:

      4a. Set $\lambda_k := 1$. Choose line search parameters $\beta := 10^{-4}$ and $\theta := 0.5$.

      4b. Multiply λ_k by θ until
      $$S^{\underline{z}}\big(\underline{x}_k + \lambda_k(\underline{x}_{k+\frac{1}{2}} - \underline{x}_k)\big) \leq S^{\underline{z}}(\underline{x}_k) + \beta\lambda_k\nabla S^{\underline{z}}(\underline{x}_k)^{\mathrm{T}}(\underline{x}_{k+\frac{1}{2}} - \underline{x}_k).$$

   5. Set $\underline{x}_{k+1} := \underline{x}_k + \lambda_k(\underline{x}_{k+\frac{1}{2}} - \underline{x}_k)$.

   6. If $\|\underline{x}_{k+1} - \underline{x}_k\|_2^2 < 10^{-15}$ and the same was true for the previous 10 iterations, then stop iteration.

The important step is the generalized projection because this technique allows to combine sparsity and nonnegativity constraints.

10.3.1. Step length selection

Following [BZZ09] we choose the step length s_k in the k-th iteration by alternating two so called Barzilai–Borwein step lengths. Such step lengths were introduced in [BB88]. The algorithm works as follows:

Algorithm 10.8. Let $s_{\min} \in (0,1)$ and $s_{\max} > 1$ be bounds for the step length. The step length in the k-th iteration ($k > 0$) of the minimization algorithm is based on an alternation parameter τ_k, on the iterates $\underline{x}_{k-1}$ and $\underline{x}_k$, and on the gradients $\nabla S^{\underline{z}}(\underline{x}_{k-1})$ and $\nabla S^{\underline{z}}(\underline{x}_k)$.

1. If $k = 0$, then set $s_k := 1$ and set the alternation parameter to $\tau_1 := 0.5$.
2. If $k > 0$ and $(\underline{x}_k - \underline{x}_{k-1})^{\mathrm{T}}\big(\nabla S^{\underline{z}}(\underline{x}_k) - \nabla S^{\underline{z}}(\underline{x}_{k-1})\big) \le 0$, then set $s_k := s_{\max}$.
3. If $k > 0$ and $(\underline{x}_k - \underline{x}_{k-1})^{\mathrm{T}}\big(\nabla S^{\underline{z}}(\underline{x}_k) - \nabla S^{\underline{z}}(\underline{x}_{k-1})\big) > 0$, then calculate

$$s_k^{(1)} := \max\left\{ s_{\min}, \min\left\{ \frac{\|\underline{x}_k - \underline{x}_{k-1}\|_2^2}{(\underline{x}_k - \underline{x}_{k-1})^{\mathrm{T}}\big(\nabla S^{\underline{z}}(\underline{x}_k) - \nabla S^{\underline{z}}(\underline{x}_{k-1})\big)}, s_{\max} \right\} \right\}$$

and

$$s_k^{(2)} := \max\left\{ s_{\min}, \min\left\{ \frac{(\underline{x}_k - \underline{x}_{k-1})^{\mathrm{T}}\big(\nabla S^{\underline{z}}(\underline{x}_k) - \nabla S^{\underline{z}}(\underline{x}_{k-1})\big)}{\|\nabla S^{\underline{z}}(\underline{x}_k) - \nabla S^{\underline{z}}(\underline{x}_{k-1})\|_2^2}, s_{\max} \right\} \right\}$$

and set

$$s_k := \begin{cases} s_k^{(1)}, & \text{if } \frac{s_k^{(2)}}{s_k^{(1)}} > \tau_k, \\ s_k^{(2)}, & \text{if } \frac{s_k^{(2)}}{s_k^{(1)}} \le \tau_k, \end{cases} \qquad \tau_{k+1} := \begin{cases} 1.1\tau_k, & \text{if } \frac{s_k^{(2)}}{s_k^{(1)}} > \tau_k, \\ 0.9\tau_k, & \text{if } \frac{s_k^{(2)}}{s_k^{(1)}} \le \tau_k. \end{cases}$$

10.3.2. Generalized projection

Let $\underline{\tilde{x}} \in \mathbb{R}^n$ with $\underline{V\tilde{x}} \ge \underline{0}$ be a given point (the current iterate), let $\underline{p} \in \mathbb{R}^n$ be a step direction (negative gradient), and let $s > 0$ be a step length. The generalized projection of $\underline{\tilde{x}} + s\underline{p}$ with respect to the functional $\alpha\underline{\Omega}$ is defined as the solution of

$$\frac{1}{2}\|\underline{x} - (\underline{\tilde{x}} + s\underline{p})\|_2^2 + s\alpha\underline{\Omega}(\underline{x}) \to \min_{\underline{x}\in\mathbb{R}^n}. \tag{10.8}$$

Thus, in each iteration of Algorithm 10.7 we have to solve this minimization problem. If the structure of $\underline{\Omega}$ is not too complex, there might be an explicit formula for the minimizer.

Assume, for a moment, that we drop the nonnegativity constraint, that is, $\Omega(\underline{x}) = \sum_{j=1}^n w_j|x_j|$. Then the techniques of Section 10.2 applied to (10.8) show that the minimizer is $\underline{x}^* \in \mathbb{R}^n$ with

$$x_j^* = \max\{0, |\tilde{x}_j + sp_j| - s\alpha w_j\}\operatorname{sgn}(\tilde{x}_j + sp_j). \tag{10.9}$$

This expression is known in the literature as soft-thresholding operation. On the other hand we could drop the sparsity constraint, that is, $\underline{\Omega}(\underline{x}) = 0$ if $\underline{Vx} \geq \underline{0}$ and $\underline{\Omega}(\underline{x}) = \infty$ else. Then one easily shows that the minimizer is

$$\underline{x}^* = \underline{V}^{-1}\underline{h} \quad \text{with} \quad h_i := \max\{0, [\underline{V}(\underline{\tilde{x}} + s\underline{p})]_i\}, \tag{10.10}$$

which is a cut-off operation for the function associated with the Haar coefficients $\underline{\tilde{x}} + s\underline{p}$. A composition of soft-thresholding and cut-off operation yields a feasible point ($\underline{Vx} \geq \underline{0}$), but we cannot expect that this point is a minimizer of (10.8). Therefore we have to solve the minimization problem numerically.

We reformulate (10.8) as a convex quadratic minimization problem over $\mathbb{R}^{2n}$: For this purpose write $\underline{x} = \underline{x}^+ - \underline{x}^-$ with $x_j^+ := \max\{0, x_j\}$ and $x_j^- := -\min\{0, x_j\}$. Then $|x_j| = x_j^+ + x_j^-$ and (10.8) is equivalent to

$$\begin{aligned} &\frac{1}{2}\begin{bmatrix}\underline{x}^+\\ \underline{x}^-\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\underline{I} & -\underline{I}\\ -\underline{I} & \underline{I}\end{bmatrix}\begin{bmatrix}\underline{x}^+\\ \underline{x}^-\end{bmatrix} + \begin{bmatrix}-(\underline{\tilde{x}} + s\underline{p}) + s\alpha\underline{w}\\ \underline{\tilde{x}} + s\underline{p} + s\alpha\underline{w}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\underline{x}^+\\ \underline{x}^-\end{bmatrix} \to \min_{[\underline{x}^+, \underline{x}^-]^{\mathrm{T}} \in \mathbb{R}^{2n}}\\ &\text{subject to } \begin{bmatrix}\underline{x}^+\\ \underline{x}^-\end{bmatrix} \geq \underline{0}, \quad \begin{bmatrix}\underline{V} & -\underline{V}\end{bmatrix}\begin{bmatrix}\underline{x}^+\\ \underline{x}^-\end{bmatrix} \geq \underline{0}. \end{aligned}$$

There are many algorithms for solving such minimization problems. We use an interior point method suggested in [NW06, Algorithm 16.4]. We do not go into the details here, but we provide as much information as is necessary to implement the method. The algorithm works as follows:

Algorithm 10.9. To shorten formulas we set

$$\underline{G} := \begin{bmatrix}\underline{I} & -\underline{I}\\ -\underline{I} & \underline{I}\end{bmatrix}, \quad \underline{C} := \begin{bmatrix}\underline{I} & \underline{0}\\ \underline{0} & \underline{I}\\ \underline{V} & -\underline{V}\end{bmatrix}, \quad \underline{\Lambda} := \operatorname{diag}(\underline{\lambda}), \quad \underline{Y} := \operatorname{diag}(\underline{y})$$

for $\underline{\lambda}, \underline{y} \in \mathbb{R}^{3n}$.

0. Choose initial points $\underline{x}_0^+ := \underline{\tilde{x}}_0^+$, $\underline{x}_0^- := \underline{\tilde{x}}_0^-$, $\underline{y}_0 := [1, \ldots, 1]^{\mathrm{T}} \in \mathbb{R}^{3n}$ (slack variables), and $\underline{\lambda}_0 \in \mathbb{R}^{3n}$ (Lagrange multipliers) with

$$[\underline{\lambda}_0]_j := \begin{cases} [-(\underline{\tilde{x}} + s\underline{p}) + s\alpha\underline{w}]_j, & j \in \{1, \ldots, n\},\\ [\underline{\tilde{x}} + s\underline{p} + s\alpha\underline{w}]_{j-n}, & j \in \{n+1, \ldots, 2n\},\\ 1, & j \in \{2n+1, \ldots, 3n\}. \end{cases}$$

These values are obtained from the starting point heuristic given in [NW06, end of Section 6.6].

1. Iterate for $k = 0, 1, 2, \ldots$:

2. Solve

$$\begin{bmatrix}\underline{G} & \underline{0} & -\underline{C}^{\mathrm{T}}\\ \underline{C} & -\underline{I} & \underline{0}\\ \underline{0} & \underline{\Lambda}_k & \underline{Y}_k\end{bmatrix}\begin{bmatrix}\begin{bmatrix}\hat{\underline{x}}^+\\ \hat{\underline{x}}^-\end{bmatrix}\\ \hat{\underline{y}}\\ \hat{\underline{\lambda}}\end{bmatrix} = \begin{bmatrix}-\underline{G}\begin{bmatrix}\underline{x}_k^+\\ \underline{x}_k^-\end{bmatrix} + \underline{C}^{\mathrm{T}}\underline{\lambda}_k - \begin{bmatrix}-(\underline{\tilde{x}} + s\underline{p}) + s\alpha\underline{w}\\ \underline{\tilde{x}} + s\underline{p} + s\alpha\underline{w}\end{bmatrix}\\ -C\begin{bmatrix}\underline{x}_k^+\\ \underline{x}_k^-\end{bmatrix} + \underline{y}_k\\ -\underline{\Lambda}_k\underline{Y}_k\underline{e}\end{bmatrix}$$

for $\hat{\underline{x}}^+, \hat{\underline{x}}^-, \hat{\underline{y}}, \hat{\underline{\lambda}}$, where $\underline{e} := [1, \dots, 1]^{\mathrm{T}} \in \mathbb{R}^{3n}$.

3. Calculate

$$\begin{aligned}
\mu &:= \frac{1}{3n} \underline{y}_k^{\mathrm{T}} \underline{\lambda}_k, \\
\hat{a} &:= \max\bigl\{a \in (0,1] : \underline{y}_k + a\hat{\underline{y}} \geq \underline{0},\ \underline{\lambda}_k + a\hat{\underline{\lambda}} \geq \underline{0}\bigr\}, \\
\hat{\mu} &:= \frac{1}{3n} (\underline{y}_k + \hat{a}\hat{\underline{y}})^{\mathrm{T}} (\underline{\lambda}_k + a\hat{\underline{\lambda}}), \\
\sigma &:= \left(\frac{\hat{\mu}}{\mu}\right)^3.
\end{aligned}$$

4. Solve

$$\begin{bmatrix} \underline{G} & \underline{0} & -\underline{C}^{\mathrm{T}} \\ \underline{C} & -\underline{I} & \underline{0} \\ \underline{0} & \underline{\Lambda}_k & \underline{Y}_k \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \underline{x}^+ \\ \underline{x}^- \end{bmatrix} \\ \underline{y} \\ \underline{\lambda} \end{bmatrix} = \begin{bmatrix} -\underline{G} \begin{bmatrix} \underline{x}_k^+ \\ \underline{x}_k^- \end{bmatrix} + \underline{C}^{\mathrm{T}} \underline{\lambda}_k - \begin{bmatrix} -(\tilde{\underline{x}} + s\underline{p}) + s\alpha\underline{w} \\ \tilde{\underline{x}} + s\underline{p} + s\alpha\underline{w} \end{bmatrix} \\ -\underline{C} \begin{bmatrix} \underline{x}_k^+ \\ \underline{x}_k^- \end{bmatrix} + \underline{y}_k \\ -\underline{\Lambda}_k \underline{Y}_k \underline{e} - \hat{\underline{\Lambda}} \hat{\underline{Y}} \underline{e} + \sigma\mu\underline{e} \end{bmatrix}$$

for $\underline{x}^+, \underline{x}^-, \underline{y}, \underline{\lambda}$, where $\underline{e} := [1, \dots, 1]^{\mathrm{T}} \in \mathbb{R}^{3n}$.

5. Calculate

$$\begin{aligned}
a_1 &:= \max\bigl\{a \in (0,1] : \underline{y}_k + 2a\underline{y} \geq \underline{0}\bigr\}, \\
a_2 &:= \max\bigl\{a \in (0,1] : \underline{\lambda}_k + 2a\underline{\lambda} \geq \underline{0}\bigr\}.
\end{aligned}$$

6. Set

$$\begin{bmatrix} \begin{bmatrix} \underline{x}_{k+1}^+ \\ \underline{x}_{k+1}^- \end{bmatrix} \\ \underline{y}_{k+1} \\ \underline{\lambda}_{k+1} \end{bmatrix} := \begin{bmatrix} \begin{bmatrix} \underline{x}_k^+ \\ \underline{x}_k^- \end{bmatrix} \\ \underline{y}_k \\ \underline{\lambda}_k \end{bmatrix} + \min\{a_1, a_2\} \begin{bmatrix} \begin{bmatrix} \underline{x}^+ \\ \underline{x}^- \end{bmatrix} \\ \underline{y} \\ \underline{\lambda} \end{bmatrix}.$$

7. Stop iteration if $\|\underline{x}_{k+1}^+ - \underline{x}_k^+\|_2^2 + \|\underline{x}_{k+1}^- - \underline{x}_k^-\|_2^2$ is small enough.

In each iteration of the algorithm we have to solve two $(8n) \times (8n)$ systems. Due to the simple structure of the matrices $\underline{G}$ and $\underline{C}$ we can reduce both to $n \times n$ systems. Indeed, the solution of the system

$$\begin{bmatrix} \underline{I} & -\underline{I} & \underline{0} & \underline{0} & \underline{0} & -\underline{I} & \underline{0} & -\underline{V}^{\mathrm{T}} \\ -\underline{I} & \underline{I} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & -\underline{I} & \underline{V}^{\mathrm{T}} \\ \underline{I} & \underline{0} & -\underline{I} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{I} & \underline{0} & -\underline{I} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{V} & -\underline{V} & \underline{0} & \underline{0} & -\underline{I} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{\Lambda}_k^1 & \underline{0} & \underline{0} & \underline{Y}_k^1 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{\Lambda}_k^2 & \underline{0} & \underline{0} & \underline{Y}_k^2 & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{\Lambda}_k^3 & \underline{0} & \underline{0} & \underline{Y}_k^3 \end{bmatrix} \begin{bmatrix} \underline{x}^+ \\ \underline{x}^- \\ \underline{y}^1 \\ \underline{y}^2 \\ \underline{y}^3 \\ \underline{\lambda}^1 \\ \underline{\lambda}^2 \\ \underline{\lambda}^3 \end{bmatrix} = \begin{bmatrix} \underline{\rho}^+ \\ \underline{\rho}^- \\ \underline{r}^1 \\ \underline{r}^2 \\ \underline{r}^3 \\ \underline{q}^1 \\ \underline{q}^2 \\ \underline{q}^3 \end{bmatrix}$$

(all occurring vectors are in $\mathbb{R}^n$) is given by solving

$$\begin{aligned}\Big(\underline{V}^{\mathrm{T}} - \big(\underline{I} + (\underline{\Lambda}_k^1)^{-1}\underline{Y}_k^1 + (\underline{\Lambda}_k^2)^{-1}\underline{Y}_k^2\big)\underline{V}^{\mathrm{T}}\big(\underline{I} + \tfrac{1}{n}(\underline{\Lambda}_k^3)^{-1}\underline{Y}_k^3\big)\Big)\underline{\lambda}_3 \\ = \underline{r}^1 - \underline{\rho}^+ - \underline{r}^2 + (\underline{\Lambda}_k^1)^{-1}\underline{q}^1 - (\underline{\Lambda}_k^2)^{-1}\underline{q}^2 + \big(\underline{I} + (\underline{\Lambda}_k^1)^{-1}\underline{Y}_k^1\big)\big(\underline{\rho}^+ + \underline{\rho}^-\big) \\ - \big(\underline{I} + (\underline{\Lambda}_k^1)^{-1}\underline{Y}_k^1 + (\underline{\Lambda}_k^2)^{-1}\underline{Y}_k^2\big)\big(\tfrac{1}{n}\underline{V}^{\mathrm{T}}\underline{r}^3 + \underline{\rho}^- + \tfrac{1}{n}\underline{V}^{\mathrm{T}}(\underline{\Lambda}_k^3)^{-1}\underline{q}^3\big)\end{aligned}$$

for $\underline{\lambda}_3$ and gradually calculating

$$\begin{aligned}\underline{\lambda}^2 &= \tfrac{1}{n}\underline{V}^{\mathrm{T}}\big(-\underline{r}^3 - (\underline{\Lambda}_k^3)^{-1}\underline{q}^3 + (\underline{\Lambda}_k^3)^{-1}\underline{Y}_k^3\underline{\lambda}^3 + n\underline{\lambda}^3\big) - \underline{\rho}^-, \\ \underline{\lambda}^1 &= -\underline{\rho}^+ - \underline{\rho}^- - \underline{\lambda}^2, \\ \underline{y}^3 &= (\underline{\Lambda}_k^3)^{-1}\big(\underline{q}^3 - \underline{Y}_k^3\underline{\lambda}^3\big), \\ \underline{y}^2 &= (\underline{\Lambda}_k^2)^{-1}\big(\underline{q}^2 - \underline{Y}_k^2\underline{\lambda}^2\big), \\ \underline{y}^1 &= (\underline{\Lambda}_k^1)^{-1}\big(\underline{q}^1 - \underline{Y}_k^1\underline{\lambda}^1\big), \\ \underline{x}^- &= \underline{r}^2 + \underline{y}^2, \\ \underline{x}^+ &= \underline{\rho}^+ + \underline{x}^- + \underline{\lambda}^1 + \underline{V}^{\mathrm{T}}\underline{\lambda}^3.\end{aligned}$$

This result can be obtained by Gaussian elimination. Note that the matrices $\underline{\Lambda}_k^i$ and $\underline{Y}_k^i$ are diagonal. Thus matrix multiplication and inversion is not expensive.

10.3.3. Problems related with the algorithm

In this subsection we comment on some problems and modifications of Algorithm 10.7.

A problem we did not mention so far is the situation that the algorithm could produce iterates $\underline{x}_k$ with $[\underline{A}\underline{x}_k]_i = 0$ for some i where $z_i > 0$. In this case the fitting functional $S^{\underline{z}}(\underline{x}_k)$ would be infinite and not differentiable. In numerical experiments we never observed this problem. Having a look at the algorithm we see that the choice of $\underline{x}_0$ and the line search prevent such situations.

A second problem concerns the stopping criterion. In Section 10.2 we derived an optimality criterion which can be checked numerically (cf. Proposition 10.6). But in Algorithm 10.7 we stop if the iterates do not change. The reason is that numerical experiments have shown that there is no convergence to exact satisfaction of the optimality criterion. Thus, we cannot give bounds or the bounds have to depend on n and on the underlying exact solution. A kind of normalization could solve this scaling problem. Nonetheless for all numerical examples provided below we check whether the optimality criterion seems to be approximately satisfied.

In each iteration of Algorithm 10.7 we have to solve the minimization problem (10.8). Algorithm 10.9 provides us with a minimizer but is expensive with respect to computation time. In numerical experiments we observed that applying the cut-off operation (10.10) to the result of the soft-thresholding operation (10.9) gives almost always the same result as Algorithm 10.9. In the rare situation that the soft-thresholding yields the Haar coefficients of a nonnegative function this observation is easy to comprehend. But in general we could not prove any relation between the minimizers of (10.8) and the outcome of the soft-thresholding and the cut-off operation. Only in very few cases the composition of the two simple operations fails to give a minimizer. In such a case

we could look at the resulting (incorrect) iterate $\underline{x}_k$ as a new starting point for Algorithm 10.7. Consequently, replacing Algorithm 10.9 by soft-thresholding and cut-off saves computation time while providing similar results.

A fourth and last point concerns the implementation. The matrix $\underline{A}$ is sparse as can be seen in Table 10.1. Thus, matrix vector multiplication could be accelerated by exploiting this fact. But for the sake of simplicity our implementation does not take advantage of sparsity.

n	non-zero elements [%]
16	39.84
32	26.17
64	16.21
128	10.35
256	7.06
512	5.22
1024	4.22
2048	3.67
4096	3.37

Table 10.1.: Share of non-zero elements in the matrix $\underline{A}$ for different discretization levels n and fixed kernel width $b = \frac{1}{8}$.

10.4. Numerical results

We are now ready to perform some numerical experiments. The aim is to compare the results obtained by the standard Tikhonov-type approach (adapted to Gaussian noise)

$$\frac{1}{2}\|\underline{Ax} - \underline{z}\|_2^2 + \alpha\underline{\Omega}(\underline{x}) \to \min_{\underline{x}\in\mathbb{R}^n}$$

with the results from the Poisson noise adapted version described in this part of the thesis. Note that the Gaussian noise adapted Tikhonov-type method with nonnegativity and sparsity constraints is also given by Algorithm 10.7 but the gradient in step 2 has to be replaced by $\underline{A}^{\mathrm{T}}(\underline{Ax}_k - \underline{z})$.

The synthetic data is produced as follows: Given an exact solution $\underline{x}^\dagger$ first calculate $\underline{y}^\dagger := \underline{Ax}^\dagger$. In case of Gaussian distributed data generate a vector $\underline{z}$ such that z_i follows a Gaussian distribution with mean $y_i^\dagger$ and standard deviation $\sigma > 0$. In case of Poisson distributed data generate a vector $\underline{z}$ such that

$$\frac{\gamma}{\max\{y_1^\dagger,\dots,y_n^\dagger\}} z_i \sim \mathrm{Poisson}\left(\frac{\gamma}{\max\{y_1^\dagger,\dots,y_n^\dagger\}} y_i^\dagger\right),$$

where $\gamma > 0$ controls the noise level (in the language of imaging: γ is the average number of photons impinging on the pixel with highest light intensity).

Note, that using the same discretization level and the same discretized operator for direct and inverse calculations is an inverse crime. But in our case it is a way to eliminate

the influence of discretization on the regularization error. We are solely interested in the influence of the fitting functional.

Due to the same reason the regularization parameter α will be chosen by hand such that $\|\underline{x}_\alpha^z - \underline{x}^\dagger\|_2$ attains its minimum. Of course, for real world problems we do not know the exact solution $\underline{x}^\dagger$, but in our experiments we have this information at hand. The interested reader finds an attempt to develop Poisson noise adapted parameter choices in [ZBZB09].

For all experiments we use the discretization level $\bar{l} = 8$, that is, $n = 512$, and the kernel width $b = \frac{1}{8}$. The weights $w_1, \dots, w_n$ in the definition of $\underline{\Omega}$ shall be one. As bounds for the step length in Algorithm 10.8 we set $s_{\min} := 10^{-10}$ and $s_{\max} := 10^5$.

All computations were carried out using the software *MATLAB* by *The MathWorks, Inc.*, version R2011a.

10.4.1. Experiment 1: astronomical imaging

Our first example can be seen as a one-dimensional version of a typical astronomic image: a bright star surrounded by smaller and less bright objects. The exact solution $x^\dagger$ is depicted in Figure 10.1. Here as well as in the sequel, variables without underline denote the step function corresponding to the same variable with underline (a vector in $\mathbb{R}^n$).

We start with relatively high Poisson noise ($\gamma = 100$). Figure 10.2 shows the exact data $Ax^\dagger$ and the noisy data z.

The dependence of the regularization error $\|\underline{x}_\alpha^z - \underline{x}^\dagger\|_2$ on the parameter α is depicted in Figure 10.3 for both the Poisson noise adapted Tikhonov functional and the standard (Gaussian noise adapted) Tikhonov functional. The regularization error is constant for large α because the corresponding regularized solutions all represent the same constant function. We do not have an explanation for the strong oscillations near the constant region. But since those oscillations do not effect the minima of the curves we are not going to investigate this phenomenon in more detail. Perhaps, it results from a combination of sparsity constraints and incomplete minimization.

The regularized solutions x_α^z obtained from the Poisson noise adapted and the standard Tikhonov method are shown in Figure 10.4 and Figure 10.5, respectively. Remember, that we choose the optimal regularization parameter α for comparing the results of the two Tikhonov-type methods. This eliminates the influence of (imperfect) parameter choice rules.

The regularization error $\|\underline{x}_\alpha^z - \underline{x}^\dagger\|_2$ is 0.63227 in the Poisson case (algorithm stopped after 609 iterations) and 0.83155 in the standard case (algorithm stopped after 58 iterations). Comparing the two graphs we see the following:

- In the Poisson case the region between $t = 0.3$ and $t = 0.5$ is reconstructed quite exact, but in the Gaussian case this region shows oscillations.
- The Poisson noise adapted method pays attention to the three smaller objects on the right-hand side such that they can be identified from the regularized solution. In contrast, the standard method blurs them too much.

If we reduce the noise by setting $\gamma = 1000$ (that is, ten times more photons as before), we obtain the regularized solutions depicted in Figure 10.7 and Figure 10.8.

Exact and noisy data are shown in Figure 10.6. Now the areas with low function values are reconstructed well by both methods. But, contrary to the Poisson case, the standard Tikhonov method produces strong oscillations between $t = 0.3$ and $t = 0.5$. The regularization errors are 0.2224 in the Poisson case (algorithm stopped after 3574 iterations) and 0.5846 for the standard method (algorithm stopped after 1036 iterations).

The experiments carried out so far indicate that the Poisson noise adapted method yields more accurate results than the standard method in case of Poisson distributed data. Of course we have to check whether the roles change if the data follows a Gaussian distribution. Therefore we generate Gaussian distributed data with standard deviation $\sigma = 0.0006$ and cut off negative values (see Figure 10.9).

The regularized solution obtained from the Poisson noise adapted method (see Figure 10.10) shows oscillations outside the interval $(0.3, 0.5)$, whereas the standard Tikhonov method (see Figure 10.11) yields a quite exact reconstruction. In the Poisson case the regularization error is 0.7864 (algorithm stopped after 133 iterations) and for the standard method it is 0.6123 (algorithm stopped after 178 iterations).

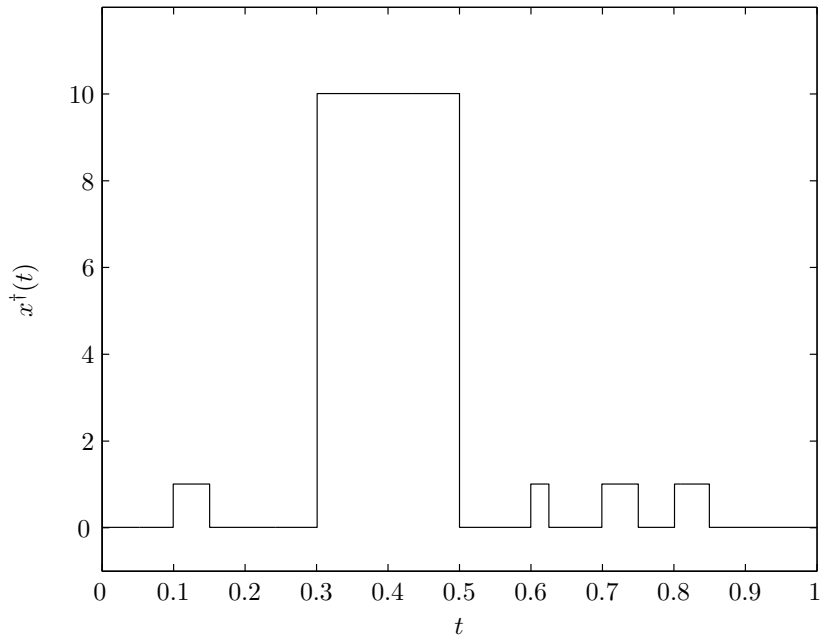

Figure 10.1.: The exact solution $x^\dagger$ represents a bright star surrounded by smaller objects.

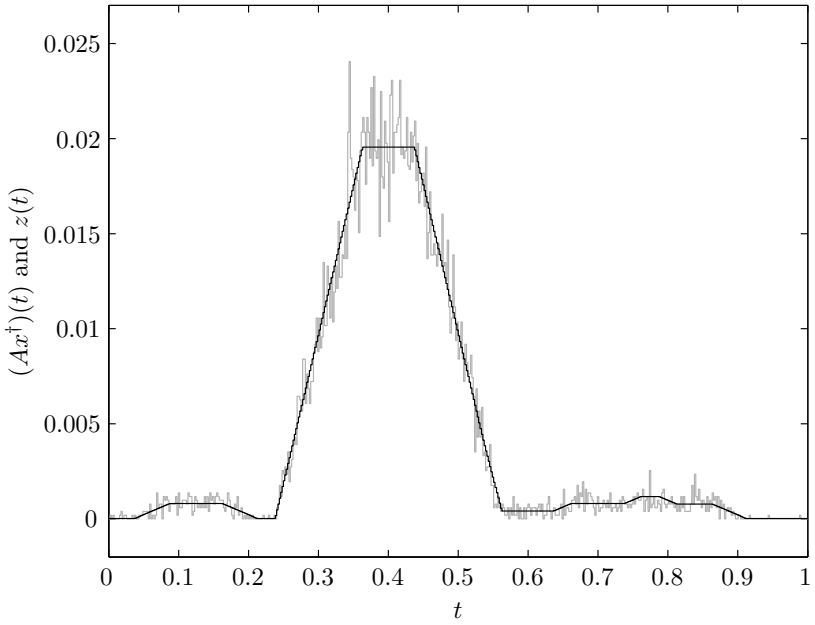

Figure 10.2.: Exact data $Ax^\dagger$ (black line) and noisy data z (gray line) corrupted by Poisson noise with $\gamma = 100$.

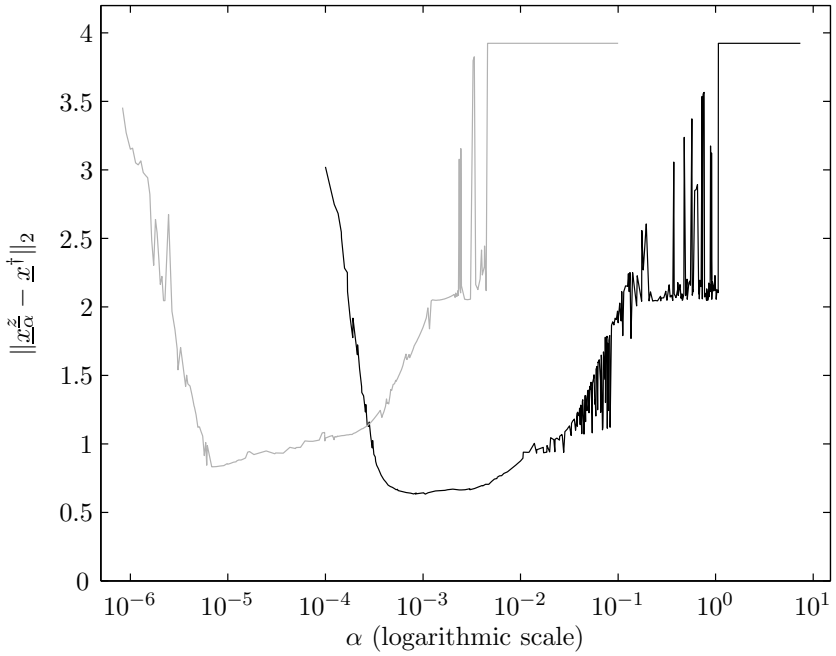

Figure 10.3.: Regularization error $\|\underline{x}_\alpha^z - \underline{x}^\dagger\|_2$ in dependence of α for the Poisson noise adapted (black line) and the standard Tikhonov method (gray line).

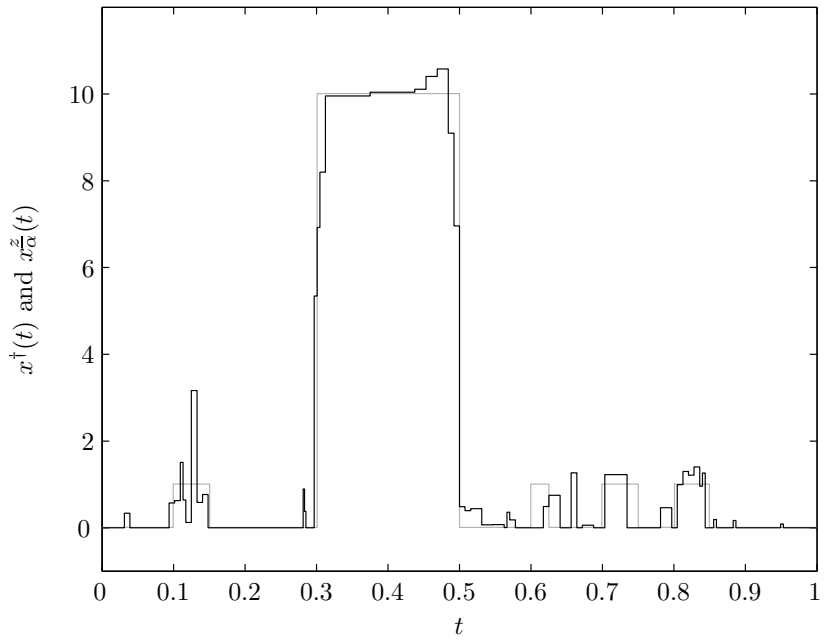

Figure 10.4.: Regularized solution x_α^z (black line) obtained from the Poisson noise adapted Tikhonov functional ($\alpha = 8.41 \cdot 10^{-4}$). The exact solution $x^\dagger$ is depicted in gray.

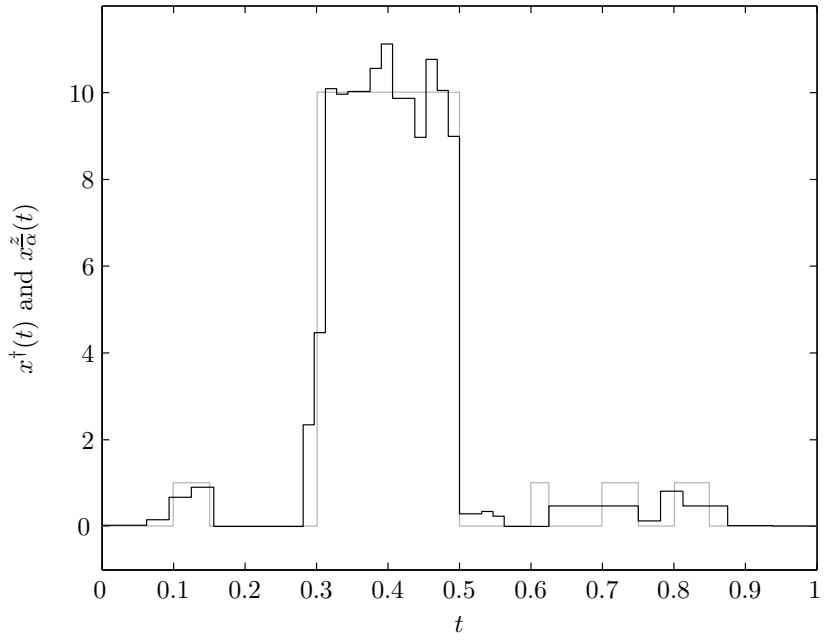

Figure 10.5.: Regularized solution x_α^z (black line) obtained from the standard Tikhonov functional ($\alpha = 6.82 \cdot 10^{-6}$). The exact solution $x^\dagger$ is depicted in gray.

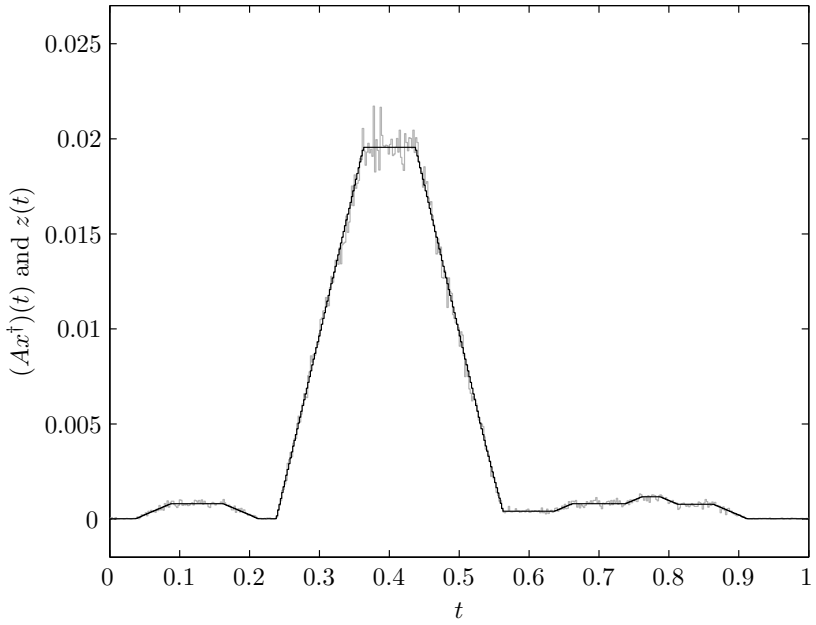

Figure 10.6.: Exact data $Ax^\dagger$ (black line) and noisy data z (gray line) corrupted by Poisson noise with $\gamma = 1000$.

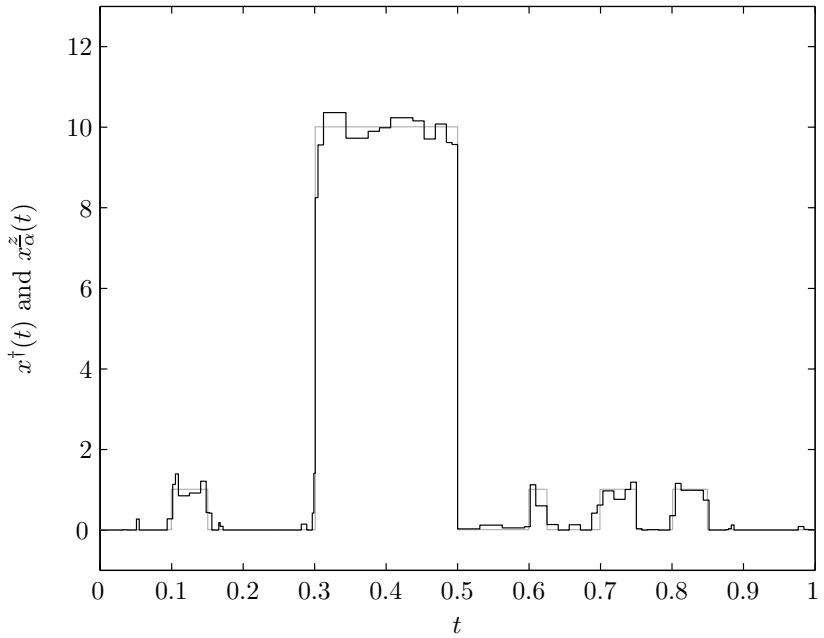

Figure 10.7.: Regularized solution x_α^z (black line) obtained from the Poisson noise adapted Tikhonov functional ($\alpha = 3.68 \cdot 10^{-4}$). The exact solution $x^\dagger$ is depicted in gray.

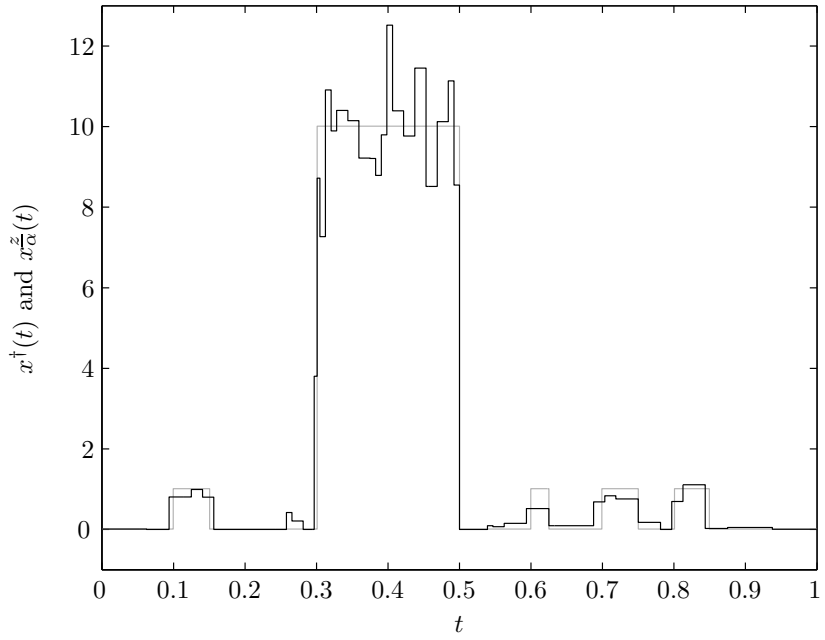

Figure 10.8.: Regularized solution x_α^z (black line) obtained from the standard Tikhonov functional ($\alpha = 8.30 \cdot 10^{-7}$). The exact solution $x^\dagger$ is depicted in gray.

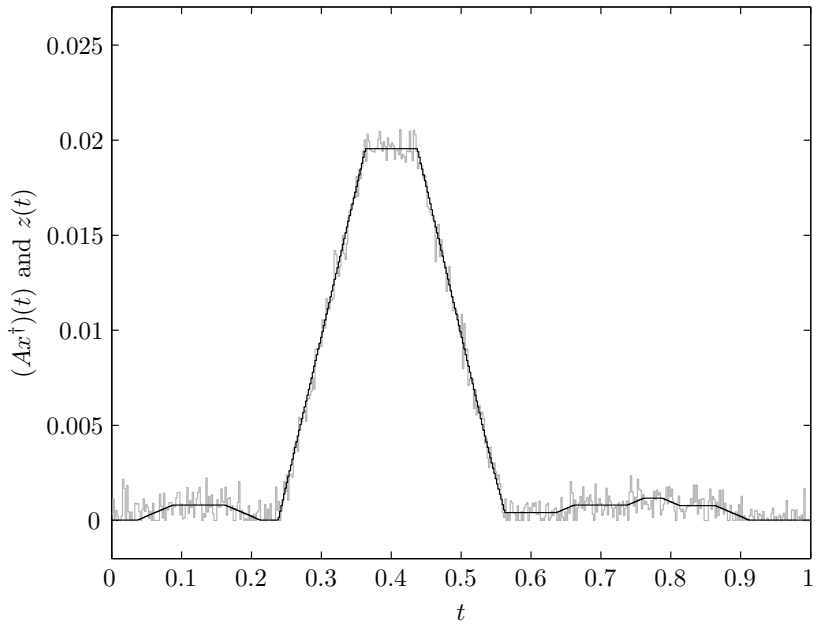

Figure 10.9.: Exact data $Ax^\dagger$ (black line) and noisy data z (gray line) corrupted by Gaussian noise with $\sigma = 0.0006$.

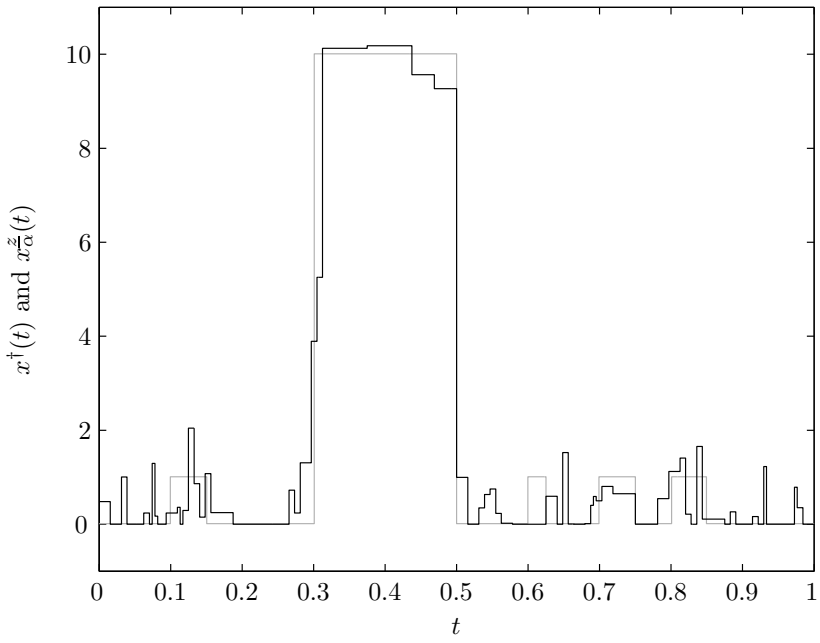

Figure 10.10.: Regularized solution x_α^z (black line) obtained from the Poisson noise adapted Tikhonov functional ($\alpha = 1.14 \cdot 10^{-3}$). The exact solution $x^\dagger$ is depicted in gray.

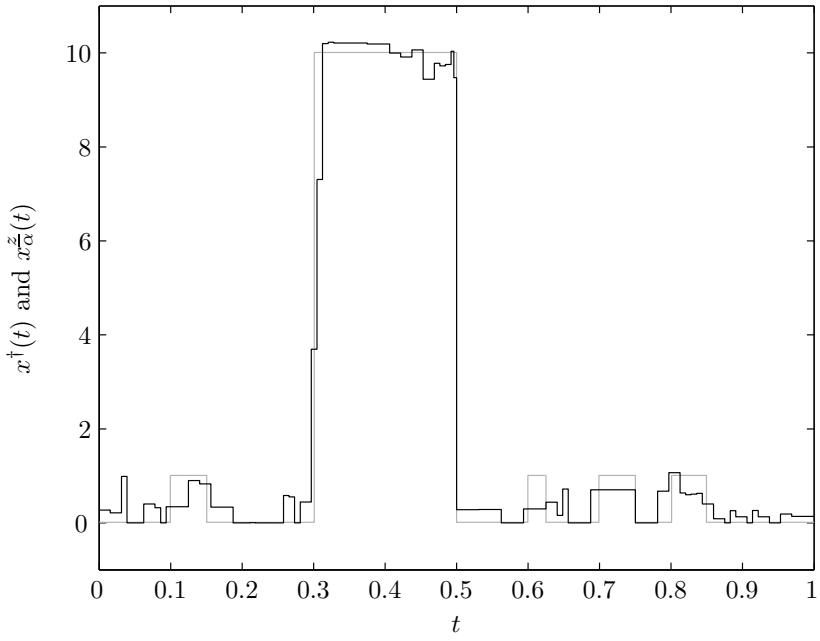

Figure 10.11.: Regularized solution x_α^z (black line) obtained from the standard Tikhonov functional ($\alpha = 1.29 \cdot 10^{-6}$). The exact solution $x^\dagger$ is depicted in gray.

10.4.2. Experiment 2: high count rates

In our second experiment we consider the typical situation in imaging with high photon counts. The characteristic property is that there is a strong background radiation and only relatively small perturbations of this high light intensity. But from these small perturbations the human eye forms an image, where the smallest (but nonetheless high) photon count is regarded as black. A concrete example is provided in Figure 10.12. All function values lie above 48.

From the exact data $\underline{Ax}^\dagger$ we generate Poisson distributed data $\underline{z}$ with photon level $\gamma = 500000$ (see Figure 10.13). Due to the high but similar values of $\underline{x}^\dagger$ the standard deviation of a Poisson distributed random variable with mean $c[\underline{Ax}^\dagger]_i$, $c > 0$, is very insensitive with respect to i. In addition a Poisson distribution with parameter λ approximates a Gaussian distribution with mean λ and variance λ if λ goes to infinity. Thus, using Gaussian distributed data instead of Poisson distributed data would lead to a similar appearance of the noisy data. Numerical simulations have shown that $\sigma = 0.00015$ would yield the same noise level as $\gamma = 500000$ in the present example.

As a consequence of these considerations we expect that the Poisson noise adapted and the standard Tikhonov method yield similar results. The regularized solutions depicted in Figure 10.14 and Figure 10.15 verify this conjecture. The regularization error is 0.1502 in the Poisson case (algorithm stopped after 81 iterations) and 0.1512 in the standard case (algorithm stopped after 70 iterations).

10.4.3. Experiment 3: organic structures

One application for imaging with low photon counts is *confocal laser scanning microscopy* (CLSM), see [Wil]. This technique is used for observing processes in living tissue. Thus, in our third and last experiment we have a look at an exact solution $x^\dagger$ which represents the typical features of organic structures: smooth and often also periodic appearance. The exact solution $x^\dagger$ given in Figure 10.16 is not sparse with respect to the Haar system. Thus, we cannot expect accurate results.

We perturb the exact data $\underline{Ax}^\dagger$ by Poisson noise with photon level $\gamma = 500$ (see Figure 10.17).

The regularized solutions are given in Figure 10.18 and Figure 10.19. As in the first experiment we see that the solution obtained from the standard Tikhonov method is too smooth in regions of small function values, whereas the Poisson noise adapted method yields quite good results. The regularization error is 0.04284 in the Poisson case (algorithm stopped after 65 iterations) and 0.04286 in the standard case (algorithm stopped after 23 iterations).

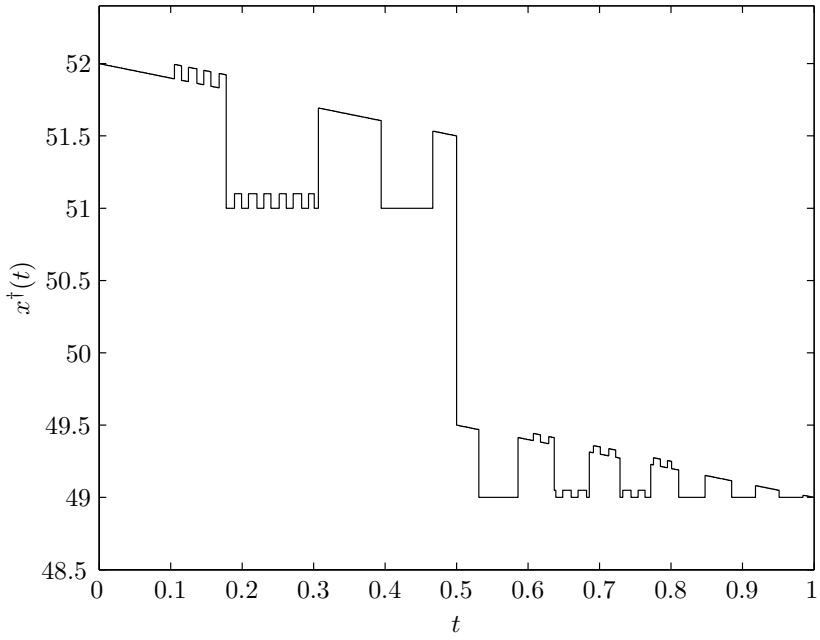

Figure 10.12.: The exact solution $x^\dagger$ has only a small range but high function values (the vertical axis does not start at zero).

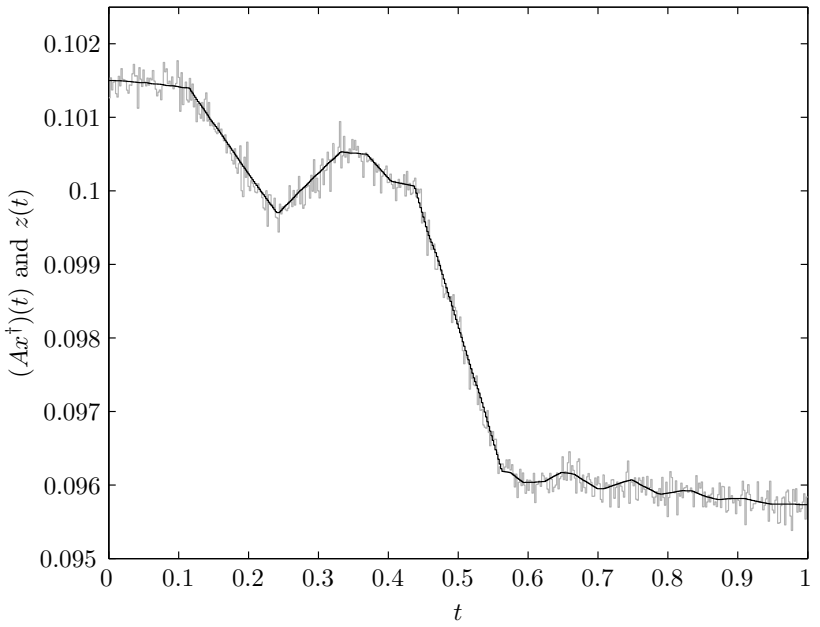

Figure 10.13.: Exact data $Ax^\dagger$ (black line) and noisy data z (gray line) corrupted by Poisson noise with $\gamma = 500000$ (the vertical axis does not start at zero).

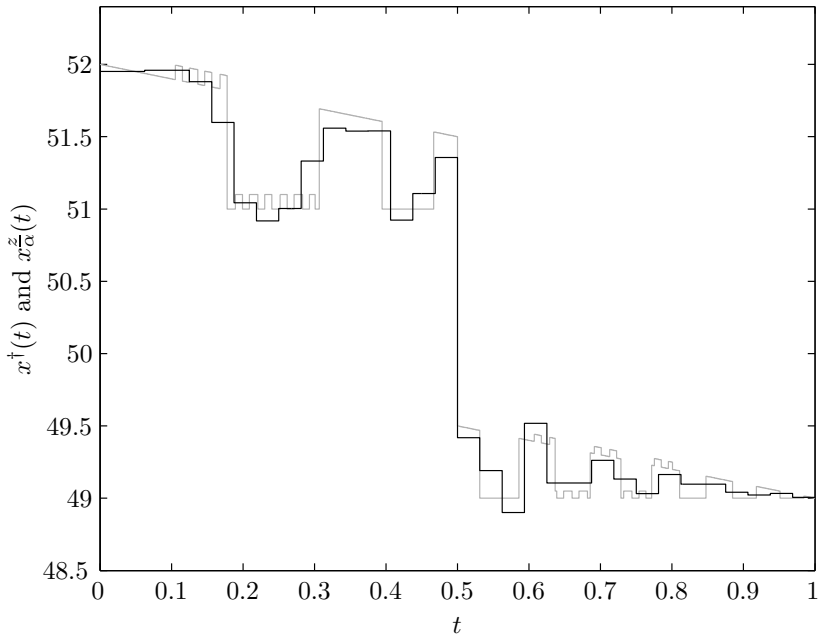

Figure 10.14.: Regularized solution x_α^z (black line) obtained from the Poisson noise adapted Tikhonov functional ($\alpha = 8.40 \cdot 10^{-6}$). The exact solution $x^\dagger$ is depicted in gray. Note, that the vertical axis does not start at zero.

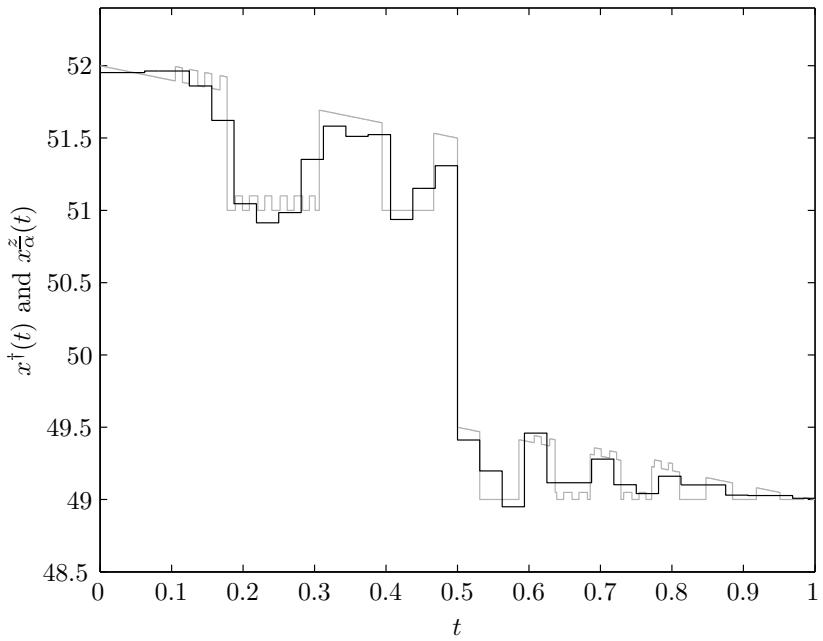

Figure 10.15.: Regularized solution x_α^z (black line) obtained from the standard Tikhonov functional ($\alpha = 9.00 \cdot 10^{-7}$). The exact solution $x^\dagger$ is depicted in gray. Note, that the vertical axis does not start at zero.

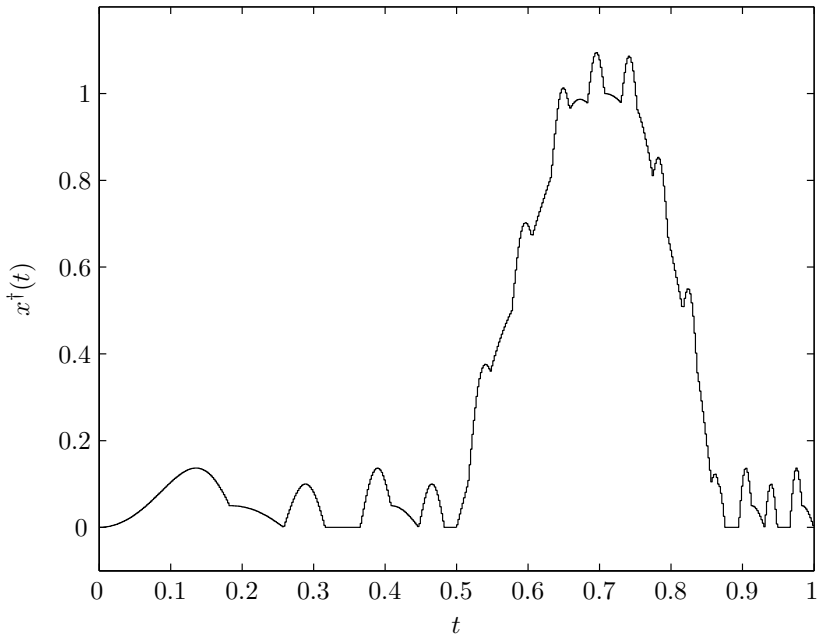

Figure 10.16.: The exact solution $x^\dagger$ is smooth and to some extent also periodic.

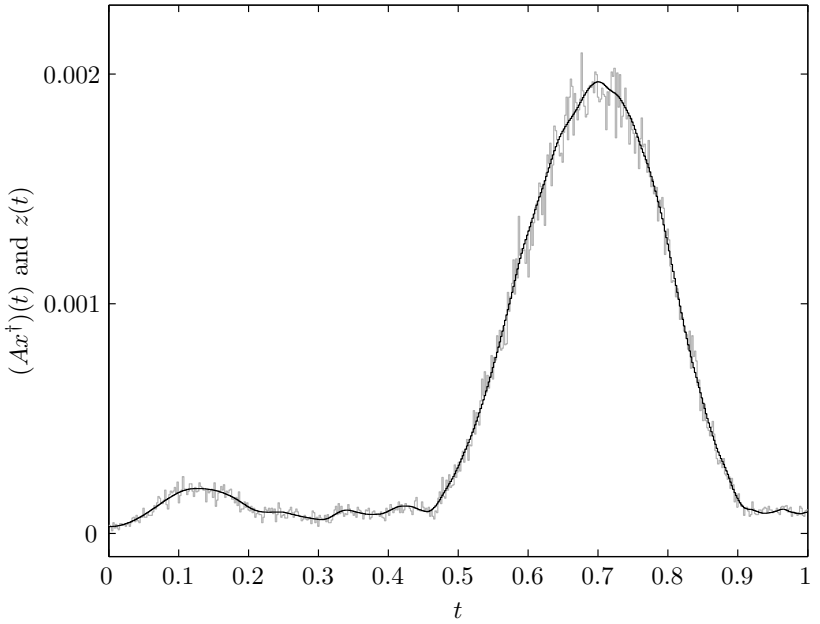

Figure 10.17.: Exact data $Ax^\dagger$ (black line) and noisy data z (gray line) corrupted by Poisson noise with $\gamma = 500$.

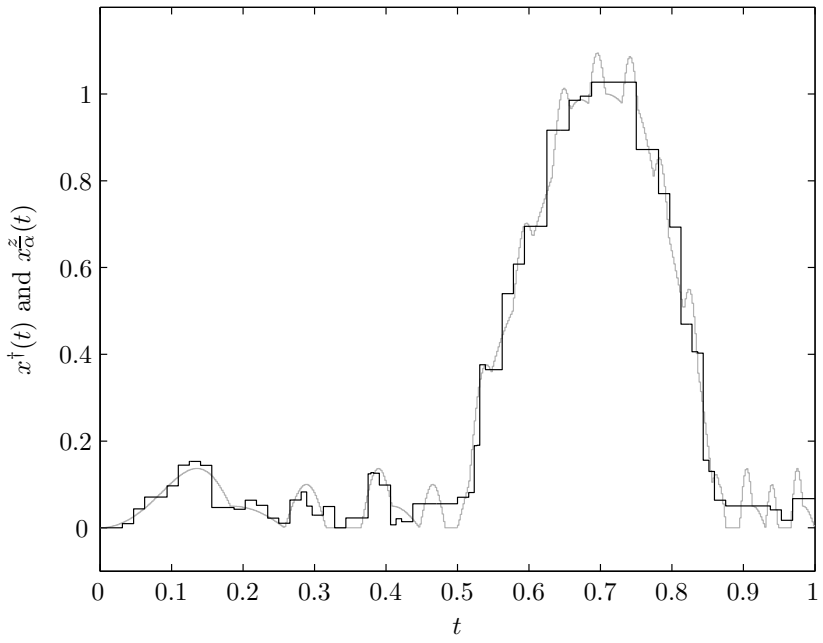

Figure 10.18.: Regularized solution x_α^z (black line) obtained from the Poisson noise adapted Tikhonov functional ($\alpha = 5.75 \cdot 10^{-4}$). The exact solution $x^\dagger$ is depicted in gray.

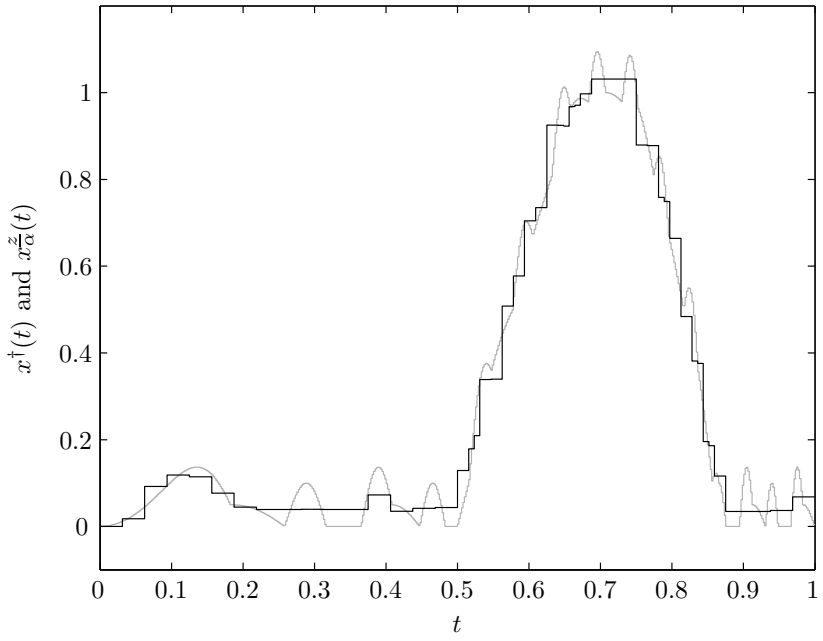

Figure 10.19.: Regularized solution x_α^z (black line) obtained from the standard Tikhonov functional ($\alpha = 4.54 \cdot 10^{-7}$). The exact solution $x^\dagger$ is depicted in gray.

10.5. Conclusions

The numerical experiments presented above have shown that the regularized solutions obtained from the Poisson noise adapted and from the standard (Gaussian noise adapted) Tikhonov method are different. We made the following observations:

- In case of Poisson distributed data the standard Tikhonov method yields solutions which are too smooth in regions of small function values and which oscillate too much in regions of high function values.
- In case of Gaussian distributed data regularized solutions obtained from the Poisson noise adapted method oscillate if the function values are small.

We try to explain these observations. The squared norm as a fitting functional penalizes deviations of $\underline{Ax}$ from the data $\underline{z}$ for all $i = 1, \ldots, n$ in the same way, regardless of the value of z_i. But if the data follows a Poisson distribution, then the deviation of $[\underline{Ax}^\dagger]_i$ from z_i is small if z_i is small and it is very high if z_i has a large value. Thus, in regions of small function values the penalization is too weak (causing overregularization) and in regions of high function values the penalization is too strong (resulting in underregularization).

On the other hand, the Kullback–Leibler distance used as the fitting functional in the Poisson noise adapted method varies the strength of penalization depending on the values z_i. This fact is advantageous in case of Poisson distributed data, since deviations of $[\underline{Ax}^\dagger]_i$ from z_i are small if z_i is small and they are high if z_i has a larger value. But for Gaussian distributed data, which shows the same variance for all $i = 1, \ldots, n$, similar effects as described above for the standard method with Poisson distributed data occur. In regions of small function values the penalization is too strong (resulting in underregularization) and in regions of large function values the penalization is too weak (causing overregularization).

We see that in case of Poisson distributed data the Poisson noise adapted Tikhonov method is superior to the standard method with respect to reconstruction quality. But the experiments have also shown that the Poisson noise adapted method requires more iterations and, thus, more computation time. Another drawback of this method is the lack of parameter choice rules. The discrepancy principle known for the standard Tikhonov method is not applicable because in case of Poisson distributed data we have now means of noise level. An attempt to develop Poisson noise adapted parameter choices is given in [ZBZB09].

Part III.

Smoothness assumptions

11. Introduction

The third and last part of this thesis on the one hand supplements Part I by discussing the variational inequality (4.3), which allows to derive convergence rates for general Tikhonov-type regularization methods. On the other hand the findings we present below are of independent interest because they provide new and extended insights into the interplay of different kinds of smoothness assumptions frequently occurring in the context of regularization methods.

To make the material accessible to readers who did not work through Part I in detail we keep the present part as self-contained as possible; only few references are made to results from Part I. In particular we restrict our attention to Banach and sometimes also to Hilbert spaces and corresponding Tikhonov-type regularization approaches, even though some of the results remain still valid in more general settings.

The rate of convergence of regularized solutions to an exact solution depends on the (abstract) smoothness of all involved quantities. Typically the operator of the underlying equation has to be differentiable, the spaces should be smooth (that is, they should have differentiable norms), and the exact solution has to satisfy some abstract smoothness assumption with respect to the operator. This last type of smoothness is usually expressed in form of source conditions (see below).

If one of the three components (operator, spaces, exact solution) lacks smoothness, the other components have to compensate this lack. Thus, the obvious idea is to combine all required types of smoothness into one sufficient condition for deriving convergence rates. Since the aim of convergence rates theory is to provide upper bounds for rates on a whole class of exact solutions, such an 'all-inclusive' condition has to be independent of noisy data used for the regularization process. Otherwise the condition cannot be checked in advance. This restriction makes the construction very challenging.

In 2007 a sufficient condition for convergence rates combining all necessary smoothness assumptions has been suggested in [HKPS07]. The authors formulated a so called variational inequality which allows to prove convergence rates without any further assumption on the operator, the spaces, or the exact solution. Inequality (4.3) in Part I is a generalization of this original variational inequality. The development from the original to the very general form is sketched in Section 12.1.4.

Our aim is to bring the cross connections between variational inequalities and classical smoothness assumptions to light. Here smoothness of the involved spaces will play only a minor role. We concentrate on solution smoothness and provide also some results related to properties of the possibly nonlinear operator of the underlying equation.

The material of this part is split into two chapters. The first discusses smoothness in Banach spaces and contains the major results. In the second chapter on smoothness in Hilbert spaces we specialize and extend some of the results obtained for Banach spaces.

12. Smoothness in Banach spaces

As noted in the introductory chapter we restrict our attention to Banach spaces, though some of the results also hold in more general situations. The setting in the present chapter is the same as in Example 1.2. That is, X and Y are Banach spaces, $F : D(F) \subseteq X \to Y$ is a possibly nonlinear operator with domain $D(F)$, and we aim to solve the equation

$$F(x) = y^0, \quad x \in D(F), \tag{12.1}$$

with exact right-hand side $y^0 \in Y$. In practice y^0 is not known to us, instead we only have some noisy measurement $y^\delta \in Y$ at hand satisfying $\|y^\delta - y^0\| \le \delta$ with noise level $\delta > 0$. To overcome ill-posedness of F we regularize the solution process by minimizing the Tikhonov functional

$$T_\alpha^{y^\delta}(x) := \tfrac{1}{p}\|F(x) - y^\delta\|^p + \alpha\Omega(x) \tag{12.2}$$

over $x \in D(F)$, where $p \ge 1$, $\alpha > 0$, and $\Omega : X \to (-\infty, \infty]$ is convex.

The assumptions made in Part I can be summarized as follows (cf. Section 2.2):

- $D(F)$ is weakly sequentially closed.
- F is sequentially continuous with respect to the weak topologies on X and Y.
- The sublevel sets $M_\Omega(c) := \{x \in X : \Omega(x) \le c\}$ are weakly sequentially compact for all $c \in \mathbb{R}$.

Since we solely consider the sequential versions of weak closedness, weak continuity, and weak compactness, we drop the 'sequential' below.

In addition we assume the existence of solutions to (12.1) which lie in the essential domain $D(\Omega) := \{x \in X : \Omega(x) < \infty\}$ of Ω. Then Proposition 3.1 guarantees the existence of Ω-minimizing solutions. Throughout this chapter let $x^\dagger$ be one fixed Ω-minimizing solution with $\partial\Omega(x^\dagger) \ne \emptyset$.

12.1. Different smoothness concepts

In this section we collect different types of smoothness assumptions which can be used to derive upper bounds for the Bregman distance $B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big)$ with $\xi^\dagger \in \Omega(x^\dagger)$ and an a priori parameter choice $\delta \mapsto \alpha(\delta)$. As in Part I the element $x_\alpha^{y^\delta} \in \operatorname{argmin}_{x \in D(F)} T_\alpha^{y^\delta}(x)$ is a regularized solution to the noisy measurement y^δ with noise level δ.

12.1.1. Structure of nonlinearity

To control the nonlinear structure of the operator F one typically assumes that $x^\dagger$ is an interior point of $D(F)$ and that F is Gâteaux differentiable at $x^\dagger$. If $x^\dagger$ is a boundary point of $D(F)$ and $D(F)$ is convex or at least starlike with respect to $x^\dagger$ (see Definition 12.1), then alternative constructions are possible. In both cases one assumes that $\lim_{t\to+0} \frac{1}{t}\big\|F\big(x^\dagger + t(x-x^\dagger)\big) - F(x^\dagger)\big\|$ exists for all $x \in D(F)$ and that there is a bounded linear operator $F'[x^\dagger] : X \to Y$ such that

$$F'[x^\dagger](x-x^\dagger) = \lim_{t\to+0} \frac{1}{t}\big\|F\big(x^\dagger + t(x-x^\dagger)\big) - F(x^\dagger)\big\| \quad \text{for all } x \in D(F).$$

Definition 12.1. A set $M \subseteq X$ is called *starlike* with respect to $\bar{x} \in X$ if for each $x \in M$ there is some $t_0 > 0$ such that $\bar{x} + t(x-\bar{x}) \in M$ for all $t \in [0, t_0]$.

The reason for linearizing F is that classical assumptions on the smoothness of the exact solution $x^\dagger$ were designed for linear operators. To extend the applicability of these classical techniques to nonlinear operators the smoothness of $x^\dagger$ is expressed with respect to $F'[x^\dagger]$. But to obtain convergence rates this way additional assumptions on the connection between F and $F'[x^\dagger]$ are required.

Literature provides different kinds of such structural assumptions connecting F with its linearization. We do not go into the details here. In the sequel we only use the simplest form

$$\|F'[x^\dagger](x-x^\dagger)\| \le c\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in M$$

with $c \ge 0$. The set $M \subseteq D(F)$ has to be sufficiently large to contain the regularized solutions $x_{\alpha(\delta)}^{y^\delta}$ for all sufficiently small $\delta > 0$ with a given parameter choice $\delta \mapsto \alpha(\delta)$. More sophisticated formulations are given for instance in [BH10, formulas (3.8) and (3.9)].

12.1.2. Source conditions

The most common assumption on the smoothness of the exact solution $x^\dagger$ is a source condition with respect to Ω and $F'[x^\dagger]$ as formulated in the following definition (with $F'[x^\dagger]$ defined as in Subsection 12.1.1).

Definition 12.2. The exact solution $x^\dagger$ satisfies a *source condition* with respect to the stabilizing functional Ω and to the operator $F'[x^\dagger]$ if there are a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ and a source element $\eta^\dagger \in Y^*$ such that

$$\xi^\dagger = F'[x^\dagger]^*\eta^\dagger.$$

Source conditions are discussed for instance in [EHN96] in Hilbert spaces. For Banach space settings we refer to more recent literature, e.g., [BO04] and [SGG+09, Proposition 3.35].

In Hilbert spaces spectral theory allows to modify the operator in the source condition to weaken or strengthen the condition (see Subsection 13.1.1). For Banach space settings there are only two source conditions, the one given above and a stronger

one involving duality mappings and re-enacting the Hilbert space source condition $\xi^\dagger = F'[x^\dagger]^* F'[x^\dagger]^* \eta^\dagger$, where X and Y are identified with there duals X^* and Y^*. The stronger source condition for Banach spaces was introduced in [Hei09, Neu09, NHH+10].

The convergence rate obtained from the source condition $\xi^\dagger = F'[x^\dagger]^* \eta^\dagger$ depends on the structure of nonlinearity of F (see Subsection 12.1.1). For a linear operator $A := F$ with $D(F) = X$ and $F'[x^\dagger] = A$ we have the following result.

Proposition 12.3. *Let $A := F$ be bounded and linear. If there are $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ such that $\xi^\dagger = A^*\eta^\dagger$, then*

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}(\delta) \quad \text{if } \delta \to 0$$

for an appropriate a priori parameter choice $\delta \mapsto \alpha(\delta)$.

Proof. A proof is given in [SGG+09] (Theorem 3.42 in combination with Proposition 3.35 there). Alternatively the assertion follows from Theorem 4.11 of this thesis in combination with Proposition 12.28 below. □

12.1.3. Approximate source conditions

Source conditions as described in the previous subsection provide only a very imprecise classification of solution smoothness; either $x^\dagger$ satisfies a source condition or it does not. In Hilbert spaces this problem is compensated by a wide scale of different source conditions. But in Banach spaces other techniques for expressing solutions smoothness have to be used.

In [Hof06] the idea of approximate source conditions has been introduced and an extension to Banach spaces has been described in [Hei08b, HH09]. Instead of deciding whether a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ satisfies a source conditions, one measures how far away $\xi^\dagger$ is from satisfying a source condition. This measuring is realized as a so called *distance function*

$$d(r) := \inf\big\{\big\|\xi^\dagger - F'[x^\dagger]\eta\big\| : \eta \in Y^*,\ \|\eta\| \le r\big\}, \quad r \ge 0. \tag{12.3}$$

Here the operator $F'[x^\dagger]$ shall be defined as in Subsection 12.1.1. Obviously d is monotonically decreasing and $0 \le d(r) \le \|\xi^\dagger\|$ for all $r \ge 0$. In case of reflexive Banach spaces the infimum is attained (see [HH09, Section 3]) and analysis can be based on the corresponding minimizers. In the following we do not assume reflexivity, but applying slightly refined techniques we obtain the same results as in [HH09], even if the infimum in (12.3) is not attained.

The following proposition states that d is convex, which in combination with $d(r) < \infty$ for all $r \ge 0$ implies continuity of d on $(0, \infty)$.

Proposition 12.4. *The distance function d defined by* (12.3) *is convex.*

Proof. The proof generalizes the corresponding one given in [FHM11] for distance functions in Hilbert spaces. For reflexive Banach spaces the assertion has been proven in [BH10] by arguments from convex analysis. Our proof is elementary and works for general Banach spaces.

Let $r, \tilde{r} \geq 0$ and $\lambda \in [0,1]$. We want to show $d(\lambda r + (1-\lambda)\tilde{r}) \leq \lambda d(r) + (1-\lambda)d(\tilde{r})$. For each $\eta \in Y^*$ with $\|\eta\| \leq r$ and each $\tilde{\eta} \in Y^*$ with $\|\tilde{\eta}\| \leq \tilde{r}$ we have $\|\lambda\eta + (1-\lambda)\tilde{\eta}\| \leq \lambda r + (1-\lambda)\tilde{r}$. Thus,

$$\begin{aligned} d\big(\lambda r + (1-\lambda)\tilde{r}\big) &\leq \big\|\xi^\dagger - F'[x^\dagger]^*\big(\lambda\eta + (1-\lambda)\tilde{\eta}\big)\big\| \\ &= \big\|\lambda(\xi^\dagger - F'[x^\dagger]^*\eta) + (1-\lambda)(\xi^\dagger - F'[x^\dagger]^*\tilde{\eta})\big\| \\ &\leq \lambda\big\|\xi^\dagger - F'[x^\dagger]^*\eta\big\| + (1-\lambda)\big\|\xi^\dagger - F'[x^\dagger]^*\tilde{\eta}\big\|. \end{aligned}$$

Since this is true for all $\eta, \tilde{\eta} \in Y^*$ with $\|\eta\| \leq r$ and $\|\tilde{\eta}\| \leq \tilde{r}$, we may pass to the infimum over η and $\tilde{\eta}$, yielding $d(\lambda r + (1-\lambda)\tilde{r}) \leq \lambda d(r) + (1-\lambda)d(\tilde{r})$. □

One easily verifies that $d(r)$ decays to zero at infinity if and only if $\xi^\dagger \in \overline{\mathcal{R}\big(F'[x^\dagger]^*\big)}$. See [HH09, Remark 4.2] for a discussion of this last condition. The case $\xi^\dagger \in \mathcal{R}\big(F'[x^\dagger]^*\big)$ can be characterized as follows.

Proposition 12.5. *Let $\xi^\dagger \in \partial\Omega(x^\dagger)$ and let d be the associated distance function defined by (12.3). There exists some $r_0 \geq 0$ with $d(r_0) = 0$ if and only if there is some $\eta^\dagger \in Y^*$ with $\|\eta^\dagger\| \leq r_0$ such that $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$.*

Proof. Assume $d(r_0) = 0$ for some $r_0 \geq 0$. Then there is a sequence $(\eta_k)_{k\in\mathbb{N}}$ in Y^* with $\|\eta_k\| \leq r_0$ and $\|\xi^\dagger - F'[x^\dagger]^*\eta_k\| \to 0$. Thus, for each $x \in X$ we have $\langle \xi^\dagger - F'[x^\dagger]^*\eta_k, x\rangle \to 0$ or, equivalently, $\langle F'[x^\dagger]^*\eta_k, x\rangle \to \langle \xi^\dagger, x\rangle$. Since $\langle F'[x^\dagger]^*\eta_k, x\rangle \leq r_0\|F'[x^\dagger]x\|$ for all k, we obtain $\langle \xi^\dagger, x\rangle \leq r_0\|F'[x^\dagger]x\|$ for all $x \in X$. The first direction of the assertion follows now from [SGG+09, Lemma 8.21].

If there is some $\eta^\dagger \in Y^*$ with $\|\eta^\dagger\| \leq r_0$ and $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$, then

$$d(r_0) \leq \|\xi^\dagger - F'[x^\dagger]^*\eta^\dagger\| = 0.$$

Thus, also the second direction of the assertion is true. □

The proposition shows that the only interesting case is $d(r) > 0$ for all $r \geq 0$. For obtaining convergence rates we furthermore assume $d(r) \to 0$ if $r \to \infty$ or, equivalently, $\xi^\dagger \in \overline{\mathcal{R}\big(F'[x^\dagger]^*\big)}$. Exploiting convexity one easily shows that d has to be strictly monotonically decreasing in this case.

We formalize the concept of approximate source conditions in a definition.

Definition 12.6. The exact solution $x^\dagger$ satisfies an *approximate source condition* with respect to the stabilizing functional Ω and to the operator $F'[x^\dagger]$ if there is a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ such that the associated distance function d defined by (12.3) decays to zero at infinity.

Depending on the decay rate of the distance function convergence rates were obtained in [HH09] in case of reflexive Banach spaces. Under the additional assumption that the Bregman distance $B^\Omega_{\xi^\dagger}(\bullet, x^\dagger)$ is q-coercive (see [HH09, Example 2.3] or Section 12.5) higher rates were shown in the same article.

Following similar arguments as in [HH09] we prove convergence rates depending on the distance function d without reflexivity assumption. We restrict ourselves to bounded linear operators $A := F$ with $D(F) = X$ and $F'[x^\dagger] = A$, since we only want to

demonstrate the ideas of the proof. The proposition can be extended to nonlinear operators (by imposing assumptions on the structure of nonlinearity) and also to q-coercive Bregman distances (cf. Proposition 12.35) along the lines of the corresponding proofs given in [HH09].

Proposition 12.7. *Assume $p > 1$ in (12.2) and that $A := F$ is bounded and linear. Further assume that $x^\dagger$ satisfies an approximate source condition with $\xi^\dagger \in \partial\Omega(x^\dagger)$ such that the associated distance function fulfills $d(r) > 0$ for all $r \geq 0$. Define functions Φ and Ψ by $\Phi(r) := \frac{d(r)}{r}$ and $\Psi(r) := r^{-p} d(r)^{p-1}$ for $r \in (0, \infty)$. Then*

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}\big(d\big(\Phi^{-1}(\delta)\big)\big) \quad \text{if } \delta \to 0$$

with the a priori parameter choice $\delta \mapsto \alpha(\delta)$ defined by $\delta^p = \alpha(\delta) d\big(\Psi^{-1}(\alpha)\big)$.

Proof. Obviously the functions Φ and Ψ are strictly monotonically decreasing with range $(0, \infty)$. Thus, the inverse functions Φ^{-1} and Ψ^{-1} are well-defined on $(0, \infty)$ and also strictly monotonically decreasing with range $(0, \infty)$. As a consequence the parameter choice is uniquely determined by $\delta^p = \alpha(\delta) d\big(\Psi^{-1}(\alpha(\delta))\big)$ (the right-hand side is strictly monotonically increasing with respect to α and has range $(0, \infty)$). From this equation we immediately obtain $\alpha(\delta) \to 0$ if $\delta \to 0$ and also $\frac{\delta^p}{\alpha(\delta)} \to 0$ if $\delta \to 0$. These facts will be used later in the proof.

For the sake of brevity we now write α instead of $\alpha(\delta)$.

The first of two major steps of the proof is to show the existence of $\bar{\delta} > 0$ such that the set

$$M := \bigcup_{\delta \in (0, \bar{\delta}]} \bigcup_{\{y^\delta : \|y^\delta - y^0\| \leq \delta\}} \operatorname*{argmin}_{x \in X} T_\alpha^{y^\delta}(x)$$

is bounded. Assume the contrary, which means that there are sequences $(\delta_k)_{k \in \mathbb{N}}$ in $(0, \infty)$ converging to zero, $(y_k)_{k \in \mathbb{N}}$ in Y with $\|y_k - y^0\| \leq \delta_k$, and $(x_k)_{k \in \mathbb{N}}$ in X with $x_k \in \operatorname{argmin}_{x \in X} T_{\alpha_k}^{y^{\delta_k}}$ such that $\|x_k\| \to \infty$ (where $\alpha_k = \alpha(\delta_k)$). Since $\alpha_k \to 0$ and $\frac{\delta_k^p}{\alpha_k} \to 0$, by Corollary 4.2 there is a weakly convergent subsequence $(x_{k_l})_{l \in \mathbb{N}}$ of (x_k). Weakly convergent sequences in Banach spaces are bounded and thus $\|x_{k_l}\| \to \infty$ cannot be true. Consequently, there is some $\bar{\delta} > 0$ such that M is bounded.

In the second step we estimate the Bregman distance $B_{\xi^\dagger}^{\Omega}\big(x_\alpha^{y^\delta}, x^\dagger\big)$. For fixed $r \geq 0$ and each $\eta \in Y^*$ with $\|\eta\| \leq r$ we have

$$\begin{aligned} B_{\xi^\dagger}^{\Omega}\big(x_\alpha^{y^\delta}, x^\dagger\big) &= \Omega(x_\alpha^{y^\delta}) - \Omega(x^\dagger) + \langle A^*\eta - \xi^\dagger, x_\alpha^{y^\delta} - x^\dagger \rangle + \langle A^*(-\eta), x_\alpha^{y^\delta} - x^\dagger \rangle \\ &\leq \Omega(x_\alpha^{y^\delta}) - \Omega(x^\dagger) + \|\xi^\dagger - A^*\eta\| \|x_\alpha^{y^\delta} - x^\dagger\| + \|\eta\| \|A(x_\alpha^{y^\delta} - x^\dagger)\| \\ &\leq \Omega(x_\alpha^{y^\delta}) - \Omega(x^\dagger) + c\|\xi^\dagger - A^*\eta\| + r\|A(x_\alpha^{y^\delta} - x^\dagger)\|, \end{aligned}$$

where $c \geq 0$ denotes the bound on the set $M - x^\dagger$, that is, $\|x - x^\dagger\| \leq c$ for all $x \in M$. Taking the infimum over all $\eta \in Y^*$ with $\|\eta\| \leq r$ yields

$$B_{\xi^\dagger}^{\Omega}\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq \Omega(x_\alpha^{y^\delta}) - \Omega(x^\dagger) + c d(r) + r\|A(x_\alpha^{y^\delta} - x^\dagger)\|.$$

Due to the minimizing property of $x_\alpha^{y^\delta}$ we further obtain

$$\begin{aligned}\Omega(x_\alpha^{y^\delta}) - \Omega(x^\dagger) &= \frac{1}{\alpha}\Big(T_\alpha^{y^\delta}(x_\alpha^{y^\delta}) - \frac{1}{p}\|A(x_\alpha^{y^\delta} - x^\dagger)\|^p\Big) - \Omega(x^\dagger) \\ &\leq \frac{\delta^p}{p\alpha} - \frac{1}{p\alpha}\|A(x_\alpha^{y^\delta} - x^\dagger)\|^p\end{aligned}$$

and in combination with the previous estimate

$$B_{\xi^\dagger}^\Omega\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha} + r\|A(x_\alpha^{y^\delta} - x^\dagger)\| - \frac{1}{p\alpha}\|A(x_\alpha^{y^\delta} - x^\dagger)\|^p + cd(r).$$

Young's inequality

$$ab \leq \tfrac{1}{p}a^p + \tfrac{p-1}{p}b^{\frac{p}{p-1}}, \quad a, b \geq 0,$$

with $a := \alpha^{-\frac{1}{p}}\|A(x_\alpha^{y^\delta} - x^\dagger)\|$ and $b := r\alpha^{\frac{1}{p}}$ yields

$$r\|A(x_\alpha^{y^\delta} - x^\dagger)\| \leq \frac{1}{p\alpha}\|A(x_\alpha^{y^\delta} - x^\dagger)\|^p + \frac{p-1}{p}\alpha^{\frac{1}{p-1}}r^{\frac{p}{p-1}}.$$

Therefore,

$$B_{\xi^\dagger}^\Omega\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha} + \frac{p-1}{p}\alpha^{\frac{1}{p-1}}r^{\frac{p}{p-1}} + cd(r)$$

for all $\delta \in (0, \bar{\delta}]$ and all $r \geq 0$. We choose $r = r_\alpha := \Psi^{-1}(\alpha)$, which is equivalent to $r_\alpha^{-p}d(r_\alpha)^{p-1} = \alpha$ and thus also to $d(r_\alpha) = \alpha^{\frac{1}{p-1}}r_\alpha^{\frac{p}{p-1}}$. With this specific r the last estimate becomes

$$B_{\xi^\dagger}^\Omega\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha} + \Big(\frac{p-1}{p} + c\Big)d(r_\alpha)$$

and taking into account that the parameter choice satisfies $\delta^p = \alpha d(r_\alpha)$, we obtain

$$B_{\xi^\dagger}^\Omega\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq (1 + c)d(r_\alpha).$$

To complete the proof it remains to show $r_\alpha = \Phi^{-1}(\delta)$ or equivalently $\Phi(r_\alpha) = \delta$, which is a simple consequence of the parameter choice:

$$\Phi(r_\alpha) = \frac{d(r_\alpha)}{r_\alpha} = \left(\frac{d(r_\alpha)^{p-1}}{r_\alpha^p}d(r_\alpha)\right)^{\frac{1}{p}} = \big(\Psi(r_\alpha)d(r_\alpha)\big)^{\frac{1}{p}} = \big(\alpha d(r_\alpha)\big)^{\frac{1}{p}} = \delta.$$

□

Remark 12.8. As already mentioned in [HH09] the proposition remains true if the distance function d in the parameter choice and in the $\mathcal{O}$-expression is replaced by some strictly decreasing majorant of d.

By the definition of Φ in Proposition 12.7 we may write $d\big(\Phi^{-1}(\delta)\big) = \delta\Phi^{-1}(\delta)$. Thus, the convergence rate stated in the proposition is the higher the faster the distance function d decays to zero at infinity. We also see that the convergence rate always lies below $\mathcal{O}(\delta)$ (because $\Phi^{-1}(\delta) \to \infty$ if $\delta \to 0$).

Convergence rates results based on source conditions provide a common rate bound for all $x^\dagger$ with a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ satisfying $\xi^\dagger \in \{F'[x^\dagger]^*\eta : \eta \in Y^*,\ \|\eta\| \le c\}$ with some fixed $c > 0$. If we replace d by a majorant of d, in analogy to source conditions Proposition 12.7 provides a common rate bound for all $x^\dagger$ which have a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ such that the associated distance function lies below the fixed majorant. Thus, we can extend the elementwise convergence rates result based on approximate source conditions to a rates result for whole classes of exact solutions $x^\dagger$.

12.1.4. Variational inequalities

Approximate source conditions overcome the coarse scale of solution smoothness provided by source conditions, but two major problems remain unsolved: approximate source conditions require additional assumptions on the nonlinearity structure of the operator F to provide convergence rates and approximate source conditions rely on norm based fitting terms in the Tikhonov functional.

To avoid assumptions on the nonlinearity of F, the new concept of variational inequalities was introduced in [HKPS07]. It was show there that an inequality

$$\langle -\xi^\dagger, x - x^\dagger\rangle \le \beta_1 B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \beta_2\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in M$$

with $\xi^\dagger \in \Omega(x^\dagger)$, $\beta_1 \in [0, 1)$, and $\beta_2 \ge 0$ holding on a sufficiently large set $M \subseteq D(F)$ yields the convergence rate

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}(\delta) \quad \text{if } \delta \to 0$$

for a suitable parameter choice $\delta \mapsto \alpha(\delta)$. In [HKPS07] a concrete set M is used, but careful inspection of the proofs shows that any set can be chosen if all regularized solutions $x_{\alpha(\delta)}^{y^\delta}$ belong to this set.

This first version of variational inequalities does not require any additional assumption on the smoothness of $x^\dagger$ or on the nonlinearity of F to provide convergence rates. But as for source conditions the scale of smoothness is very coarse; either a variational inequality is satisfied or not. Considering a certain assumption on the nonlinearity of F in combination with a source condition the authors of [HH09] derived an inequality of the form

$$\langle -\xi^\dagger, x - x^\dagger\rangle \le \beta_1 B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \beta_2\|F(x) - F(x^\dagger)\|^\kappa \quad \text{for all } x \in M$$

with $\kappa \in (0, 1]$ and obtained the convergence rate

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}(\delta^\kappa) \quad \text{if } \delta \to 0$$

from it. Thus, introducing the exponent κ extends the scale of convergence rates to powers of δ.

A further step of generalization was undertaken in [BH09] (published in final form as [BH10]). There the exponent κ has been replaced by some strictly monotonically increasing and concave function $\varphi : [0, \infty) \to [0, \infty)$ satisfying $\varphi(0) = 0$. Thus, the variational inequality reads as

$$\langle -\xi^\dagger, x - x^\dagger\rangle \le \beta_1 B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \varphi\big(\|F(x) - F(x^\dagger)\|\big) \quad \text{for all } x \in M$$

and the corresponding convergence rate is

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}(\varphi(\delta)) \quad \text{if } \delta \to 0.$$

As in Part I we prefer to write variational inequalities in the form

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + \varphi\big(\|F(x) - F(x^\dagger)\|\big) \quad \text{for all } x \in M,$$

where $\beta = 1 - \beta_1 \in (0, 1]$. The advantage is that the right-hand side satisfies

$$\Omega\big(x_{\alpha(\delta)}^{y^\delta}\big) - \Omega(x^\dagger) + \varphi\big(\big\|F\big(x_{\alpha(\delta)}^{y^\delta}\big) - F(x^\dagger)\big\|\big) = \mathcal{O}(\varphi(\delta)) \quad \text{if } \delta \to 0$$

for a suitable a priori parameter choice (cf. Sections 4.1 and 4.2). Thus, assuming a variational inequality means that the error measure $B_{\xi^\dagger}^{\Omega}(\bullet, x^\dagger)$ shall be bounded by a term realizing the desired convergence rate. Note that a variational inequality with $\beta > 1$ implies a variational inequality with $\beta = 1$ (or below). Therefore it suffices to consider $\beta \in (0, 1]$, which has the advantage that the difference $\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) - (\Omega(x) - \Omega(x^\dagger))$ is a concave function in x. This observation will turn out very useful in the subsequent sections.

Remark 12.9. For $p > 1$ the function φ used in the present part is different from the function φ in the variational inequality (4.3) of Part I. The difference lies in the fact that we wrote variational inequalities in Part I with the term $\varphi\big(\frac{2^{1-p}}{p}\|F(x) - F(x^\dagger)\|^p\big)$ (cf. Proposition 2.13) instead of $\varphi(\|F(x) - F(x^\dagger)\|)$, which we use now. But this is simply a matter of scaling. In fact the additional exponent p occurring in the form used in Part I is compensated by the function $\psi(\delta) = \frac{1}{p}\delta^p$ in the data model (see Assumption 4.1 and Example 4.3).

In [Pös08] variational inequalities were adapted to Tikhonov-type regularization with more general fitting functionals not based on the norm in Y. And [Gra10a] contains a variational inequality yielding convergence rates for general fitting functionals and error measures other than the Bregman distance $B_{\xi^\dagger}^{\Omega}(\bullet, x^\dagger)$. To our knowledge the most general form of variational inequalities, including all previous versions as special cases, is the one suggested in Part I (see (4.3) and the discussion thereafter).

Next to the articles cited above, variational inequalities are also applied in [Hei08a, KH10, AR11] to obtain convergence rates. One also finds a multiplicative form of variational inequalities in the literature (see [KH10]), which we do not discuss here.

A difficult question is which classes of functions φ should be considered in a variational inequality. The difficulty arises from the fact that only the local behavior of φ around zero has influence an the obtained convergence rate. In Section 4.2 we struggled through the technicalities of such local considerations (cf. Assumption 4.9), but now we restrict our attention to the more pleasing case that φ is monotonically increasing and concave on its whole domain $[0, \infty)$.

Of course the question arises whether convex functions φ should be considered, too. As we indicate now, the answer is 'no'. If φ is convex with $\varphi(0) = 0$ then $t \mapsto \frac{\varphi(t)}{t}$ is monotonically increasing on $(0, \infty)$ and thus $\lim_{t \to +0} \frac{\varphi(t)}{t} < \infty$. In case the limit is not zero, the function φ behaves linearly near zero and thus the corresponding convergence

rate is the same as obtained from a variational inequality with linear (that is concave) φ. As the following proposition shows, under reasonable assumptions on the operator F and on the set M the remaining case $\lim_{t\to+0}\frac{\varphi(t)}{t}=0$ can only occur in the singular situation that $x^\dagger$ minimizes Ω over M.

Proposition 12.10. *Let $\xi^\dagger\in\partial\Omega(x^\dagger)$, $M\subseteq X$, and $\beta\in(0,1]$. Assume that M is starlike and that there is an operator $F'[x^\dagger]$ as defined in Subsection 12.1.1. If $\varphi:[0,\infty)\to[0,\infty)$ satisfies $\varphi(0)=0$ and $\lim_{t\to+0}\frac{\varphi(t)}{t}=0$, then the variational inequality*

$$\beta B_{\xi^\dagger}^{\Omega}(x,x^\dagger)\le\Omega(x)-\Omega(x^\dagger)+\varphi\big(\|F(x)-F(x^\dagger)\|\big)\quad\text{for all } x\in M$$

implies $\Omega(x^\dagger)\le\Omega(x)$ for all $x\in M$.

Proof. We apply the variational inequality to $x:=x^\dagger+t(\tilde x-x^\dagger)$ with $\tilde x\in M$ and $t\in(0,t_0(\tilde x)]$, where $t_0(\tilde x)$ is small enough to ensure $x\in M$. With out loss of generality we assume $t_0(\tilde x)\le 1$. Multiplying the resulting inequality by $\frac{1}{t}$ yields

$$\frac{1-\beta}{t}\big(\Omega(x^\dagger)-\Omega(x^\dagger+t(\tilde x-x^\dagger))\big)-\beta\langle\xi^\dagger,\tilde x-x^\dagger\rangle\le\frac{\varphi\big(\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|\big)}{t}.$$

The convexity of Ω gives

$$\begin{aligned}\Omega(x^\dagger)-\Omega(x^\dagger+t(\tilde x-x^\dagger))&=\Omega(x^\dagger)-\Omega((1-t)x^\dagger+t\tilde x)\\&\ge\Omega(x^\dagger)-(1-t)\Omega(x^\dagger)-t\Omega(\tilde x)=t\Omega(x^\dagger)-t\Omega(\tilde x)\end{aligned}$$

and together with the assumption $\xi^\dagger\in\partial\Omega(x^\dagger)$ and the previous inequality we obtain

$$\begin{aligned}\Omega(x^\dagger)-\Omega(\tilde x)&\le(1-\beta)\big(\Omega(x^\dagger)-\Omega(\tilde x)\big)-\beta\langle\xi^\dagger,\tilde x-x^\dagger\rangle\\&\le\frac{\varphi\big(\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|\big)}{t}\end{aligned}$$

for all $t\in(0,t_0(\tilde x)]$.

If there is some $t\in(0,t_0(\tilde x)]$ with $\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|=0$ then $\Omega(x^\dagger)-\Omega(\tilde x)\le 0$ follows. If $\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|>0$ for all $t\in(0,t_0(\tilde x)]$ then the inequality can be written as

$$\Omega(x^\dagger)-\Omega(\tilde x)\le\frac{\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|}{t}\frac{\varphi\big(\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|\big)}{\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|}.$$

Since $\frac{1}{t}\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|\to\|F'[x^\dagger](\tilde x-x^\dagger)\|$ and $\|F(x^\dagger+t(\tilde x-x^\dagger))-F(x^\dagger)\|\to 0$ if $t\to 0$, the right-hand side goes to zero if $t\to 0$ and thus we have shown $\Omega(x^\dagger)-\Omega(\tilde x)\le 0$ for all $\tilde x\in M$. □

Under slightly stronger assumptions on Ω and F the result of the proposition was also obtained in [HY10] for monomials $\varphi(t)=t^\kappa$. In fact it was shown there that $\kappa>1$ cannot occur in a variational inequality.

The investigation of cross connections between variational inequalities and other smoothness concepts in subsequent sections will show that variational inequalities can be divided into two distinct classes depending on the concave function φ. Variational

inequalities with $\lim_{t\to+0}\frac{\varphi(t)}{t}=\infty$ show a common behavior and variational inequalities with $\lim_{t\to+0}\frac{\varphi(t)}{t}<\infty$ form the second class. In the latter case the function φ behaves linearly near zero and as the following proposition shows this second class of variational inequalities can be reduced to variational inequalities with linear φ. Note that a linear φ then yields the same convergence rate as the original φ.

Proposition 12.11. *Let $\xi^\dagger\in\partial\Omega(x^\dagger)$, $M\subseteq X$, and $\beta\in(0,1]$. Assume that $\varphi:[0,\infty)\to[0,\infty)$ satisfies $\varphi(0)=0$ and $c:=\lim_{t\to+0}\frac{\varphi(t)}{t}<\infty$. Then the variational inequality*

$$\beta B_{\xi^\dagger}^{\Omega}(x,x^\dagger)\le\Omega(x)-\Omega(x^\dagger)+\varphi\big(\|F(x)-F(x^\dagger)\|\big)\quad\text{for all } x\in M$$

implies

$$\beta B_{\xi^\dagger}^{\Omega}(x,x^\dagger)\le\Omega(x)-\Omega(x^\dagger)+c\|F(x)-F(x^\dagger)\|\quad\text{for all } x\in M.$$

Proof. By the concavity of φ and by the assumption $\varphi(0)=0$ the function $t\mapsto\frac{\varphi(t)}{t}$ is monotonically decreasing. Thus, for each $t>0$ we have $\varphi(t)=t\frac{\varphi(t)}{t}\le ct$. The assertion follows now with $t=\|F(x)-F(x^\dagger)\|$. □

We formalize the concept of variational inequalities as a definition.

Definition 12.12. Let $M\subseteq D(F)$ and let $\varphi:[0,\infty)\to[0,\infty)$ be a monotonically increasing and concave function with $\varphi(0)=0$ and $\lim_{t\to+0}\varphi(t)=0$, which is strictly increasing in a neighborhood of zero. The exact solution $x^\dagger$ satisfies a *variational inequality* on M with respect to φ if there are $\xi^\dagger\in\partial\Omega(x^\dagger)$ and $\beta\in(0,1]$ such that

$$\beta B_{\xi^\dagger}^{\Omega}(x,x^\dagger)\le\Omega(x)-\Omega(x^\dagger)+\varphi\big(\|F(x)-F(x^\dagger)\|\big)\quad\text{for all } x\in M. \tag{12.4}$$

Proposition 12.13. *Let $x^\dagger$ satisfy a variational inequality with respect to a sufficiently large set M and a function φ. Then*

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta},x^\dagger\big)=\mathcal{O}(\varphi(\delta))\quad\text{if } \delta\to0$$

with a parameter choice $\delta\mapsto\alpha(\delta)$ depending on φ but not on M. In the context of this proposition the set M is sufficiently large if there is some $\bar\delta>0$ such that

$$\bigcup_{\delta\in(0,\bar\delta]}\ \bigcup_{\{y^\delta:\|y^\delta-y^0\|\le\delta\}}\operatorname*{argmin}_{x\in D(F)}T_{\alpha(\delta)}^{y^\delta}(x)\subseteq M.$$

Proof. The proposition is a special case of Theorem 4.11 (mind Remark 12.9). □

Examples for a 'sufficiently large' set M and details on the parameter choice are discussed in Sections 4.1 and 4.2.

As obvious from Proposition 12.13, the best rate is obtained from a variational inequality with $\varphi(t)=ct$, $c>0$. The following proposition shows that such a variational inequality is indeed stronger than a variational inequality with a different concave function φ.

Proposition 12.14. *Let $x^\dagger$ satisfy a variational inequality*

$$\beta B^\Omega_{\xi^\dagger}(x,x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + c\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in M$$

with $\beta \in (0,1]$, $c > 0$, $M \subseteq D(F)$ and let φ be as in Definition 12.12. If there are constants $\tilde\beta \in (0,\beta]$ and $\tilde c > 0$ such that

$$\tilde\beta B^\Omega_{\xi^\dagger}(x,x^\dagger) - \big(\Omega(x) - \Omega(x^\dagger)\big) \le \tilde c \quad \text{for all } x \in M,$$

then

$$\tilde\beta B^\Omega_{\xi^\dagger}(x,x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + \frac{\tilde c}{\varphi\big(\frac{\tilde c}{c}\big)}\varphi\big(\|F(x) - F(x^\dagger)\|\big) \quad \text{for all } x \in M.$$

Proof. Set $h(x) := \tilde\beta B^\Omega_{\xi^\dagger}(x,x^\dagger) - \big(\Omega(x) - \Omega(x^\dagger)\big)$ for $x \in M$. We only have to consider $x \in M$ with $h(x) \in (0,\tilde c]$ (all other $x \in M$ satisfy $h(x) \le 0$). Since φ is monotonically increasing we may estimate

$$\begin{aligned} h(x) &= \frac{h(x)}{\varphi\big(\frac1c h(x)\big)}\varphi\big(\tfrac1c h(x)\big) \le \frac{h(x)}{\varphi\big(\frac1c h(x)\big)}\varphi\big(\tfrac1c \beta B^\Omega_{\xi^\dagger}(x,x^\dagger) - \tfrac1c\big(\Omega(x) - \Omega(x^\dagger)\big)\big) \\ &\le \frac{h(x)}{\varphi\big(\frac1c h(x)\big)}\varphi\big(\|F(x) - F(x^\dagger)\|\big). \end{aligned}$$

The concavity of φ implies that $t \mapsto \frac{t}{\varphi(t)}$ is monotonically increasing. Thus

$$\frac{h(x)}{\varphi\big(\frac1c h(x)\big)} = c\frac{\frac1c h(x)}{\varphi\big(\frac1c h(x)\big)} \le c\frac{\frac{\tilde c}{c}}{\varphi\big(\frac{\tilde c}{c}\big)} = \frac{\tilde c}{\varphi\big(\frac{\tilde c}{c}\big)}.$$

Combining the two estimates we obtain

$$h(x) \le \frac{\tilde c}{\varphi\big(\frac{\tilde c}{c}\big)}\varphi\big(\|F(x) - F(x^\dagger)\|\big)$$

for all $x \in M$ with $h(x) \in (0,\tilde c]$ and thus for all $x \in M$. □

Remark 12.15. In Proposition 12.14 we assumed the existence of $\tilde c > 0$ and $\tilde\beta > 0$ such that

$$\tilde\beta B^\Omega_{\xi^\dagger}(x,x^\dagger) - \big(\Omega(x) - \Omega(x^\dagger)\big) \le \tilde c \quad \text{for all } x \in M.$$

This assumption is not very strong and due to the weak compactness of the sublevel sets of Ω we believe that this assumption is always satisfied, but we have no proof.

In case of $\Omega = \frac1q\|\bullet\|^q$ with $q \ge 1$ and $M = X$ the assumption holds at least for $\tilde\beta \le \frac{1}{2^{q-1}+1}$. Indeed, using $\xi^\dagger \in \partial\Omega(x^\dagger)$ and the inequality

$$\|\tilde x - x\|^q \le 2^{q-1}\big(\|\tilde x\|^q + \|x\|^q\big) \quad \text{for } \tilde x, x \in X$$

(see [SGG+09, Lemma 3.20]) we obtain

$$\begin{aligned} \langle -\xi^\dagger, x - x^\dagger\rangle = \langle \xi^\dagger, (2x^\dagger - x) - x^\dagger\rangle &\le \frac1q\big(\|2x^\dagger - x\|^q - \|x^\dagger\|^q\big) \\ &\le \frac1q\big(2^{q-1}\|x\|^q + \big(2^{2q-1} - 1\big)\|x^\dagger\|^q\big). \end{aligned}$$

Thus,

$$\begin{aligned}\tilde{\beta}B^{\Omega}_{\xi^\dagger}(x,x^\dagger)-\big(\Omega(x)-\Omega(x^\dagger)\big) &= \frac{1-\tilde{\beta}}{q}\big(\|x^\dagger\|^q-\|x\|^q\big)+\tilde{\beta}\langle-\xi^\dagger,x-x^\dagger\rangle\\ &\le \frac{\big(2^{q-1}+1\big)\tilde{\beta}-1}{q}\|x\|^q+\frac{\big(2^{2q-1}-1\big)\tilde{\beta}+1}{q}\|x^\dagger\|^q\\ &\le \frac{\big(2^{2q-1}-1\big)\tilde{\beta}+1}{q}\|x^\dagger\|^q=:\tilde{c}\end{aligned}$$

for all $x\in M$.

For $\Omega=\frac{1}{q}\|\bullet\|^q$ one can even show that the assumption of Proposition 12.14 is satisfied for all $\tilde{\beta}\in(0,1)$. But this requires advanced techniques (duality mappings and their relation to subdifferentials) which we do not want to introduce here.

One should be aware of the fact that due to the concavity of φ the best possible convergence rate obtainable from a variational inequality is $B^{\Omega}_{\xi^\dagger}\big(x^{y^\delta}_{\alpha(\delta)},x^\dagger\big)=\mathcal{O}(\delta)$. As we will see in Chapter 13 higher rates can be obtained in Hilbert spaces by extending the concept of variational inequalities slightly.

12.1.5. Approximate variational inequalities

Before variational inequalities with exponent κ introduced in [HH09] have been extended to variational inequalities with a more general function φ (cf. Subsection 12.1.4) another concept for expressing various types of smoothness was introduced in [Gei09] and [FH10]: approximate variational inequalities.

The idea is to overcome the limited scale of convergence rates obtainable via variational inequalities with φ of power-type by measuring the violation of a fixed benchmark variational inequality, thereby preserving the applicability to nonlinear operators and to non-metric fitting terms in the Tikhonov-type functional. The benchmark inequality should provide as high rates as possible because as for approximate source conditions the convergence rate obtained from the approximate variant is limited by the rates provided by the benchmark inequality. Thus, $\varphi(t)=t$ is a suitable benchmark.

Given $\xi^\dagger\in\partial\Omega(x^\dagger)$, $\beta\in(0,1]$, and $M\subseteq D(F)$ with $x^\dagger\in M$ we define the *distance function* $D_\beta:[0,\infty)\to[0,\infty]$ by

$$D_\beta(r):=\sup_{x\in M}\big(\beta B^{\Omega}_{\xi^\dagger}(x,x^\dagger)-\big(\Omega(x)-\Omega(x^\dagger)\big)-r\|F(x)-F(x^\dagger)\|\big),\quad r\ge 0. \tag{12.5}$$

Obviously D_β is monotonically decreasing and since $x^\dagger\in M$, we have $D_\beta(r)\ge 0$ for all $r\ge 0$. It might happen that $D_\beta(r)=\infty$ for some r. As a supremum of affine functions D_β is a convex function and thus continuous on the interior of its essential domain $\{r\ge 0: D_\beta(r)<\infty\}$.

From (12.5) we immediately see that there is some $r_0\ge 0$ with $D_\beta(r_0)=0$ if and only if the variational inequality

$$\beta B^{\Omega}_{\xi^\dagger}(x,x^\dagger)\le\Omega(x)-\Omega(x^\dagger)+r_0\|F(x)-F(x^\dagger)\|\quad\text{for all } x\in M$$

is satisfied. If D_β decays to zero at infinity we can derive convergence rates depending on the speed of this decay. Thus, in analogy to approximate source conditions we formulate the following definition.

Definition 12.16. Let $M \subseteq D(F)$ with $x^\dagger \in M$. The exact solution $x^\dagger$ satisfies an *approximate variational inequality* with respect to the stabilizing functional Ω and to the operator F if there are a subgradient $\xi^\dagger \in \Omega(x^\dagger)$ and a constant $\beta \in (0,1]$ such that the associated distance function D_β defined by (12.5) decays to zero at infinity.

Exploiting convexity one easily shows that D_β is strictly monotonically decreasing on its essential domain if it decays to zero at infinity but never attains zero.

In preparation of the convergence rates result connected with approximate variational inequalities we formulate a lemma.

Lemma 12.17. *Assume $p > 1$ in (12.2) and let $x^\dagger$ satisfy an approximate variational inequality with $\xi^\dagger \in \partial\Omega(x^\dagger)$, $\beta \in (0,1]$, and M sufficiently large. Here the set M is sufficiently large if there is some $\bar\delta > 0$ such that*

$$\bigcup_{\delta\in(0,\bar\delta]} \bigcup_{\{y^\delta:\|y^\delta - y^0\|\le\delta\}} \operatorname*{argmin}_{x\in D(F)} T^{y^\delta}_{\alpha(\delta)}(x) \subseteq M.$$

Then

$$\beta B^\Omega_{\xi^\dagger}\big(x^{y^\delta}_\alpha, x^\dagger\big) \le \frac{\delta^p}{p\alpha} + \frac{p-1}{p}\alpha^{\frac{1}{p-1}} r^{\frac{p}{p-1}} + D_\beta(r)$$

for all $r \ge 0$, all $\alpha > 0$, and all $\delta \in (0,\bar\delta]$.

Proof. By the definition of $D_\beta(r)$ we have

$$\beta B^\Omega_{\xi^\dagger}\big(x^{y^\delta}_\alpha, x^\dagger\big) \le \Omega(x^{y^\delta}_\alpha) - \Omega(x^\dagger) + r\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\| + D_\beta(r) \quad \text{for all } x \in M.$$

Using the minimizing property of $x^{y^\delta}_\alpha$ we obtain

$$\begin{aligned}\Omega(x^{y^\delta}_\alpha) - \Omega(x^\dagger) &= \frac{1}{\alpha}\Big(T^{y^\delta}_\alpha(x^{y^\delta}_\alpha) - \frac{1}{p}\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\|^p\Big) - \Omega(x^\dagger)\\ &\le \frac{\delta^p}{p\alpha} - \frac{1}{p\alpha}\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\|^p\end{aligned}$$

and the combination of both estimates yields

$$\beta B^\Omega_{\xi^\dagger}\big(x^{y^\delta}_\alpha, x^\dagger\big) \le \frac{\delta^p}{p\alpha} + r\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\| - \frac{1}{p\alpha}\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\|^p + D_\beta(r).$$

Applying Young's inequality

$$ab \le \tfrac{1}{p}a^p + \tfrac{p-1}{p}b^{\frac{p}{p-1}}, \quad a, b \ge 0,$$

with $a := \alpha^{-\frac{1}{p}}\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\|$ and $b := r\alpha^{\frac{1}{p}}$ we see

$$r\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\| \le \frac{1}{p\alpha}\|F(x^{y^\delta}_\alpha) - F(x^\dagger)\|^p + \frac{p-1}{p}\alpha^{\frac{1}{p-1}} r^{\frac{p}{p-1}}$$

and therefore

$$\beta B^\Omega_{\xi^\dagger}\big(x^{y^\delta}_\alpha, x^\dagger\big) \le \frac{\delta^p}{p\alpha} + \frac{p-1}{p}\alpha^{\frac{1}{p-1}} r^{\frac{p}{p-1}} + D_\beta(r).$$

□

We provide two expressions for the convergence rate obtainable from an approximate variational inequality: the one already given in [Gei09, FH10] and a version which exploits the technique of conjugate functions. In Proposition 12.22 below we show that both expressions describe the same rate.

Proposition 12.18. *Let the assumptions of Lemma 12.17 be satisfied.*

- *Assume $D_\beta(r) > 0$ for all $r \geq 0$ and define functions Φ and Ψ by $\Phi(r) := \frac{D_\beta(r)}{r}$ and $\Psi(r) := r^{-p} D_\beta(r)^{p-1}$ for $r \in (0, \infty)$. Then*

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}\big(D_\beta\big(\Phi^{-1}(\delta)\big)\big) \quad \text{if } \delta \to 0$$

  with the a priori parameter choice $\delta \mapsto \alpha(\delta)$ defined by $\delta^p = \alpha(\delta) D_\beta\big(\Psi^{-1}(\alpha)\big)$.

- *Assume $D_\beta(0) > 0$. Then*

$$B_{\xi^\dagger}^{\Omega}\big(x_{\alpha(\delta)}^{y^\delta}, x^\dagger\big) = \mathcal{O}\big(-D_\beta^*(-\delta)\big) \quad \text{if } \delta \to 0$$

  with the a priori parameter choice $\alpha(\delta) := \frac{\delta^p}{r(\delta)}$, where $r(\delta) \in \operatorname{argmin}_{r>0}(\delta r + D_\beta(r))$.

Proof. We show the first assertion. Because $D_\beta(r) \to 0$ if $r \to \infty$, there is some $r_0 \geq 0$ such that $D_\beta < \infty$ on (r_0, ∞). Obviously the functions Φ and Ψ are strictly monotonically decreasing on (r_0, ∞) with range $(0, D_\beta(r_0))$. Thus, the inverse functions Φ^{-1} and Ψ^{-1} are well-defined on $(0, D_\beta(r_0))$ and also strictly monotonically decreasing with range (r_0, ∞). As a consequence the parameter choice is uniquely determined by $\delta^p = \alpha(\delta) D_\beta\big(\Psi^{-1}(\alpha(\delta))\big)$ for small $\delta > 0$. (the right-hand side is strictly monotonically increasing with respect to $\alpha \in (0, D_\beta(r_0))$ and has range $\big(0, D_\beta(r_0)^2\big)$). For the sake of brevity we now write α instead of $\alpha(\delta)$.

Lemma 12.17 provides

$$\beta B_{\xi^\dagger}^{\Omega}\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha} + \frac{p-1}{p} \alpha^{\frac{1}{p-1}} r^{\frac{p}{p-1}} + D_\beta(r)$$

for all $\delta \in (0, \bar{\delta}]$ and all $r \geq 0$. We choose $r = r_\alpha := \Psi^{-1}(\alpha)$, which is equivalent to $r_\alpha^{-p} D_\beta(r_\alpha)^{p-1} = \alpha$ and thus also to $D_\beta(r_\alpha) = \alpha^{\frac{1}{p-1}} r_\alpha^{\frac{p}{p-1}}$. With this specific r the last estimate becomes

$$\beta B_{\xi^\dagger}^{\Omega}\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha} + \Big(\frac{p-1}{p} + 1\Big) D_\beta(r_\alpha)$$

and taking into account that the parameter choice satisfies $\delta^p = \alpha D_\beta(r_\alpha)$, we obtain

$$\beta B_{\xi^\dagger}^{\Omega}\big(x_\alpha^{y^\delta}, x^\dagger\big) \leq 2 D_\beta(r_\alpha).$$

To complete the proof of the first assertion it remains to show $r_\alpha = \Phi^{-1}(\delta)$ or equivalently $\Phi(r_\alpha) = \delta$, which is a simple consequence of the parameter choice:

$$\Phi(r_\alpha) = \frac{D_\beta(r_\alpha)}{r_\alpha} = \left(\frac{D_\beta(r_\alpha)^{p-1}}{r_\alpha^p} D_\beta(r_\alpha)\right)^{\frac{1}{p}} = \big(\Psi(r_\alpha) D_\beta(r_\alpha)\big)^{\frac{1}{p}} = \big(\alpha D_\beta(r_\alpha)\big)^{\frac{1}{p}} = \delta.$$

Now we come to the second assertion. We first show that the parameter choice $\alpha(\delta) := \frac{\delta^p}{r(\delta)}$ with $r(\delta) \in \operatorname{argmin}_{r>0} h_\delta(r)$ is well-defined, where we set $h_\delta(r) := \delta r + D_\beta(r)$ for $r \geq 0$. That is, we have to ensure the existence of $\tilde{\delta} > 0$ such that $\operatorname{argmin}_{r>0} h_\delta(r) \neq \emptyset$ for all $\delta \in (0, \tilde{\delta}]$. The functions h_δ are lower semi-continuous, $h_\delta(r) \to \infty$ if $r \to \infty$, and their essential domain coincides with the essential domain of D_β. Thus, if $D_\beta(0) = \infty$ then $\operatorname{argmin}_{r>0} h_\delta(r) \neq \emptyset$.

For the case $D_\beta(0) < \infty$ we give an indirect proof. So assume that there is no $\tilde{\delta}$ with the described property. Then there exists a sequence $(\delta_k)_{k\in\mathbb{N}}$ in $(0, \infty)$ converging to zero and satisfying $\operatorname{argmin}_{r>0} h_{\delta_k}(r) = \emptyset$ for all $k \in \mathbb{N}$. Therefore $h_{\delta_k}(r) \geq h_{\delta_k}(0)$ for all $r \geq 0$ and all $k \in \mathbb{N}$. Together with the monotonicity of D_β we thus obtain $0 \leq D_\beta(0) - D_\beta(r) \leq \delta_k r$ and $k \to \infty$ yields $D_\beta(r) = D_\beta(0)$ for all $r \geq 0$. But since $D_\beta(r) \to 0$ if $r \to \infty$ this means $D_\beta(r) = 0$ for all $r \geq 0$, which contradicts the assumption $D_\beta(0) > 0$. Therefore the proposed parameter choice is well-defined.

Finally we estimate the Bregman distance. From Lemma 12.17 we know

$$\beta B^{\Omega}_{\xi^\dagger}\big(x^{y^\delta}_{\alpha(\delta)}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha(\delta)} + \frac{p-1}{p}\alpha(\delta)^{\frac{1}{p-1}} r(\delta)^{\frac{p}{p-1}} + D_\beta(r(\delta))$$

with $r(\delta) \in \operatorname{argmin}_{r>0}(\delta r + D_\beta(r))$. Observing that $\alpha(\delta) = \frac{\delta^{p-1}}{r(\delta)}$ minimizes

$$\alpha \mapsto \frac{\delta^p}{p\alpha} + \frac{p-1}{p}\alpha^{\frac{1}{p-1}} r(\delta)^{\frac{p}{p-1}}$$

over $\alpha \in (0, \infty)$ we obtain

$$\beta B^{\Omega}_{\xi^\dagger}\big(x^{y^\delta}_{\alpha(\delta)}, x^\dagger\big) \leq \delta r(\delta) + D_\beta(r(\delta)) = \inf_{r>0}\big(\delta r + D_\beta(r)\big) = -\sup_{r>0}\big(-\delta r - D_\beta(r)\big).$$

By the lower semi-continuity of D_β we may extend the supremum to $r \geq 0$ without changing its value. Setting D_β to $+\infty$ on $(-\infty, 0)$, the supremum does not change if we allow $r \in \mathbb{R}$. Therefore

$$\beta B^{\Omega}_{\xi^\dagger}\big(x^{y^\delta}_{\alpha(\delta)}, x^\dagger\big) \leq -\sup_{r\in\mathbb{R}}\big(-\delta r - D_\beta(r)\big) = -D^*_\beta(-\delta).$$

□

Remark 12.19. For obtaining the second rate expression in Proposition 12.18 we had to exclude the case $D_\beta(0) = 0$. But this case is only of minor interest since it allows to show arbitrarily high convergence rates. Indeed, Lemma 12.17 with $r = 0$ provides

$$\beta B^{\Omega}_{\xi^\dagger}\big(x^{y^\delta}_{\alpha}, x^\dagger\big) \leq \frac{\delta^p}{p\alpha} \quad \text{for all } \alpha > 0$$

and thus, by applying a suitable parameter choice, any desirable rate can be proven. In addition we have

$$\Omega(x^\dagger) - \Omega(x) \leq \beta B^{\Omega}_{\xi^\dagger}(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x) \leq D_\beta(0) = 0 \quad \text{for all } x \in M,$$

that is, $x^\dagger$ minimizes Ω over M.

Remark 12.20. The function $-D_\beta^*(-\bullet)$ in Proposition 12.18 looks somewhat unusual, but it satisfies the typical properties one expects from a function in a convergence rates result. It is nonnegative, concave, upper semi-continuous, and monotonically increasing. Near zero it is even strictly monotonically increasing if we assume $D_\beta(0) > 0$. Further, $-D_\beta^*(-0) = 0$ and $-D_\beta^*(-t) \to 0$ if $t \to +0$. On $(-\infty, 0)$ this function attains the value $-\infty$. All these properties are shown in the proof of Lemma 12.32.

Remark 12.21. As for approximate source conditions Proposition 12.18 remains true if the distance function D_β is everywhere replaced by some decreasing majorant of D_β.

Finally we show that both $\mathcal{O}$-expressions stated in Proposition 12.18 describe the same convergence rate.

Proposition 12.22. *Let $x^\dagger$ satisfy an approximate variational inequality with $\xi^\dagger \in \partial\Omega(x^\dagger)$, $\beta \in (0, 1]$, and $M \subseteq D(F)$. Assume $D_\beta(r) > 0$ for all $r \geq 0$ and define the function Φ on $(0, \infty)$ by $\Phi(r) := \frac{D_\beta(r)}{r}$. Then*

$$D_\beta\big(\Phi^{-1}(\delta)\big) \leq -D_\beta^*(-\delta) \leq 2D_\beta\big(\Phi^{-1}(\delta)\big)$$

for all sufficiently small $\delta > 0$.

Proof. The proof is based on ideas presented in [Mat08]. Without loss of generality we assume $D_\beta < \infty$ on $[0, \infty)$; see also the first paragraph in the proof of Proposition 12.18. Then Φ^{-1} is well-defined on $(0, \infty)$.

By the definition of Φ we have $D_\beta\big(\Phi^{-1}(\delta)\big) = \delta\Phi^{-1}(\delta)$. Thus $r \geq \Phi^{-1}(\delta)$ implies

$$D_\beta\big(\Phi^{-1}(\delta)\big) = \delta\Phi^{-1}(\delta) \leq \delta r \leq \delta r + D_\beta(r).$$

On the other hand, for $r \leq \Phi^{-1}(\delta)$ by the monotonicity of D_β we obtain

$$D_\beta\big(\Phi^{-1}(\delta)\big) \leq D_\beta(r) \leq \delta r + D_\beta(r).$$

Both estimates together yield

$$D_\beta\big(\Phi^{-1}(\delta)\big) \leq \inf_{r\in[0,\infty)} \big(\delta r + D_\beta(r)\big) = -D_\beta^*(-\delta).$$

The second asserted inequality in the proposition follows from

$$-D_\beta^*(-\delta) = \inf_{r\in[0,\infty)} \big(\delta r + D_\beta(r)\big) \leq \delta\Phi^{-1}(\delta) + D_\beta\big(\Phi^{-1}(\delta)\big) = 2D_\beta\big(\Phi^{-1}(\delta)\big).$$

□

Approximate variational inequalities are a highly abstract tool and thus are not well suited for 'everyday' convergence rates analysis. But since they are an intermediate technique between variational inequalities and approximate source conditions they will turn out very useful for analyzing the relations between these two more accessible tools. To enlighten the role of approximate variational inequalities we show in Section 12.4 that each approximate variational inequality can be written as a variational inequality and vice versa.

12.1.6. Projected source conditions

The last smoothness concept we want to discuss is different from the previous ones because it is used in conjunction with constrained Tikhonov regularization. But as we show in Section 12.3 it provides a nice interpretation of variational inequalities. The results connected with projected source conditions are joint work with Bernd Hofmann (Chemnitz) and were published in [FH11].

In applications one frequently encounters Tikhonov regularization with convex constraints:

$$T_\alpha^{y^\delta}(x) = \tfrac{1}{p}\|F(x) - y^\delta\|^p + \alpha\Omega(x) \to \min_{x\in C}, \tag{12.6}$$

where $C \subseteq D(F)$ is a convex set. Also for such constrained problems source conditions yield convergence rates, but weaker assumptions on the smoothness of the exact solution $x^\dagger$ are adequate, too. Here the definition of $x^\dagger$ has to be adapted slightly: we assume $x^\dagger \in C$ and $\Omega(x^\dagger) = \min\{\Omega(x) : x \in C,\ F(x) = y^0\}$ instead of $\Omega(x^\dagger) = \min\{\Omega(x) : x \in D(F),\ F(x) = y^0\}$. In particular $\operatorname{argmin}\{\Omega(x) : x \in C,\ F(x) = y^0\} \neq \emptyset$ shall hold, which is true if C is closed and therefore, due to convexity, also weakly closed.

In [Neu88] constrained Tikhonov regularization in Hilbert spaces X and Y with $p = 2$, $\Omega = \|\bullet\|^2$, and a bounded linear operator $A = F$ is considered. It was shown there that the assumption $x^\dagger = P_C(A^*w)$ for some $w \in Y$ yields the convergence rate $\|x_{\alpha(\delta)}^{y^\delta} - x^\dagger\|^2 = \mathcal{O}(\delta)$ with a suitable parameter choice $\delta \mapsto \alpha(\delta)$. Here $P_C : X \to X$ denotes the metric projector onto the (closed) convex set C. Note that the condition $x^\dagger = P_C(A^*w)$ can be equivalently written as $A^*w - x^\dagger \in N_C(x^\dagger)$ with $N_C(x^\dagger) := \{\tilde{x} \in X : \langle \tilde{x}, x - x^\dagger\rangle \leq 0 \text{ for all } x \in C\}$ being the normal cone of C at $x^\dagger$.

The extension of such projected source conditions to Tikhonov regularization with a nonlinear operator F in Hilbert spaces X and Y is described in [CK94]. There the stabilizing functional $\Omega = \|\bullet - \bar{x}\|^2$ with fixed a priori guess $\bar{x} \in X$ is used. The corresponding projected source condition reads as $F'[x^\dagger]^*w - (x^\dagger - \bar{x}) \in N_C(x^\dagger)$ or, equivalently, $x^\dagger = P_C(\bar{x} + F'[x^\dagger]^*w)$, where in the context of [CK94] $F'[x^\dagger]$ denotes the Fréchet derivative of F at $x^\dagger$.

In Banach spaces the normal cone of C at $x^\dagger$ is defined by

$$N_C(x^\dagger) := \{\xi \in X^* : \langle \xi, x - x^\dagger\rangle \leq 0 \text{ for all } x \in X\}.$$

Using this definition we are able to extend projected source conditions to Banach spaces.

Definition 12.23. The exact solution $x^\dagger$ satisfies a *projected source condition* with respect to the operator $F'[x^\dagger]$ (cf. Subsection 12.1.1) if there are $\xi^\dagger \in \Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ such that

$$F'[x^\dagger]^*\eta^\dagger - \xi^\dagger \in N_C(x^\dagger). \tag{12.7}$$

If $x^\dagger$ belongs to the interior of C then $N_C(x^\dagger) = \{0\}$ and (12.7) reduces to the usual source condition $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$.

Before we derive convergence rates from a projected source condition we want to motivate the term 'projected' also for Banach spaces. To this end we assume that X is reflexive, strictly convex (see Definition B.6), and smooth (see Definition B.7). Then for each $x \in X$ there is a uniquely determined element $J(x) \in X^*$ such that

$\langle J(x), x\rangle = \|x\|^2 = \|J(x)\|^2$. The corresponding mapping $J : X \to X^*$, $x \mapsto J(x)$ has similar properties as the Riesz isomorphism in Hilbert spaces and is known as *duality mapping* on X. For details on duality mappings we refer to [BP86, Chapter 1, § 2.4]. The mapping J is bijective (see [BP86, Proposition 2.16]) and its inverse J^{-1} is the duality mapping $J_* : X^* \to X$ on X^* (see [Zei85, Proposition 47.19]). If $C \subseteq D(F)$ is closed and convex then the assumptions on X guarantee that the metric projector $P_C : X \to X$ onto C is well-defined. In other words, for each $x \in X$ there is exactly one element $x_C \in C$ such that $\|x_C - x\| = \min_{\tilde{x}\in C} \|\tilde{x} - x\|$ (see [BP86, Chapter 3, § 3.2]).

Following [Kie02, Proposition 2.2] we have $x^\dagger = P_C(x)$ with $x \in X$ if and only if $J(x - x^\dagger) \in N_C(x^\dagger)$. Since

$$F'[x^\dagger]^*\eta^\dagger - \xi^\dagger = J\big(x^\dagger + J_*\big(F'[x^\dagger]^*\eta^\dagger - \xi^\dagger\big) - x^\dagger\big)$$

we see that the projected source condition (12.7) is equivalent to

$$x^\dagger = P_C\big(x^\dagger + J_*(F'[x^\dagger]^*\eta^\dagger - \xi^\dagger)\big).$$

Thus, the term 'projected' is indeed appropriate.

Eventually, we formulate a convergence rates result based on a projected source condition. We restrict our attention to bounded linear operators $A = F$. For nonlinear operators additional assumptions on the structure of nonlinearity are required to obtain convergence rates, but the major steps of the proof are the same.

Proposition 12.24. *Let $A := F$ be bounded and linear. If there are $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ such that $A^*\eta^\dagger - \xi^\dagger \in N_C(x^\dagger)$, then*

$$B^\Omega_{\xi^\dagger}\big(x^{y^\delta}_{\alpha(\delta)}, x^\dagger\big) = \mathcal{O}(\delta) \quad \text{if } \delta \to 0$$

for an appropriate a priori parameter choice $\delta \mapsto \alpha(\delta)$.

Proof. By the definition of $N_C(x^\dagger)$ we have

$$\langle -\xi^\dagger, x - x^\dagger\rangle = \langle -\eta^\dagger, A(x - x^\dagger)\rangle \le \|\eta^\dagger\|\|A(x - x^\dagger)\| \quad \text{for all } x \in C.$$

Thus,

$$B^\Omega_{\xi^\dagger}(x, x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + \|\eta^\dagger\|\|A(x - x^\dagger)\| \quad \text{for all } x \in C.$$

This is a variational inequality as introduced in Definition 12.12 with $\beta = 1$, $\varphi(t) = \|\eta^\dagger\|t$, and $M = C$. Since $x^{y^\delta}_\alpha \in C$ for all $\alpha > 0$ and all $\delta > 0$ in constrained Tikhonov regularization, the set M is sufficiently large in the sense of Proposition 12.13. Thus, Proposition 12.13 applies and yields the desired rate. □

12.2. Auxiliary results on variational inequalities

In this section we present two results related to variational inequalities with $\varphi(t) = t$, which will be used in subsequent sections but which are also of independent interest.

The first proposition relates variational inequalities with the nonlinear operator F to variational inequalities formulated with the linearization $F'[x^\dagger]$ (cf. Subsection 12.1.1). Note that whenever we use the operator $F'[x^\dagger]$ we assume $D(F)$ to be starlike with respect to $x^\dagger$, else the operator is not well-defined.

Proposition 12.25. *Let $M \subseteq D(F)$ be starlike with respect to $x^\dagger$. If there are $\xi^\dagger \in \partial\Omega(x^\dagger)$, $\beta \in (0,1]$, and $c \geq 0$ such that*

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + c\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in M \tag{12.8}$$

then

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + c\|F'[x^\dagger](x - x^\dagger)\| \quad \text{for all } x \in M. \tag{12.9}$$

Proof. For fixed $x \in M$ there is $t_0 > 0$ such that $x^\dagger + t(x - x^\dagger) \in M$ for all $t \in [0, t_0]$. Thus for all $t \in (0, \min\{1, t_0\}]$ the given variational inequality (12.8) and the convexity of Ω imply

$$\begin{aligned}
&\frac{c}{t}\|F(x^\dagger + t(x - x^\dagger)) - F(x^\dagger)\| \\
&\quad \geq \frac{\beta}{t} B_{\xi^\dagger}^{\Omega}\big(x^\dagger + t(x - x^\dagger), x^\dagger\big) + \frac{1}{t}\big(\Omega(x^\dagger) - \Omega\big(x^\dagger + t(x - x^\dagger)\big)\big) \\
&\quad = \frac{1 - \beta}{t}\big(\Omega(x^\dagger) - \Omega\big((1 - t)x^\dagger + tx\big)\big) + \beta\langle -\xi^\dagger, x - x^\dagger\rangle \\
&\quad \geq (1 - \beta)\big(\Omega(x^\dagger) - \Omega(x)\big) + \beta\langle -\xi^\dagger, x - x^\dagger\rangle \\
&\quad = \beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x).
\end{aligned}$$

If we let $t \to +0$ we derive

$$\begin{aligned}
c\|F'[x^\dagger](x - x^\dagger)\| &= \lim_{t \to +0} \frac{c}{t}\|F(x^\dagger + t(x - x^\dagger)) - F(x^\dagger)\| \\
&\geq \beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x)
\end{aligned}$$

for all $x \in M$. Therefore the variational inequality (12.9) is valid. □

The reverse direction, from a variational inequality (12.9) with $F'[x^\dagger]$ back to a variational inequality (12.8) with F, requires additional assumptions on the structure of nonlinearity of F as discussed in Subsection 12.1.1.

The second result shows that for linear operators $A := F$ and under some regularity assumption the constant β in a variational inequality with linear φ plays only a minor role. That is, if a variational inequality holds for one $\beta \in (0,1]$ then it holds for all $\beta \in (0,1]$.

As preparation we state the following lemma, which is a separation theorem for convex sets. The lemma will be used in subsequent sections, too.

Lemma 12.26. *Let $E_1, E_2 \subseteq X \times \mathbb{R}$ be convex sets. If one of them has nonempty interior and the interior does not intersect with the other set then there exist $\xi \in X^*$ and $\tau \in \mathbb{R}$ with $(\xi, \tau) \neq (0, 0)$ such that*

$$\sup_{(x,t) \in E_1} \big(\langle \xi, x\rangle + \tau t\big) \leq \inf_{(x,t) \in E_2} \big(\langle \xi, x\rangle + \tau t\big).$$

Proof. The assertion is an immediate consequence of [BGW09, Theorem 2.1.2]. □

Note that replacing Lemma 12.26 by a separation theorem which works with less strong assumptions would allow to weaken the assumptions in some of the subsequent propositions and theorems. A more general separation theorem involving the notion of quasi convexity can be found in [CDB05]. See also [BCW08] for details on quasi convexity.

Proposition 12.27. *Let $A := F$ be a bounded linear operator with $D(F) = X$, let $M \subseteq X$ be convex, and let $\beta \in (0,1]$ and $c \geq 0$. Further assume that $N_{D(\Omega)}(x^\dagger) \cap (-N_M(x^\dagger)) = \{0\}$ and that at least one of the sets $D(\Omega)$ or M has interior points. Then there exists a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$ with*

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + c\|A(x - x^\dagger)\| \quad \text{for all } x \in M \tag{12.10}$$

if and only if

$$0 \leq \Omega(x) - \Omega(x^\dagger) + c\|A(x - x^\dagger)\| \quad \text{for all } x \in M. \tag{12.11}$$

Proof. If there is some $\xi^\dagger \in \partial\Omega(x^\dagger)$ such that (12.10) is true, then (12.11) is obviously also satisfied.

Assume (12.11). We apply Lemma 12.26 to the sets

$$E_1 := \{(x,t) \in X \times \mathbb{R} : t \leq \Omega(x^\dagger) - \Omega(x)\},$$
$$E_2 := \{(x,t) \in X \times \mathbb{R} : x \in M,\ t \geq c\|A(x - x^\dagger)\|\}.$$

To see that the assumptions of that lemma are satisfied first note that $\operatorname{int} E_1 \neq \emptyset$ if $\operatorname{int} D(\Omega) \neq \emptyset$ and $\operatorname{int} E_2 \neq \emptyset$ if $\operatorname{int} M \neq \emptyset$. Without loss of generality we assume $\operatorname{int} E_2 \neq \emptyset$ (the case $\operatorname{int} E_1 \neq \emptyset$ can be treated analogously). We have to show $E_1 \cap (\operatorname{int} E_2) = \emptyset$, which is true if $E_1 \cap E_2$ is a subset of the boundary of E_2. So let $(x,t) \in E_1 \cap E_2$ and set $(x_k, t_k) := (x, t - \frac{1}{k})$ for $k \in \mathbb{N}$. Then $(x_k, t_k) \to (x,t)$ (with respect to the norm topology) and using the definition of E_1 and the inequality (12.11) we obtain

$$t_k = t - \frac{1}{k} < t \leq \Omega(x^\dagger) - \Omega(x) \leq c\|A(x - x^\dagger)\| = c\|A(x_k - x^\dagger)\|,$$

that is, $(x_k, t_k) \notin E_2$. In other words, (x,t) is indeed a boundary point of E_2.

Taking into account $(x^\dagger, 0) \in E_1 \cap E_2$ Lemma 12.26 provides $\xi \in X^*$ and $\tau \in \mathbb{R}$ with

$$\langle \xi, x - x^\dagger \rangle + \tau t \leq 0 \quad \text{for all } (x,t) \in E_1, \tag{12.12}$$
$$\langle \xi, x - x^\dagger \rangle + \tau t \geq 0 \quad \text{for all } (x,t) \in E_2. \tag{12.13}$$

In case $\tau < 0$ inequality (12.13) yields $\langle -\frac{1}{\tau}\xi, x - x^\dagger \rangle \geq t$ for all $t \geq c\|A(x - x^\dagger)\|$ and all $x \in M$. This is obviously not possible (since $\langle -\frac{1}{\tau}\xi, x - x^\dagger \rangle < \infty$) and therefore $\tau \geq 0$ has to be true. If $\tau = 0$ then (12.12) implies $\langle \xi, x - x^\dagger \rangle \leq 0$ for all $x \in D(\Omega)$ and (12.13) implies $\langle \xi, x - x^\dagger \rangle \geq 0$ for all $x \in M$. Thus $\xi \in N_{D(\Omega)}(x^\dagger) \cap (-N_M(x^\dagger))$, which implies $\xi = 0$. This contradicts $(\xi, \tau) \neq (0,0)$.

It remains the case $\tau > 0$. Inequality (12.12) yields $\langle \frac{1}{\tau}\xi, x - x^\dagger \rangle \leq -t$ for all $(x,t) \in E_1$. With $t := \Omega(x^\dagger) - \Omega(x)$ this gives $\xi^\dagger := \frac{1}{\tau}\xi \in \partial\Omega(x^\dagger)$. From (12.13) with $t := c\|A(x - x^\dagger)\|$ we obtain $\langle -\xi^\dagger, x - x^\dagger \rangle \leq c\|A(x - x^\dagger)\|$ for all $x \in M$. Thus,

$$\begin{aligned}\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq B_{\xi^\dagger}^{\Omega}(x, x^\dagger) &= \Omega(x) - \Omega(x^\dagger) + \langle -\xi^\dagger, x - x^\dagger \rangle \\ &\leq \Omega(x) - \Omega(x^\dagger) + c\|A(x - x^\dagger)\|\end{aligned}$$

for all $x \in M$ and arbitrary $\beta \in (0,1]$. ∎

Note that Proposition 12.27 is in general not true for nonlinear operators F. In Section 12.5 we will see the influence of β on the convergence rate for variational inequalities with a function φ which is not linear.

12.3. Variational inequalities and (projected) source conditions

The aim of this section is to clarify the relation between variational inequalities and projected source conditions. Since usual source conditions are a special case of projected ones, the results also apply to usual source conditions. The results of this section are joint work with Bernd Hofmann (Chemnitz) and were published in [FH11].

In [SGG+09, Proposition 3.35] it is shown that a source condition implies a variational inequality (see also [HY10, Section 4]) if one imposes some assumption on the structure of nonlinearity of F. The function φ in the variational inequality depends on the chosen nonlinearity condition. For the sake of completeness we repeat the result here together with its proof, but we use only a very simple nonlinearity condition.

Proposition 12.28. *Let $M \subseteq D(F)$ and assume*

$$\|F'[x^\dagger](x - x^\dagger)\| \leq c\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in M$$

with a constant $c \geq 0$ and $F'[x^\dagger]$ as defined in Subsection 12.1.1. If there are $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ such that $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$ then

$$B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + c\|\eta^\dagger\|\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in M.$$

Proof. Let $x \in M$. The assertion follows from

$$\begin{aligned} B_{\xi^\dagger}^{\Omega}(x, x^\dagger) &= \Omega(x) - \Omega(x^\dagger) - \langle \xi^\dagger, x - x^\dagger \rangle \leq \Omega(x) - \Omega(x^\dagger) + \|\eta^\dagger\|\|F'[x^\dagger](x - x^\dagger)\| \\ &\leq \Omega(x) - \Omega(x^\dagger) + c\|\eta^\dagger\|\|F(x) - F(x^\dagger)\|. \end{aligned}$$

□

For projected source conditions we can prove a similar result.

Proposition 12.29. *Let $C \subseteq D(F)$ be convex and assume*

$$\|F'[x^\dagger](x - x^\dagger)\| \leq c\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in C$$

with a constant $c \geq 0$ and $F'[x^\dagger]$ as defined in Subsection 12.1.1. If there are $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ such that $F'[x^\dagger]^*\eta^\dagger - \xi^\dagger \in N_C(x^\dagger)$ then

$$B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + c\|\eta^\dagger\|\|F(x) - F(x^\dagger)\| \quad \text{for all } x \in C.$$

Proof. Let $x \in C$. By the definition of $N_C(x^\dagger)$ the projected source condition implies $\langle -\xi^\dagger, x - x^\dagger \rangle \leq \langle -\eta^\dagger, F'[x^\dagger](x - x^\dagger) \rangle$ and therefore

$$\begin{aligned} B_{\xi^\dagger}^{\Omega}(x, x^\dagger) &\leq \Omega(x) - \Omega(x^\dagger) + \|\eta^\dagger\|\|F'[x^\dagger](x - x^\dagger)\| \\ &\leq \Omega(x) - \Omega(x^\dagger) + c\|\eta^\dagger\|\|F(x) - F(x^\dagger)\|. \end{aligned}$$

Thus, the assertion is true. □

The reverse direction, from a variational inequality with linear function φ back to a source condition, has been show in [SGG+09, Proposition 3.38] under the additional assumption that F and Ω are Gâteaux differentiable at $x^\dagger$ (see also [HY10, Proposition 4.4]).

We generalize this finding in two points: on the one hand we do not require Gâteaux differentiability of F or Ω and on the other hand we derive a (projected) source condition also in the case $N_C(x^\dagger) \neq \{0\}$. In [SGG+09, Proposition 3.38] only $N_C(x^\dagger) = \{0\}$ is considered as a consequence of Gâteaux differentiability. In this connection reference should be made to the fact that the projected source condition $F'[x^\dagger]^*\eta^\dagger - \xi^\dagger \in N_C(x^\dagger)$ reduces to the usual source condition $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$ if $N_C(x^\dagger) = \{0\}$. Solely the assumption that $M = C$ is convex in our context is less general than the cited result, where M is starlike but not convex (except if F is linear).

To show that a variational inequality implies a projected source condition we start with linear operators $A = F$.

Lemma 12.30. *Let $A := F$ be bounded and linear with $D(F) = X$ and let $C \subseteq X$ be convex. Further assume $N_{D(\Omega)}(x^\dagger) \cap (-N_C(x^\dagger)) = \{0\}$ and that at least one of the sets $D(\Omega)$ or C has interior points. If*

$$0 \leq \Omega(x) - \Omega(x^\dagger) + c\|A(x - x^\dagger)\| \quad \text{for all } x \in C$$

with $c \geq 0$, then there are $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ with $\|\eta^\dagger\| \leq c$ such that $A^*\eta^\dagger - \xi^\dagger \in N_C(x^\dagger)$.

Proof. We apply Lemma 12.26 to the sets

$$E_1 := \{(x,t) \in X \times \mathbb{R} : x \in C,\ t \leq \Omega(x^\dagger) - \Omega(x)\},$$
$$E_2 := \{(x,t) \in X \times \mathbb{R} : t \geq c\|A(x - x^\dagger)\|\}$$

(the assumptions of that lemma can be verified analogously to the proof of Proposition 12.27). Together with $(x^\dagger, 0) \in E_1 \cap E_2$ Lemma 12.26 provides $\xi \in X^*$ and $\tau \in \mathbb{R}$ such that

$$\langle \xi, x - x^\dagger \rangle + \tau t \leq 0 \quad \text{for all } (x,t) \in E_1, \tag{12.14}$$
$$\langle \xi, x - x^\dagger \rangle + \tau t \geq 0 \quad \text{for all } (x,t) \in E_2. \tag{12.15}$$

If $\tau < 0$ then (12.15) implies $\langle -\frac{1}{\tau}\xi, x - x^\dagger \rangle \geq t$ for all $t \geq c\|A(x - x^\dagger)\|$ and all $x \in X$, which is obviously not possible (since $\langle -\frac{1}{\tau}\xi, x - x^\dagger \rangle < \infty$). In case $\tau = 0$ inequality (12.15) gives $\langle \xi, x - x^\dagger \rangle \geq 0$ for all $x \in X$. But this contradicts $(\xi, \tau) \neq (0,0)$. Thus, $\tau > 0$ has to be true.

From (12.15) with $t := c\|A(x - x^\dagger)\|$ we obtain for all $x \in X$

$$\langle -\tfrac{1}{\tau}\xi, x - x^\dagger \rangle \leq c\|A(x - x^\dagger)\|.$$

Hence there is some $\eta^\dagger \in Y^*$ such that $\frac{1}{\tau}\xi = A^*\eta^\dagger$ and $\|\eta^\dagger\| \leq c$ (see [SGG+09, Lemma 8.21]). Inequality (12.14) with $t := \Omega(x^\dagger) - \Omega(x)$ now yields $\Omega(x^\dagger) - \Omega(x) \leq \langle -A^*\eta^\dagger, x - x^\dagger \rangle$ for all $x \in C$.

To obtain a subgradient of Ω at $x^\dagger$ we apply Lemma 12.26 to the sets

$$\tilde{E}_1 := \{(x,t) \in X \times \mathbb{R} : t \le \Omega(x^\dagger) - \Omega(x)\},$$
$$\tilde{E}_2 := \{(x,t) \in X \times \mathbb{R} : x \in C,\ t \ge \langle -A^*\eta^\dagger, x - x^\dagger\rangle\}$$

(again, one easily verifies the assumptions of that lemma). With $(x^\dagger, 0) \in \tilde{E}_1 \cap \tilde{E}_2$ this yields $\tilde{\xi} \in X^*$ and $\tilde{\tau} \in \mathbb{R}$ such that

$$\langle \tilde{\xi}, x - x^\dagger\rangle + \tilde{\tau} t \le 0 \quad \text{for all } (x,t) \in \tilde{E}_1, \tag{12.16}$$
$$\langle \tilde{\xi}, x - x^\dagger\rangle + \tilde{\tau} t \ge 0 \quad \text{for all } (x,t) \in \tilde{E}_2. \tag{12.17}$$

If $\tilde{\tau} < 0$ then (12.17) implies $\langle -\frac{1}{\tilde{\tau}}\tilde{\xi}, x - x^\dagger\rangle \ge t$ for all $t \ge \langle -A^*\eta^\dagger, x - x^\dagger\rangle$ and all $x \in C$, which is obviously not possible (since $\langle -\frac{1}{\tilde{\tau}}\tilde{\xi}, x - x^\dagger\rangle < \infty$). In case $\tilde{\tau} = 0$ inequality (12.16) gives $\langle \tilde{\xi}, x - x^\dagger\rangle \le 0$ for all $x \in D(\Omega)$ and (12.17) gives $\langle \tilde{\xi}, x - x^\dagger\rangle \ge 0$ for all $x \in C$. Thus $\xi \in N_{D(\Omega)}(x^\dagger) \cap (-N_C(x^\dagger))$, which implies $\xi = 0$. This contradicts $(\xi, \tau) \ne (0,0)$. Therefore $\tilde{\tau} > 0$ is true.

With $t := \Omega(x^\dagger) - \Omega(x)$ from (12.16) we obtain $\langle \frac{1}{\tilde{\tau}}\tilde{\xi}, x - x^\dagger\rangle \le \Omega(x) - \Omega(x^\dagger)$ for all $x \in X$, that is, $\xi^\dagger := \frac{1}{\tilde{\tau}}\tilde{\xi} \in \partial\Omega(x^\dagger)$. Eventually, (12.17) with $t := \langle -A^*\eta^\dagger, x - x^\dagger\rangle$ yields $\langle -\xi^\dagger, x - x^\dagger\rangle \le \langle -A^*\eta^\dagger, x - x^\dagger\rangle$ for all $x \in C$. Thus, we have found $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$ such that $A^*\eta^\dagger - \xi^\dagger \in N_C(x^\dagger)$. □

Theorem 12.31. *Let $C \subseteq D(F)$ be convex and let $F'[x^\dagger]$ be as in Subsection 12.1.1. Further assume $N_{D(\Omega)}(x^\dagger) \cap (-N_C(x^\dagger)) = \{0\}$ and that at least one of the sets $D(\Omega)$ or C has interior points. If there are $\xi^\dagger \in \partial\Omega(x^\dagger)$, $\beta \in (0,1]$, and $c \ge 0$ such that*

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + c\|F(x) - F(x^\dagger)\| \quad \textit{for all } x \in C,$$

then there are $\tilde{\xi}^\dagger \in \partial\Omega(x^\dagger)$ and $\tilde{\eta}^\dagger \in Y^*$ with $\|\tilde{\eta}^\dagger\| \le c$ such that $F'[x^\dagger]^*\tilde{\eta}^\dagger - \tilde{\xi}^\dagger \in N_C(x^\dagger)$.

Proof. From Proposition 12.25 we obtain the variational inequality

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + c\|F'[x^\dagger](x - x^\dagger)\| \quad \text{for all } x \in C.$$

Thus, the assertion follows from Lemma 12.30 with $A = F'[x^\dagger]$. □

12.4. Variational inequalities and their approximate variant

In this section we establish a strong connection between variational inequalities and approximate variational inequalities. In fact we show that the two concepts are equivalent in a certain sense.

Given $\beta \in (0,1]$, $c \ge 0$, $M \subseteq D(F)$, and $\xi^\dagger \in \partial\Omega(x^\dagger)$ we already mentioned in Subsection 12.1.5 that the corresponding variational inequality with $\varphi(t) = ct$ is satisfied if and only if the distance function D_β associated with β, M, and $\xi^\dagger$ becomes zero at some point $r_0 \ge 0$. By Proposition 12.11 variational inequalities with φ being almost linear near zero can be reduced to variational inequalities with linear φ. Thus, it only remains to clarify the connection between variational inequalities for which

$\lim_{t\to+0} \frac{\varphi(t)}{t} = \infty$ and approximate variational inequalities with a distance function D_β satisfying $D_\beta > 0$ on $[0,\infty)$. Nonetheless the results below also cover the trivial situation already discussed in Subsection 12.1.5.

The relation between variational inequalities and their approximate variant will be formulated using conjugate functions (see Definition B.4). Thus it is sensible to work with functions defined on the whole real line. To this end we extend the concave functions φ to $\mathbb{R}$ by setting $\varphi(t) := -\infty$ for $t < 0$ and the convex functions D_β are extended by $D_\beta(r) := +\infty$ for $r < 0$. The conjugate functions of $-\varphi$ and D_β will be denoted by $(-\varphi)^*$ and D_β^*.

At first we show how to obtain a variational inequality from an approximate variational inequality. In a second lemma an upper bound for D_β given a variational inequality is derived. And finally we combine and interpret the two results in form of a theorem.

Lemma 12.32. *Let $x^\dagger$ satisfy an approximate variational inequality with $\beta \in (0,1]$, $M \subseteq D(F)$, and $\xi^\dagger \in \partial\Omega(x^\dagger)$. Assume $D_\beta(0) > 0$ for the corresponding distance function D_β. Then $-D_\beta^*(-\bullet)$ satisfies the assumptions imposed on φ in Definition 12.12 and $x^\dagger$ satisfies a variational inequality* (12.4) *with β, M, $\xi^\dagger$, and $\varphi = -D_\beta^*(-\bullet)$.*

Proof. Fix $x \in M$ and observe

$$\begin{aligned}
&\beta B_{\xi^\dagger}^\Omega(x,x^\dagger) + \Omega(x^\dagger) - \Omega(x) \\
&\quad = \beta B_{\xi^\dagger}^\Omega(x,x^\dagger) + \Omega(x^\dagger) - \Omega(x) - r\|F(x) - F(x^\dagger)\| + r\|F(x) - F(x^\dagger)\| \\
&\quad \le D_\beta(r) + r\|F(x) - F(x^\dagger)\|
\end{aligned}$$

for all $r \ge 0$. Since $D_\beta(r) = +\infty$ for $r < 0$ the inequality holds for all $r \in \mathbb{R}$. Passing to the infimum over $r \in \mathbb{R}$ we obtain

$$\begin{aligned}
&\beta B_{\xi^\dagger}^\Omega(x,x^\dagger) + \Omega(x^\dagger) - \Omega(x) \\
&\quad \le \inf_{r\in\mathbb{R}} \big(D_\beta(r) + r\|F(x) - F(x^\dagger)\|\big) = -\sup_{r\in\mathbb{R}} \big(-\|F(x) - F(x^\dagger)\|r - D_\beta(r)\big) \\
&\quad = -D_\beta^*(-\|F(x) - F(x^\dagger)\|)
\end{aligned}$$

and therefore

$$\beta B_{\xi^\dagger}^\Omega(x,x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + (-D_\beta^*)(-\|F(x) - F(x^\dagger)\|) \quad \text{for all } x \in M.$$

It remains to show that $-D_\beta^*(-\bullet)$ satisfies the assumptions imposed on φ in Definition 12.12. Conjugate functions are convex and lower semi-continuous. Thus, $-D_\beta^*(-\bullet)$ is concave and upper semi-continuous. From

$$-D_\beta^*(-t) = \inf_{r\ge0} \big(D_\beta(r) + rt\big) \quad \text{for all } t \ge 0$$

we see $-D_\beta^*(-t) \in [0,\infty)$ for $t \ge 0$ and $-D_\beta^*(-0) = 0$ (by assumption $D_\beta(r) \to 0$ if $r \to \infty$). The upper semi-continuity thus implies $\lim_{t\to+0} -D_\beta^*(-t) = 0$. We also immediately see $-D_\beta^*(-\bullet) = -\infty$ on $(-\infty,0)$ and that $-D_\beta^*(-\bullet)$ is monotonically increasing.

Now assume that there is no $\varepsilon > 0$ such that $-D_\beta^*(-\bullet)$ is strictly monotonically increasing on $[0, \varepsilon]$. Concavity, monotonicity, and upper semi-continuity of $-D_\beta^*(-\bullet)$ then imply $-D_\beta^*(-\bullet) = 0$ on $[0, \infty)$ and the Fenchel–Moreau–Rockafellar theorem (see [ABM06, Theorem 9.3.2]) implies

$$D_\beta(0) = D_\beta^{**}(0) = \sup_{t \in \mathbb{R}} \big(0 \cdot (-t) - D_\beta^*(-t)\big) = \sup_{t \geq 0} \big(-D_\beta^*(-t)\big) = 0.$$

This contradicts $D_\beta(0) > 0$ and thus $-D_\beta^*(-\bullet)$ is strictly monotonically increasing near zero. □

From a variational inequality with $\varphi = -D_\beta^*(-\bullet)$ we obtain the convergence rate $\mathcal{O}\big(-D_\beta^*(-\delta)\big)$ via Proposition 12.13. This is the same rate as obtained directly from the distance function D_β via Proposition 12.18.

The assumption $D_\beta(0) > 0$ in Lemma 12.32 was already discussed in Remark 12.19.

Lemma 12.33. *Let $x^\dagger$ satisfy a variational inequality with $\beta \in (0, 1]$, $M \subseteq D(F)$, $\xi^\dagger \in \partial\Omega(x^\dagger)$, and some function φ as described in Definition 12.12. Further assume $x^\dagger \in M$. Then $x^\dagger$ satisfies an approximate variational inequality with β, M, and $\xi^\dagger$, where $D_\beta \leq (-\varphi)^*(-\bullet)$ on $[0, \infty)$ is true for the corresponding distance function.*

Proof. For each $r \geq 0$ we have

$$\begin{aligned} D_\beta(r) &= \sup_{x \in M} \big(\beta B_{\xi^\dagger}^\Omega(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x) - r\|F(x) - F(x^\dagger)\|\big) \\ &\leq \sup_{x \in M} \big(\varphi(\|F(x) - F(x^\dagger)\|) - r\|F(x) - F(x^\dagger)\|\big) \\ &\leq \sup_{t \geq 0} \big(\varphi(t) - rt\big) = \sup_{t \in \mathbb{R}} \big(-rt - (-\varphi)(t)\big) = (-\varphi)^*(-r), \end{aligned}$$

where we used $\varphi(t) = -\infty$ for $t < 0$.

We now show $(-\varphi)^*(-r) \to 0$ if $r \to \infty$. At first we consider the case $c := \lim_{t \to +0} \frac{\varphi(t)}{t} < \infty$. Then by concavity $\varphi(t) \leq ct$ for all $t \geq 0$. For $r \geq c$ we thus obtain

$$(-\varphi)^*(-r) = \sup_{t \geq 0} \big(\varphi(t) - rt\big) \leq \sup_{t \geq 0} \big(\varphi(t) - ct\big) = 0.$$

If on the other hand $\lim_{t \to +0} \frac{\varphi(t)}{t} = \infty$ then for each sufficiently large $r > 0$ there is a uniquely determined $t_r > 0$ such that $\frac{\varphi(t_r)}{t_r} = r$. In addition, $t_r \to 0$ if $r \to \infty$. Therefore $\varphi(t) - rt \leq \varphi(t_r) - rt \leq \varphi(t_r)$ for $t \in [0, t_r]$ and $\varphi(t) - rt = t\big(\frac{\varphi(t)}{t} - r\big) \leq t\big(\frac{\varphi(t_r)}{t_r} - r\big) = 0$ for $t \geq t_r$, yielding

$$(-\varphi)^*(-r) = \sup_{t \geq 0} \big(\varphi(t) - rt\big) \leq \sup_{t \in [0, t_r]} \big(\varphi(t) - rt\big) + \sup_{t \geq t_r} \big(\varphi(t) - rt\big) \leq \varphi(t_r).$$

Since $t_r \to 0$ if $r \to \infty$ we obtain $(-\varphi)^*(-r) \to 0$ if $r \to \infty$ and thus also $D_\beta(r) \to 0$ if $r \to \infty$. □

As for the previous lemma we check that the assertion of the lemma is consistent with the convergence rates results based on variational inequalities and approximate

variational inequalities. More precisely we have to show that with the estimate $D_\beta \leq (-\varphi)^*(-\bullet)$ Proposition 12.18 yields the convergence rate $\mathcal{O}(\varphi(\delta))$, which one directly obtains from the variational inequality with φ using Proposition 12.13. Obviously it suffices to show $-D_\beta^*(-\bullet) \leq \varphi$. From the definition of conjugate functions we know

$$-D_\beta^*(-\delta) = \inf_{r\geq 0}\big(\delta r + D_\beta(r)\big) \quad \text{and} \quad (-\varphi)^*(-r) = \sup_{t\geq 0}\big(\varphi(t) - rt\big).$$

Thus,

$$-D_\beta^*(-\delta) = \inf_{r\geq 0}\big(\delta r + D_\beta(r)\big) \leq \inf_{r\geq 0}\Big(\delta r + \sup_{t\geq 0}\big(\varphi(t) - rt\big)\Big) = \inf_{r\geq 0}\sup_{t\geq 0}\big((\delta - t)r + \varphi(t)\big).$$

By [ABM06, Theorem 9.7.1] we may interchange inf and sup, which yields the desired inequality:

$$-D_\beta^*(-\delta) \leq \sup_{t\geq 0}\Big(\varphi(t) + \inf_{r\geq 0}\big((\delta - t)r\big)\Big) = \sup_{t\in[0,\delta]} \varphi(t) = \varphi(\delta).$$

Theorem 12.34. *The exact solution $x^\dagger$ satisfies a variational inequality with $\beta \in (0, 1]$, $x^\dagger \in M \subseteq D(F)$, $\xi^\dagger \in \partial\Omega(x^\dagger)$, and some function φ as described in Definition 12.12 if and only if it satisfies an approximate variational inequality with the same components β, M, $\xi^\dagger$ and with $D_\beta(0) > 0$ for the associated distance function.*

In this case

$$D_\beta = \min_{\varphi\in\Phi} (-\varphi)^*(-\bullet) \quad \textit{(pointwise minimum)}, \tag{12.18}$$

where $\Phi \neq \emptyset$ denotes the set of all functions φ with properties as described in Definition 12.12 for which $x^\dagger$ satisfies a variational inequality with β, M, and $\xi^\dagger$. The minimum is attained for $\varphi = -D_\beta^*(-\bullet) \in \Phi$.

Proof. The assertion summarizes Lemma 12.32 and Lemma 12.33. We briefly discuss the equality (12.18). From Lemma 12.33 we know

$$D_\beta \leq \inf_{\varphi\in\Phi} (-\varphi)^*(-\bullet) \quad \text{(pointwise infimum)}$$

and Lemma 12.32 provides $\bar\varphi := -D_\beta^*(-\bullet) \in \Phi$. Because

$$\begin{aligned}(-\bar\varphi)^*(-r) &= \big(D_\beta^*(-\bullet)\big)^*(-r) = \sup_{t\in\mathbb{R}}\big(-rt - D_\beta^*(-t)\big) = \sup_{t\in\mathbb{R}}\big(rt - D_\beta^*(t)\big)\\ &= D_\beta^{**}(r) = D_\beta,\end{aligned}$$

the infimum is attained at $\bar\varphi$ and equals D_β. □

Relation (12.18) describes a one-to-one correspondence between the distance function D_β and the set Φ of functions φ for which a variational inequality holds. Thus, the concept of variational inequalities provides exactly the same amount of 'smoothness information' as the concept of approximate variational inequalities. But both concepts have their right to exist: variational inequalities are not too abstract but hard to analyze and approximate variational inequalities are more abstract but can be analyzed with methods from convex analysis. The accessibility of approximate variational inequalities through methods of convex analysis is a great advantage, as we will see in the next section.

12.5. Where to place approximate source conditions?

In the preceding two sections we completely revealed the connections between (projected) source conditions, variational inequalities, and approximate variational inequalities in Banach spaces. It remains to place the concept of approximate source conditions somewhere in the picture.

We start with a result from [Gei09, FH10] stating that approximate source conditions yield approximate variational inequalities in reflexive Banach spaces, but we give a proof also covering non-reflexive Banach spaces. A similar result (in reflexive Banach spaces) was shown in [BH10] where instead of an approximate variational inequality a variational inequality is derived from an approximate source condition. To avoid some technicalities we formulate the result for linear operators $A = F$. The case of nonlinear operators is discussed in [BH10].

Proposition 12.35. *Let $A := F$ be bounded and linear and let $x^\dagger$ satisfy an approximate source condition with $\xi^\dagger \in \partial\Omega(x^\dagger)$ and distance function d. Further assume that there are a set $M \subseteq X$ containing $x^\dagger$ and constants $q > 1$ and $c \geq 0$ such that*

$$\tfrac{1}{q}\|x - x^\dagger\|^q \leq cB_{\xi^\dagger}^{\Omega}(x, x^\dagger) \quad \text{for all } x \in M. \tag{12.19}$$

Then for each $\beta \in (0,1)$ the exact solution $x^\dagger$ satisfies an approximate variational inequality with β, M, and $\xi^\dagger$. The distance function D_β fulfills

$$D_\beta(r) \leq \tfrac{q-1}{q}\big(\tfrac{c}{1-\beta}\big)^{\frac{1}{q-1}} d(r)^{\frac{q}{q-1}} \quad \text{for all } r \geq 0. \tag{12.20}$$

Proof. For fixed $x \in M$ and $r \geq 0$ and for all $\eta \in Y^*$ with $\|\eta\| \leq r$ we have

$$\begin{aligned}
&\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x) - r\|A(x - x^\dagger)\| \\
&\quad = -(1-\beta)B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \langle A^*\eta - \xi^\dagger, x - x^\dagger\rangle + \langle -\eta, A(x - x^\dagger)\rangle - r\|A(x - x^\dagger)\| \\
&\quad \leq -(1-\beta)B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \|A^*\eta - \xi^\dagger\|\|x - x^\dagger\|.
\end{aligned}$$

Passing to the infimum over all $\eta \in Y^*$ with $\|\eta\| \leq r$ and applying (12.19) yields

$$\begin{aligned}
&\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x) - r\|A(x - x^\dagger)\| \\
&\quad \leq d(r)\|x - x^\dagger\| - (1-\beta)B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \\
&\quad \leq d(r)\big(qcB_{\xi^\dagger}^{\Omega}(x, x^\dagger)\big)^{\frac{1}{q}} - (1-\beta)B_{\xi^\dagger}^{\Omega}(x, x^\dagger).
\end{aligned}$$

Now we apply Young's inequality

$$ab \leq \tfrac{q-1}{q}a^{\frac{q}{q-1}} + \tfrac{1}{q}b^q, \quad a, b \geq 0,$$

with $a := \big(\frac{c}{1-\beta}\big)^{\frac{1}{q}} d(r)$ and $b := \big(q(1-\beta)B_{\xi^\dagger}^{\Omega}(x, x^\dagger)\big)^{\frac{1}{q}}$ and obtain

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \Omega(x^\dagger) - \Omega(x) - r\|A(x - x^\dagger)\| \leq \tfrac{q-1}{q}\big(\tfrac{c}{1-\beta}\big)^{\frac{1}{q-1}} d(r)^{\frac{q}{q-1}}$$

for all $x \in M$ and all $r \geq 0$. The assertion thus follows by taking the supremum over $x \in M$ on the left-hand side. □

A Bregman distance $B_{\xi^\dagger}^{\Omega}(\bullet, x^\dagger)$ which satisfies (12.19) with $M = X$ is called *q-coercive*. The proposition provides an estimate for the distance function D_β. From this estimate we obtain a convergence rate with the help of Proposition 12.18 and this rate is the same as the one obtained in [HH09, Theorem 4.6] directly from the distance function d and assuming q-coercivity of the Bregman distance. Precisely, setting $\Phi(r) := \frac{1}{r} d(r)^{\frac{q}{q-1}}$ for $r \geq 0$, [HH09, Theorem 4.6] and Proposition 12.18 provide the rates

$$\mathcal{O}\Big(d\big(\Phi^{-1}(\delta)\big)^{\frac{q}{q-1}}\Big) \quad \text{and} \quad \mathcal{O}\Big(d\big(\Phi^{-1}(\tfrac{1}{a}\delta)\big)^{\frac{q}{q-1}}\Big),$$

respectively, where $a := \frac{q-1}{q}\big(\frac{c}{1-\beta}\big)^{\frac{1}{q-1}}$ is the constant from Proposition 12.35. These two $\mathcal{O}$-expressions describe the same rate since

$$\min\{1, a\} d\big(\Phi^{-1}(\tfrac{1}{a}\delta)\big)^{\frac{q}{q-1}} \leq d\big(\Phi^{-1}(\delta)\big)^{\frac{q}{q-1}} \leq \max\{1, a\} d\big(\Phi^{-1}(\tfrac{1}{a}\delta)\big)^{\frac{q}{q-1}},$$

as we show now: By the definition of Φ we have $d\big(\Phi^{-1}(\delta)\big)^{\frac{q}{q-1}} = \delta\Phi^{-1}(\delta)$. In case $a \geq 1$ the monotonicity of $d\big(\Phi^{-1}(\bullet)\big)^{\frac{q}{q-1}}$ and of Φ^{-1} implies

$$d\big(\Phi^{-1}(\tfrac{1}{a}\delta)\big)^{\frac{q}{q-1}} \leq d\big(\Phi^{-1}(\delta)\big)^{\frac{q}{q-1}} = \delta\Phi^{-1}(\delta) \leq \delta\Phi^{-1}(\tfrac{1}{a}\delta) = a d\big(\Phi^{-1}(\tfrac{1}{a}\delta)\big)^{\frac{q}{q-1}}.$$

For $a \geq 0$ a similar reasoning applies.

Thus, when working with approximate variational inequalities instead of approximate source conditions we do not lose anything.

Note that Proposition 12.35 requires $\beta < 1$. If $\beta \to 1$ then the constant in (12.20) goes to infinity. This observation suggests that the proposition does not hold for $\beta = 1$. After formulating a technical lemma we give a connection between D_β and d in case $\beta = 1$.

Lemma 12.36. *Let $A := F$ be bounded and linear and let $\beta \in (0, 1]$, $M \subseteq X$, and $\xi^\dagger \in \partial\Omega(x^\dagger)$. Further assume that M is convex and that $x^\dagger \in M$ and let D_β be defined by* (12.5)*. Then*

$$D_\beta(r) = \inf\big\{h(\eta) : \eta \in Y^*,\ \|\eta\| \leq r\big\} \quad \textit{for all } r \geq 0$$

with

$$h(\eta) = (1-\beta)\Omega(x^\dagger) - \langle A^*\eta - \beta\xi^\dagger, x^\dagger\rangle + \sup_{x \in M}\big(\langle A^*\eta - \beta\xi^\dagger, x\rangle - (1-\beta)\Omega(x)\big).$$

Proof. Defining two functions $f : X \to (-\infty, \infty]$ and $g_r : Y \to [0, \infty]$ by

$$f(x) := \Omega(x) - \Omega(x^\dagger) - \beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) + \delta_M(x) \quad \text{and} \quad g_r(y) := r\|y - Ax^\dagger\|,$$

where δ_M denotes the indicator function of the set M (see Definition B.5), we may write

$$D_\beta(r) = \sup_{x \in X}\big(-f(x) - g_r(Ax)\big) = -\inf_{x \in X}\big(f(x) + g_r(Ax)\big)$$

for all $r \geq 0$. From [BGW09, Theorem 3.2.4] we thus obtain

$$D_\beta(r) = -\sup_{\eta \in Y^*}\big(-f^*(-A^*\eta) - g_r^*(\eta)\big) = \inf_{\eta \in Y^*}\big(f^*(-A^*\eta) + g_r^*(\eta)\big)$$

for all $r \geq 0$ (the applicability of the cited theorem can be easily verified).

The conjugate function g_r^* of g_r is

$$\begin{aligned} g_r^*(\eta) &= \sup_{y \in Y} \big(\langle \eta, y \rangle - r\|y - Ax^\dagger\| \big) = \langle \eta, Ax^\dagger \rangle + \sup_{y \in Y} \big(\langle \eta, y - Ax^\dagger \rangle - r\|y - Ax^\dagger\| \big) \\ &= \langle \eta, Ax^\dagger \rangle + \sup_{y \in Y} \big(\langle \eta, y \rangle - r\|y\| \big) = \langle \eta, Ax^\dagger \rangle + \delta_{\overline{B}_r(0)}(\eta) \end{aligned}$$

with $\delta_{\overline{B}_r(0)}$ being the indicator function of the closed r-ball in Y^* (for the last equality see [ABM06, Example 9.3.1]). Therefore

$$\begin{aligned} D_\beta(r) &= \inf \big\{ f^*(-A^*\eta) + \langle \eta, Ax^\dagger \rangle : \eta \in Y^*,\ \|\eta\| \leq r \big\} \\ &= \inf \big\{ f^*(A^*\eta) - \langle A^*\eta, x^\dagger \rangle : \eta \in Y^*,\ \|\eta\| \leq r \big\}. \end{aligned}$$

It remains to calculate the conjugate function f^* of f:

$$\begin{aligned} f^*(\xi) &= \sup_{x \in X} \big(\langle \xi, x \rangle + (1-\beta)(\Omega(x^\dagger) - \Omega(x)) - \beta \langle \xi^\dagger, x - x^\dagger \rangle - \delta_M(x) \big) \\ &= (1-\beta)\Omega(x^\dagger) + \beta \langle \xi^\dagger, x^\dagger \rangle + \sup_{x \in M} \big(\langle \xi - \beta\xi^\dagger, x \rangle - (1-\beta)\Omega(x) \big). \end{aligned}$$

□

Theorem 12.37. *Let $A := F$ be bounded and linear, set $\beta := 1$, assume that $M \subseteq X$ is convex and that $x^\dagger \in M$, and let $\xi^\dagger \in \partial\Omega(x^\dagger)$. Further, let d and $D_\beta = D_1$ be defined by* (12.3) *and* (12.5)*, respectively. If M is bounded then*

$$D_1(r) \leq \overline{c} d(r) \quad \text{for all } r \geq 0$$

with some $\overline{c} > 0$. If $x^\dagger$ is an interior point of M then

$$D_1(r) \geq \underline{c} d(r) \quad \text{for all } r \geq 0$$

with some $\underline{c} > 0$.

Proof. From Lemma 12.36 we know

$$D_1(r) = \inf \left\{ \sup_{x \in M} \langle A^*\eta - \xi^\dagger, x - x^\dagger \rangle : \eta \in Y^*,\ \|\eta\| \leq r \right\}. \tag{12.21}$$

If M is bounded, that is $\|x - x^\dagger\| \leq \overline{c}$ for all $x \in M$, then

$$\sup_{x \in M} \langle A^*\eta - \xi^\dagger, x - x^\dagger \rangle \leq \overline{c} \|A^*\eta - \xi^\dagger\|$$

and therefore $D_1 \leq \overline{c} d$. If $x^\dagger \in \operatorname{int} M$ then there is some $\underline{c} > 0$ such that $\overline{B}_{\underline{c}}(x^\dagger) \subseteq M$. Hence

$$\begin{aligned} \sup_{x \in M} \langle A^*\eta - \xi^\dagger, x - x^\dagger \rangle &\geq \sup_{x \in \overline{B}_{\underline{c}}(x^\dagger)} \langle A^*\eta - \xi^\dagger, x - x^\dagger \rangle = \underline{c} \sup_{x \in \overline{B}_1(0)} \langle A^*\eta - \xi^\dagger, x \rangle \\ &= \underline{c} \|A^*\eta - \xi^\dagger\|, \end{aligned}$$

which shows $D_1 \geq \underline{c} d$. □

Combining the results from the theorem and from Proposition 12.35 we see that the influence of β and M on the distance function D_β is crucial. In case $\beta = 1$ the distance function D_β does not depend directly on Ω, but only on the subgradient $\xi^\dagger \in \partial\Omega(x^\dagger)$. Thus, it is unlikely that this distance function contains any information about the q-coercivity of $B^\Omega_{\xi^\dagger}(\bullet, x^\dagger)$. This lack of information is reflected in the slow decay of D_1 in comparison with the decay of D_β for $\beta < 1$ (see Proposition 12.35).

Remark 12.38. Equation (12.21) suggests to regard D_1 as a distance function for an *approximate projected source condition*, which expresses how far away the subgradient $\xi^\dagger$ is from satisfying the projected source condition $A^*\eta^\dagger - \xi^\dagger \in N_M(x^\dagger)$, which is equivalent to $\langle A^*\eta - \xi^\dagger, x - x^\dagger\rangle \le 0$ for all $x \in M$. We do not follow this idea any further.

Finally we derive an alternative definition of D_β in case $\beta < 1$ which shows some similarity with the definition of the distance function d in an approximate source condition. Thus, one could think about redefining the term 'approximate source condition' by using a certain Bregman distance instead of the norm on X^* for measuring the distance between $A^*\overline{B}_r(0)$ and $\xi^\dagger$. As we will see in Section 13.2 the alternative definition of D_β and the definition of d coincide at least in Hilbert spaces up to the constant $1-\beta$ if $\Omega = \frac{1}{2}\|\bullet\|^2$ and $M = X$.

Theorem 12.39. *Let $A := F$ be bounded and linear and let $\beta \in (0,1)$, $M \subseteq X$, and $\xi^\dagger \in \partial\Omega(x^\dagger)$. Further assume that M is convex and that $x^\dagger \in M$ and let D_β be defined by* (12.5)*. Then $x^\dagger \in \partial(\Omega + \delta_M)^*(\xi^\dagger)$ and*

$$D_\beta(r) = (1-\beta)\inf\Big\{ B^{(\Omega+\delta_M)^*}_{x^\dagger}\big(\xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), \xi^\dagger\big) : \eta \in Y^*,\ \|\eta\| \le r \Big\}$$

for all $r \ge 0$, where δ_M is the indicator function of the set M (see Definition B.5) and $(\Omega + \delta_M)^* : X^* \to (-\infty, \infty]$ denotes the conjugate function of $\Omega + \delta_M$.

Proof. Since $\xi^\dagger \in \partial\Omega(x^\dagger)$ we have

$$\Omega(x) \ge \Omega(x^\dagger) + \langle \xi^\dagger, x - x^\dagger\rangle \quad \text{for all } x \in X$$

and therefore (taking into account $x^\dagger \in M$) also

$$\Omega(x) + \delta_M(x) \ge \Omega(x^\dagger) + \delta_M(x^\dagger) + \langle \xi^\dagger, x - x^\dagger\rangle \quad \text{for all } x \in X.$$

That is, $\xi^\dagger \in \partial(\Omega + \delta_M)(x^\dagger)$. The assertion $x^\dagger \in \partial(\Omega + \delta_M)^*(\xi^\dagger)$ now follows from [ABM06, Theorem 9.5.1] and thus the Bregman distance $B^{(\Omega+\delta_M)^*}_{x^\dagger}(\bullet, \xi^\dagger)$ is well-defined.

By Lemma 12.36 we have

$$D_\beta(r) = \inf\big\{ h(\eta) : \eta \in Y^*,\ \|\eta\| \le r \big\} \quad \text{for all } r \ge 0$$

with

$$h(\eta) = (1-\beta)\Omega(x^\dagger) - \langle A^*\eta - \beta\xi^\dagger, x^\dagger\rangle + \sup_{x \in M}\big(\langle A^*\eta - \beta\xi^\dagger, x\rangle - (1-\beta)\Omega(x)\big).$$

Thus, it suffices to show

$$h(\eta) = (1-\beta) B_{x^\dagger}^{(\Omega+\delta_M)^*}\big(\xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), \xi^\dagger\big).$$

First, we observe

$$\begin{aligned}
h(\eta) &= (1-\beta)\Omega(x^\dagger) - \langle A^*\eta - \beta\xi^\dagger, x^\dagger\rangle \\
&\qquad + \sup_{x\in X}\big(\langle A^*\eta - \beta\xi^\dagger, x\rangle - (1-\beta)(\Omega(x) + \delta_M(x))\big) \\
&= (1-\beta)\Big(\Omega(x^\dagger) - \langle \xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), x^\dagger\rangle \\
&\qquad + \sup_{x\in X}\big(\langle \xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), x\rangle - (\Omega + \delta_M)(x)\big)\Big) \\
&= (1-\beta)\Big((\Omega + \delta_M)(x^\dagger) - \langle \xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), x^\dagger\rangle \\
&\qquad + (\Omega + \delta_M)^*\big(\xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger)\big)\Big).
\end{aligned}$$

By [ABM06, Proposition 9.5.1] we have

$$(\Omega + \delta_M)(x^\dagger) = \langle \xi^\dagger, x^\dagger\rangle - (\Omega + \delta_M)^*(\xi^\dagger)$$

and therefore

$$\begin{aligned}
h(\eta) &= (1-\beta)\Big((\Omega + \delta_M)^*\big(\xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger)\big) \\
&\qquad - (\Omega + \delta_M)^*(\xi^\dagger) - \langle \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), x^\dagger\rangle\Big) \\
&= (1-\beta) B_{x^\dagger}^{(\Omega+\delta_M)^*}\big(\xi^\dagger + \tfrac{1}{1-\beta}(A^*\eta - \xi^\dagger), \xi^\dagger\big).
\end{aligned}$$

□

12.6. Summary and conclusions

In the previous sections we presented in detail the cross-connections between different concepts for expressing several kinds of smoothness required for proving convergence rates in Tikhonov regularization. It turned out that each concept can be expressed by a variational inequality, that is, variational inequalities are the most general smoothness concept among the ones considered here. In particular we have seen the following relations:

- Each approximate variational inequality can be written as a variational inequality and vice versa. That is, both concepts are equivalent.
- For linear operators $A = F$ each (projected) source condition can be written as a variational inequality with a linear function φ and vice versa.
- For linear operators $A = F$ each approximate source condition can be written as a variational inequality. The reverse direction is also true for a certain class of variational inequalities.

Note that passing from some smoothness assumption to a variational inequality does not change the provided convergence rate.

Since (projected) source conditions and approximate source conditions contain no information about the nonlinearity structure of F but variational inequalities do so, establishing connections between usual/projected/approximate source conditions and variational inequalities requires additional assumptions on this nonlinearity structure. In fact, the major advantage of variational inequalities is their ability to express all necessary kinds of smoothness (solution, operator, spaces) for proving convergence rates in one manageable condition.

The results of the previous sections suggest to distinguish two types of variational inequalities; the ones with linear φ and the ones with a function φ which is not linear. The former are closely related to (projected) source conditions whereas the latter correspond to approximate source conditions. This distinction will be confirmed in the subsequent chapter on smoothness assumptions in Hilbert spaces, because there we show that approximate source conditions contain more precise information about solution smoothness than usual source conditions (see Section 13.4).

Finally we should mention that the connection between approximate variational inequalities and approximate source condition is based on Fenchel duality (see the proof of Lemma 12.36). Thus we may regard variational inequalities as a dual formulation of source conditions.

13. Smoothness in Hilbert spaces

Basing on the previous chapter we now give some refinements of the smoothness concepts considered in Banach spaces and of their cross connections. If the underlying spaces X and Y are Hilbert spaces and if we consider linear operators $A := F$ then the technique of source conditions can be extended to so called *general source conditions*. This generalization on the other hand allows to extend approximate source conditions and (approximate) variational inequalities in some way. In addition we accentuate the difference between general source conditions and other smoothness concepts. This is done by proving that approximate source conditions provide optimal convergence rates under certain assumptions, whereas general source conditions (almost) never provide optimal rates.

We restrict the setting for Tikhonov regularization of Chapter 12 as follows. Let X and Y be Hilbert spaces, assume that $A := F$ is bounded and linear with $D(F) = X$, choose $p := 2$, and set $\Omega := \frac{1}{2}\|\bullet\|^2$. This results in minimizing the Tikhonov functional

$$T_\alpha^{y^\delta}(x) = \tfrac{1}{2}\|Ax - y^\delta\|^2 + \tfrac{\alpha}{2}\|x\|^2$$

over $x \in X$. The unique Ω-minimizing solution to $Ax = y^0$ is denoted by $x^\dagger$ and the unique Tikhonov regularized solutions are $x_\alpha^{y^\delta}$. There is plenty of literature on this original form of Tikhonov regularization; we only mention the standard reference [EHN96].

Note that $\partial\Omega(x^\dagger) = \{x^\dagger\}$ and that the Bregman distance $B_{x^\dagger}^\Omega(\bullet, x^\dagger)$ reduces to $\frac{1}{2}\|\bullet - x^\dagger\|^2$. In addition we have the well-known estimate

$$\|x_\alpha^{y^\delta} - x^\dagger\| \le \|x_\alpha^{y^\delta} - x_\alpha\| + \|x_\alpha - x^\dagger\| \le \frac{\delta}{2\sqrt{\alpha}} + \|x_\alpha - x^\dagger\|,$$

where $x_\alpha := x_\alpha^{y^0}$ (cf. [EHN96, equation (4.16)]). Thus, upper bounds for $\|x_\alpha - x^\dagger\|$ yield upper bounds for $\|x_\alpha^{y^\delta} - x^\dagger\|$: If $\|x_\alpha - x^\dagger\| \le f(\alpha)$ with a continuous and strictly monotonically increasing function $f : [0, \infty) \to [0, \infty)$ which satisfies $f(0) = 0$, then the a priori parameter choice $\alpha(\delta) = g^{-1}(\delta)$ with $g(t) := \sqrt{t} f(t)$ yields

$$\|x_\alpha^{y^\delta} - x^\dagger\| \le \frac{3}{2} f\big(g^{-1}(\delta)\big). \tag{13.1}$$

This observation justifies that we only consider convergence rates for $\|x_\alpha - x^\dagger\|$ as $\alpha \to 0$ in this chapter.

From time to time we use the relation

$$x_\alpha = (A^*A + \alpha I)^{-1} A^* A x^\dagger, \tag{13.2}$$

which is an immediate consequence of the first order optimality condition for $T_\alpha^{y^0}$.

Some of the ideas presented in this chapter were already published in [Fle11].

13.1. Smoothness concepts revisited

In this section we specialize the smoothness concepts introduced in Section 12.1 to the Hilbert space setting under consideration. In addition we implement the idea of general source conditions.

Since the operator $A = F$ is linear, operator smoothness is no longer an issue. We also do not dwell on projected source conditions although some ideas also apply to them.

We frequently need the following definition.

Definition 13.1. A function $f : [0, \infty) \to [0, \infty)$ is called *index function* if it is continuous and strictly monotonically increasing and if it satisfies $f(0) = 0$.

13.1.1. General source conditions

In Banach spaces we considered only the source condition $x^\dagger \in \mathcal{R}(A^*)$, which is equivalent to $x^\dagger \in \mathcal{R}\big((A^*A)^{\frac{1}{2}}\big)$. By means of spectral theory we may generalize the concept to monomials, that is $x^\dagger \in \mathcal{R}\big((A^*A)^\mu\big)$ with $\mu > 0$, and even to index functions, that is

$$x^\dagger \in \mathcal{R}\big(\vartheta(A^*A)\big) \tag{13.3}$$

with an index function ϑ. Condition (13.3) is known as *general source condition* (see, e.g., [MH08] and references therein).

Proposition 13.2. *Let $x^\dagger$ satisfy* (13.3) *with a concave index function ϑ. Then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}(\vartheta(\alpha)) \quad \text{if } \alpha \to 0.$$

Proof. A proof is given in [MP03b, proof of Theorem 2]. For completeness we repeat it here.

Using the relation (13.2) and the representation $x^\dagger = \vartheta(A^*A)w$ with $w \in X$ we have

$$\begin{aligned}\|x_\alpha - x^\dagger\| &= \big\|\big((A^*A + \alpha I)^{-1}A^*A - I\big)\vartheta(A^*A)w\big\| \le \|w\| \sup_{t\in(0,\|A^*A\|]} \left|\frac{t}{t+\alpha} - 1\right| \vartheta(t) \\ &= \|w\| \sup_{t\in(0,\|A^*A\|]} \left(\frac{\alpha}{t+\alpha}\vartheta(t) + \frac{t}{t+\alpha}\vartheta(0)\right).\end{aligned}$$

Concavity and monotonicity of ϑ further imply

$$\|x_\alpha - x^\dagger\| \le \|w\| \sup_{t\in(0,\|A^*A\|]} \vartheta\left(\frac{\alpha t}{t+\alpha}\right) \le \|w\| \sup_{t\in(0,\|A^*A\|]} \vartheta(\alpha) = \|w\|\vartheta(\alpha).$$

□

13.1.2. Approximate source conditions

We extend the concept of approximate source conditions introduced in the previous chapter on smoothness in Banach spaces to general benchmark source conditions. That

is, for a given benchmark index function ψ we define the distance function $d_\psi : [0,\infty) \to [0,\infty)$ by

$$d_\psi(r) := \min\{\|x^\dagger - \psi(A^*A)w\| : w \in X,\ \|w\| \le r\}. \tag{13.4}$$

The properties of d_ψ are the same as for the distance function d in Subsection 12.1.3.

We say that $x^\dagger$ satisfies an approximate source condition if the distance function d_ψ decays to zero at infinity. The convergence rates result for approximate source conditions in Hilbert spaces reads as follows.

Proposition 13.3. *Let $x^\dagger$ satisfy an approximate source condition with a concave benchmark index function ψ.*

- *Assume $d_\psi > 0$ on $[0,\infty)$ and define the function Φ on $[0,\infty)$ by $\Phi(r) := \frac{d_\psi(r)}{r}$. Then*
$$\|x_\alpha - x^\dagger\| = \mathcal{O}\big(d_\psi\big(\Phi^{-1}(\psi(\alpha))\big)\big) \quad \text{if } \alpha \to 0.$$
- *Without further assumptions*
$$\|x_\alpha - x^\dagger\| = \mathcal{O}\big(-d_\psi^*(-\psi(\alpha))\big) \quad \text{if } \alpha \to 0.$$

Proof. The first rate expression can be derived from [HM07, Theorem 5.5]. But for the sake of completeness we repeat the proof here.

For each $r \ge 0$ and each $w \in X$ with $\|w\| \le r$ we observe

$$\begin{aligned}\|x_\alpha - x^\dagger\| &= \big\|\big((A^*A + \alpha I)^{-1}A^*A - I\big)x^\dagger\big\| \\ &\le \big\|\big((A^*A + \alpha I)^{-1}A^*A - I\big)(x^\dagger - \psi(A^*A)w)\big\| \\ &\quad + \big\|\big((A^*A + \alpha I)^{-1}A^*A - I\big)\psi(A^*A)w\big\|.\end{aligned}$$

The second summand can be estimated by

$$\big\|\big((A^*A + \alpha I)^{-1}A^*A - I\big)\psi(A^*A)w\big\| \le \|w\|\psi(\alpha) \le r\psi(\alpha)$$

(cf. proof of Proposition 13.2) and the first summand by

$$\begin{aligned}&\big\|\big((A^*A + \alpha I)^{-1}A^*A - I\big)(x^\dagger - \psi(A^*A)w)\big\| \\ &\quad \le \|x^\dagger - \psi(A^*A)w\| \sup_{t\in(0,\|A^*A\|]} \left|\frac{t}{t+\alpha} - 1\right| \\ &\quad = \|x^\dagger - \psi(A^*A)w\| \sup_{t\in(0,\|A^*A\|]} \frac{\alpha}{t+\alpha} = \|x^\dagger - \psi(A^*A)w\|.\end{aligned}$$

Thus, $\|x_\alpha - x^\dagger\| \le \|x^\dagger - \psi(A^*A)w\| + r\psi(\alpha)$ and taking the infimum over all $w \in X$ with $\|w\| \le r$ we obtain

$$\|x_\alpha - x^\dagger\| \le d_\psi(r) + r\psi(\alpha) \quad \text{for all } r \ge 0. \tag{13.5}$$

The inverse function Φ^{-1} is well-defined on $(0,\infty)$ because d_ψ is strictly decreasing by the assumption $d_\psi > 0$. Therefore we may choose $r = \Phi^{-1}(\psi(\alpha))$ in (13.5), yielding

$$\|x_\alpha - x^\dagger\| \le d_\psi\big(\Phi^{-1}(\psi(\alpha))\big) + \psi(\alpha)\Phi^{-1}(\psi(\alpha)) = 2d_\psi\big(\Phi^{-1}(\psi(\alpha))\big).$$

The second rate expression in the proposition can be obtained from (13.5) by taking the infimum over $r \geq 0$. That is,

$$\|x_\alpha - x^\dagger\| \leq \inf_{r \geq 0} \big(d_\psi(r) + r\psi(\alpha)\big) = -d_\psi^*(-\psi(\alpha)).$$

□

Note that both $\mathcal{O}$-expressions in the proposition describe the same convergence rate. This can be proven analogously to Proposition 12.22.

As for approximate source conditions in Banach spaces, majorants of d_ψ also yield convergence rates if d_ψ is replaced by a majorant in the proposition.

13.1.3. Variational inequalities

Specializing the variational inequality (12.4) to the present setting we obtain

$$\tfrac{\beta}{2}\|x - x^\dagger\|^2 \leq \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2}\|x^\dagger\|^2 + \varphi(\|A(x - x^\dagger)\|) \quad \text{for all } x \in M.$$

A set $M \subsetneq X$ is useful for nonlinear operators F, since the nonlinearity only has to be controlled on M. Another application for $M \subsetneq X$ is constrained Tikhonov regularization as shown in Section 12.3. Here we consider only linear operators and unconstrained Tikhonov regularization. Thus, for simplicity we set $M := X$.

Noting that $\|A(x - x^\dagger)\| = \|(A^*A)^{\frac{1}{2}}(x - x^\dagger)\|$ for all $x \in X$ we may generalize variational inequalities to the form

$$\tfrac{\beta}{2}\|x - x^\dagger\|^2 \leq \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2}\|x^\dagger\|^2 + \varphi\big(\|\psi(A^*A)(x - x^\dagger)\|\big) \quad \text{for all } x \in X \tag{13.6}$$

with index functions φ and ψ. To avoid confusion we refer to φ as *modifier function* and to ψ as *benchmark function*. As in Banach spaces we assume that the modifier function φ has the properties described in Definition 12.12 and that $\beta \in (0, 1]$.

In Proposition 12.27 we have already seen that for linear modifier functions φ the constant β has no influence on the question whether a variational inequality with φ is satisfied or not (the result is also true if A is replaced by $\psi(A^*A)$ in the proposition). In addition we show in Section 13.2 that in the present Hilbert space setting $\beta \in (0, 1)$ has only negligible influence on variational inequalities with general concave φ.

The following convergence rates result covers only the benchmark function $\psi(t) = t^{\frac{1}{2}}$. But the considerations in Section 13.2 will show that also variational inequalities with other (concave) benchmark functions yield convergence rates.

Proposition 13.4. *Let $x^\dagger$ satisfy a variational inequality* (13.6) *with modifier function φ and benchmark function $\psi(t) = t^{\frac{1}{2}}$. Then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\left(\sqrt{\big(-\varphi(\sqrt{2}\bullet)\big)^*\big(-\tfrac{1}{\alpha}\big)}\right) \quad \text{if } \alpha \to 0.$$

Proof. Using the variational inequality (13.6) with $x = x_\alpha$ and exploiting that x_α is a minimizer of the Tikhonov functional with exact data y^0 we obtain

$$\begin{aligned}\tfrac{\beta}{2}\|x_\alpha - x^\dagger\|^2 &\le \tfrac{1}{2}\|x_\alpha\|^2 - \tfrac{1}{2}\|x^\dagger\|^2 + \varphi(\|A(x_\alpha - x^\dagger)\|)\\ &= \tfrac{1}{\alpha}\big(\tfrac{1}{2}\|A(x_\alpha - x^\dagger)\|^2 + \tfrac{\alpha}{2}\|x_\alpha\|^2 - \tfrac{\alpha}{2}\|x^\dagger\|^2\big)\\ &\qquad + \varphi(\|A(x_\alpha - x^\dagger)\|) - \tfrac{1}{2\alpha}\|A(x_\alpha - x^\dagger)\|^2\\ &\le \varphi(\|A(x_\alpha - x^\dagger)\|) - \tfrac{1}{2\alpha}\|A(x_\alpha - x^\dagger)\|^2\\ &\le \sup_{t\ge 0}\big(\varphi(t) - \tfrac{1}{2\alpha}t^2\big) = \sup_{t\ge 0}\big(\varphi(\sqrt{2t}) - \tfrac{1}{\alpha}t\big) = \big(-\varphi(\sqrt{2\bullet})\big)^*\big(-\tfrac{1}{\alpha}\big).\end{aligned}$$

□

Note that with Proposition 12.13 we obtain

$$\|x_\alpha^{y^\delta} - x^\dagger\| = \mathcal{O}\big(\sqrt{\varphi(\delta)}\big) \quad \text{if } \delta \to 0. \tag{13.7}$$

13.1.4. Approximate variational inequalities

Finally we specialize the concept of approximate variational inequalities to the present Hilbert space setting and extend it in the same way as done for variational inequalities.

The distance function $D_{\psi,\beta} : [0,\infty) \to [0,\infty)$ defined by

$$D_{\psi,\beta}(r) := \sup_{x\in X}\big(\tfrac{\beta}{2}\|x - x^\dagger\|^2 - \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\|\big)$$

for $\beta \in (0,1)$ has the same properties as D_β in Subsection 12.1.5 except that it does not attain the value $+\infty$. We say that the exact solution $x^\dagger$ satisfies an approximate variational inequality if $D_{\psi,\beta}$ decays to zero at infinity. Note that we exclude $\beta = 1$ since in this case the distance function $D_{\psi,1}$ decays to zero at infinity if and only if it attains the value zero at some point, which is equivalent to the source condition $x^\dagger \in \mathcal{R}(\psi(A^*A))$. The next proposition makes this observation precise.

Proposition 13.5. *The distance function $D_{\psi,1}$ satisfies $D_{\psi,1}(r) \in \{0,\infty\}$ for all $r \ge 0$.*

Proof. First observe

$$D_{\psi,1}(r) = \sup_{x\in X}\big(\langle -x^\dagger, x - x^\dagger\rangle - r\|\psi(A^*A)(x - x^\dagger)\|\big) \quad \text{for all } r \ge 0.$$

Assume that $D_{\psi,1}(r) > 0$ for some $r \ge 0$. Then there is $x \in X$ with

$$\langle -x^\dagger, x - x^\dagger\rangle - r\|\psi(A^*A)(x - x^\dagger)\| > 0.$$

For each $t \ge 0$ we thus obtain

$$\begin{aligned}D_{\psi,1}(r) &\ge \langle -x^\dagger, x^\dagger + t(x - x^\dagger) - x^\dagger\rangle - r\|\psi(A^*A)(x^\dagger + t(x - x^\dagger) - x^\dagger)\|\\ &= t\big(\langle -x^\dagger, x - x^\dagger\rangle - r\|\psi(A^*A)(x - x^\dagger)\|\big)\end{aligned}$$

and $t \to \infty$ yields $D_{\psi,1}(r) = \infty$. □

Also in Banach spaces distance functions for approximate variational inequalities with $\beta = 1$ showed different behavior in comparison to $\beta < 1$. For details see Theorem 12.37 and the discussion thereafter.

As for variational inequalities we state a convergence rates result only for the benchmark function $\psi(t) = t^{\frac{1}{2}}$. But in Section 13.2 we show that also approximate variational inequalities with other (concave) benchmark functions yield convergence rates.

For approximate variational inequalities in Banach spaces we provided two different formulations of convergence rates. We do the same for the present Hilbert space setting.

Proposition 13.6. *Let $x^\dagger$ satisfy an approximate variational inequality with $\beta \in (0,1)$ and benchmark function $\psi(t) = t^{\frac{1}{2}}$.*

- *If $D_{\psi,\beta} > 0$ on $[0,\infty)$ then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\Big(\sqrt{D_{\psi,\beta}\big(\Phi^{-1}(\sqrt{\alpha})\big)}\Big) \quad \text{if } \alpha \to 0,$$

  where $\Phi(r) := \frac{\sqrt{D_{\psi,\beta}(r)}}{r}$.

- *Without further assumptions*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\Big(\sqrt{-\big(D_{\psi,\beta}(\sqrt{2}\bullet)\big)^*(-\alpha)}\Big) \quad \text{if } \alpha \to 0.$$

Proof. Analogously to the proof of Lemma 12.17 with $\delta = 0$ and $p = 2$ we obtain

$$\tfrac{\beta}{2}\|x_\alpha - x^\dagger\|^2 \le \tfrac{1}{2}\alpha r^2 + D_{\psi,\beta}(r) \quad \text{for all } r \ge 0. \tag{13.8}$$

Setting $r_\alpha := \Phi^{-1}(\sqrt{\alpha})$ we have $\alpha r_\alpha^2 = D_{\psi,\beta}(r_\alpha)$ and thus

$$\tfrac{\beta}{2}\|x_\alpha - x^\dagger\|^2 \le \tfrac{1}{2}\alpha r_\alpha^2 + D_{\psi,\beta}(r_\alpha) = \tfrac{3}{2}D_{\psi,\beta}(r_\alpha) = \tfrac{3}{2}D_{\psi,\beta}\big(\Phi^{-1}(\sqrt{\alpha})\big).$$

The second rate expression can be derived from (13.8) by taking the infimum over $r \ge 0$. Then

$$\begin{aligned}\tfrac{\beta}{2}\|x_\alpha - x^\dagger\|^2 &\le \inf_{r\ge 0}\big(\tfrac{1}{2}\alpha r^2 + D_{\psi,\beta}(r)\big) = -\sup_{s\ge 0}\big(-\alpha s - D_{\psi,\beta}(\sqrt{2s})\big)\\ &= -\sup_{s\in\mathbb{R}}\big(-\alpha s - D_{\psi,\beta}(\sqrt{2s})\big) = -\big(D_{\psi,\beta}(\sqrt{2}\bullet)\big)^*(-\alpha),\end{aligned}$$

where we set $D_{\psi,\beta}$ to $+\infty$ on $(-\infty, 0)$. □

Properties of the function $-D_{\psi,\beta}(-\bullet)$ were discussed in Remark 12.20 and analogously to the proof of Proposition 12.22 one shows that the two rate expressions in Proposition 13.6 describe the same convergence rate.

13.2. Equivalent smoothness concepts

In the previous chapter on smoothness concepts in Banach spaces we have seen that the concepts of variational inequalities and approximate variational inequalities are equivalent, see Theorem 12.34. We restate this result for the present Hilbert space setting.

Corollary 13.7. *The exact solution $x^\dagger$ satisfies a variational inequality with $\beta \in (0,1)$, some modifier function φ, and some benchmark function ψ if and only if it satisfies an approximate variational inequality with the same constant β, with the same benchmark ψ, and with $D_{\beta,\psi}(0) > 0$ for the associated distance function.*

In this case

$$D_{\beta,\psi} = \min_{\varphi\in\Phi} (-\varphi)^*(-\bullet) \quad \textit{(pointwise minimum)},$$

where $\Phi \neq \emptyset$ denotes the set of all modifier functions φ for which $x^\dagger$ satisfies a variational inequality with β, φ, and ψ. The minimum is attained for $\varphi = -D^*_{\psi,\beta}(-\bullet)$.

Proof. The proof is the same as for Theorem 12.34 except that Y has to be replaced by X and A has to be replaced by $\psi(A^*A)$. □

The cross connections between (approximate) variational inequalities and approximate source conditions in Banach spaces were less straight, see Section 12.5. But in Hilbert spaces we obtain a satisfactory result from Theorem 12.39, which shows all-encompassing equivalence between (approximate) variational inequalities and approximate source conditions.

Corollary 13.8. *Let $\beta \in (0,1)$ and let ψ be an index function. Then the distance function $D_{\psi,\beta}$ (approximate variational inequality) and the distance function d_ψ (approximate source condition) satisfy*

$$D_{\psi,\beta}(r) = \tfrac{1}{2(1-\beta)} d^2_\psi(r) \quad \textit{for all } r \geq 0. \tag{13.9}$$

Proof. Replacing Y by X and A by $\psi(A^*A)$ in the proof of Theorem 12.39 we obtain

$$D_{\psi,\beta}(r) = (1-\beta) \inf\left\{ B^{\Omega^*}_{x^\dagger}\left(x^\dagger + \tfrac{1}{1-\beta}(\psi(A^*A)w - x^\dagger), x^\dagger\right) : w \in X,\ \|w\| \leq r \right\}$$

for all $r \geq 0$. Since $\Omega^* = \left(\frac{1}{2}\|\bullet\|^2\right)^* = \frac{1}{2}\|\bullet\|^2$ the Bregman distance $B^{\Omega^*}_{x^\dagger}(\bullet, x^\dagger)$ reduces to $\frac{1}{2}\|\bullet - x^\dagger\|^2$. Therefore

$$D_{\psi,\beta}(r) = (1-\beta) \inf\left\{ \tfrac{1}{2(1-\beta)^2} \|\psi(A^*A)w - x^\dagger\|^2 : w \in X,\ \|w\| \leq r \right\},$$

which is equivalent to $D_{\psi,\beta}(r) = \frac{1}{2(1-\beta)} d^2_\psi(r)$. □

The corollary especially shows that two distance functions D_{ψ,β_1} and D_{ψ,β_2} differ only by a constant factor:

$$D_{\psi,\beta_2} = \frac{1-\beta_1}{1-\beta_2} D_{\psi,\beta_1}. \tag{13.10}$$

Another consequence is that all convergence rates results based on approximate source conditions also apply to (approximate) variational inequalities. That is, (approximate)

variational inequalities yield convergences rates for general benchmark functions ψ and also for general linear regularization methods (cf. [HM07] for corresponding convergence rates results based on approximate source conditions).

Of course the convergence rates obtained from the three equivalent smoothness concepts should coincide if $\psi(t) = t^{\frac{1}{2}}$. Before we verify this assertion we show that the rate obtained from an approximate variational inequality via Proposition 13.6 does not depend on β. So let $\psi(t) = t^{\frac{1}{2}}$ and $\beta_1, \beta_2 \in (0,1)$. Then, taking into account (13.10), Proposition 13.6 provides the convergence rates

$$\mathcal{O}\Big(\sqrt{D_{\psi,\beta_1}\big(\Phi^{-1}(\sqrt{\alpha})\big)}\Big) \quad \text{and} \quad \mathcal{O}\left(\sqrt{D_{\psi,\beta_1}\Big(\Phi^{-1}\Big(\sqrt{\tfrac{1-\beta_2}{1-\beta_1}}\sqrt{\alpha}\Big)\Big)}\right),$$

where $\Phi^{-1}(r) := \frac{1}{r}\sqrt{D_{\psi,\beta_1}(r)}$. Analogously to the discussion subsequent to Proposition 12.35 one can show that both expressions describe the same convergence rate.

Now assume that a variational inequality with φ is satisfied. Then by Corollary 13.7 we have the estimate $D_{\psi,\beta} \le (-\varphi)^*(-\bullet)$ and Proposition 13.6 provides the convergence rate

$$\mathcal{O}\Big(\sqrt{-\big(D_{\psi,\beta}(\sqrt{2}\bullet)\big)^*(-\alpha)}\Big),$$

which can be bounded by

$$\begin{aligned}
\sqrt{-\big(D_{\beta}(\sqrt{2}\bullet)\big)^*(-\alpha)} &= \sqrt{\inf_{r\ge 0}\big(\tfrac{1}{2}\alpha r^2 + D_{\psi,\beta}(r)\big)} \le \sqrt{\inf_{r\ge 0}\Big(\tfrac{1}{2}\alpha r^2 + \sup_{t\ge 0}\big(\varphi(t) - rt\big)\Big)} \\
&= \sqrt{\sup_{t\ge 0}\Big(\varphi(t) + \inf_{r\ge 0}\big(\tfrac{1}{2}\alpha r^2 - tr\big)\Big)} = \sqrt{\sup_{t\ge 0}\big(\varphi(t) - \tfrac{1}{2\alpha}t^2\big)} \\
&= \sqrt{\big(-\varphi(\sqrt{2}\bullet)\big)^*(-\tfrac{1}{\alpha})}.
\end{aligned}$$

Interchanging inf and sup is allowed by [ABM06, Theorem 9.7.1]. Consequently, via Proposition 13.6 we obtain the same rate as directly from the variational inequality via Proposition 13.4.

If $x^\dagger$ satisfies an approximate variational inequality with distance function $D_{\psi,\beta}$, then it also satisfies a variational inequality with $\varphi = -D^*_{\psi,\beta}(-\bullet)$ by Corollary 13.7. Similar arguments as in the previous paragraph show that the rate obtained directly from the approximate variational inequality via Proposition 13.6 can also be obtained from the variational inequality via Proposition 13.4.

Eventually, Corollary 13.8 provides the identity $D_{\psi,\beta} = \frac{1}{2(1-\beta)}d_\psi^2$. Thus, setting $\Phi(r) := \frac{1}{r}d(r)$ for $r > 0$, from Proposition 13.3 and Proposition 13.6 we obtain the convergence rates

$$\mathcal{O}\big(d\big(\Phi^{-1}(\sqrt{\alpha})\big)\big) \quad \text{and} \quad \mathcal{O}\Big(d\Big(\Phi^{-1}\Big(\tfrac{1}{\sqrt{2(1-\beta)}}\sqrt{\alpha}\Big)\Big)\Big),$$

respectively. Analogously to the discussion subsequent to Proposition 12.35 one can show that both expressions describe the same convergence rate.

As a consequence of the results obtained in this section we only consider general source conditions and approximate source conditions in the remaining sections.

We close with an alternative proof of Corollary 13.8. The two assertions of the following lemma also imply the relation $D_{\psi,\beta}(r) = \frac{1}{2(1-\beta)} d_\psi^2(r)$, but they give some more information on this connection. The proof of the lemma is elementary and works without duality theory.

Lemma 13.9. *Assume that A is injective and that $x^\dagger \notin \mathcal{R}(\psi(A^*A))$. Further let $\beta \in (0,1)$, ψ be an index function, and $r \geq 0$.*

- *If $x_r \in X$ is a maximizer in the definition of $D_{\psi,\beta}(r)$, then*

$$D_{\psi,\beta}(r) = \tfrac{1-\beta}{2} \|x_r - x^\dagger\|^2.$$

- *If w_r is a minimizer in the definition of d_ψ, then*

$$x_r := x^\dagger + \tfrac{1}{1-\beta}\big(\psi(A^*A)w_r - x^\dagger\big)$$

  is a maximizer in the definition of $D_{\psi,\beta}$.

Proof. We start with the first assertion. If $x_r = x^\dagger$ then $D_{\psi,\beta}(r) = 0$ by the definition of $D_{\psi,\beta}(r)$. So assume that $x_r \neq x^\dagger$. Then the gradient of

$$x \mapsto \tfrac{\beta}{2}\|x - x^\dagger\|^2 - \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\|$$

at x_r has to be zero, that is,

$$\beta(x_r - x^\dagger) - x_r - r\frac{\psi^2(A^*A)(x_r - x^\dagger)}{\|\psi(A^*A)(x_r - x^\dagger)\|} = 0. \tag{13.11}$$

Applying $\langle \bullet, x_r - x^\dagger \rangle$ at both sides we get

$$-r\|\psi(A^*A)(x_r - x^\dagger)\| = -\beta\|x_r - x^\dagger\|^2 + \langle x_r, x_r - x^\dagger \rangle$$

and therefore

$$\begin{aligned} D_{\psi,\beta}(r) &= \tfrac{\beta}{2}\|x_r - x^\dagger\|^2 - \tfrac{1}{2}\|x_r\|^2 + \tfrac{1}{2}\|x^\dagger\|^2 - \beta\|x_r - x^\dagger\|^2 + \langle x_r, x_r - x^\dagger \rangle \\ &= \tfrac{1-\beta}{2}\|x_r - x^\dagger\|^2. \end{aligned}$$

We come to the second assertion. By the definition of w_r there exists some Lagrange multiplier $\lambda \geq 0$ with

$$\psi(A^*A)\big(\psi(A^*A)w_r - x^\dagger\big) = -\lambda w_r. \tag{13.12}$$

For $\lambda = 0$ we would get $x^\dagger = \psi(A^*A)w_r$, which contradicts $x^\dagger \notin \mathcal{R}(\psi(A^*A))$. Thus $\lambda > 0$ and therefore $\|w_r\| = r$. Defining x_r as in the lemma and using (13.12) one easily verifies (13.11), which is equivalent to the assertion that x_r is a maximizer in the definition of $D_{\psi,\beta}(r)$. □

13.3. From general source conditions to distance functions

Obviously a source condition $x^\dagger \in \mathcal{R}(\psi(A^*A))$ implies $d_\psi(r) = 0$ for all sufficiently large r, where d_ψ denotes the distance function associated with the concept of approximate source conditions.

We now briefly discuss the behavior of d_ψ if $x^\dagger \notin \mathcal{R}(\psi(A^*A))$ but $x^\dagger \in \mathcal{R}(\vartheta(A^*A))$. Note that from [MH08, HMvW09] we know that for $x^\dagger \in (\ker A)^\perp$ there is always an index function ϑ with $x^\dagger \in \mathcal{R}(\vartheta(A^*A))$. The first assertion of the following theorem was shown in [HM07, Theorem 5.9].

Theorem 13.10. *Let A be a compact and injective operator and assume that $x^\dagger = \vartheta(A^*A)w$ with $\|w\| = 1$ and $x^\dagger \notin \mathcal{R}(\psi(A^*A))$.*

- *If $\frac{\psi}{\vartheta}$ (with $\frac{\psi}{\vartheta}(0) := 0$) is an index function, then*

$$d_\psi(r) \le r\psi\left(\left(\frac{\psi}{\vartheta}\right)^{-1}\left(\frac{1}{r}\right)\right)$$

  for all sufficiently large r.

- *If $\vartheta^2 \circ \left(\psi^2\right)^{-1}$ is concave, then*

$$d_\psi(r) \le \left(-\vartheta \circ \psi^{-1}\right)^*(-r)$$

  for all $r \ge 0$.

Proof. The first assertion was shown in [HM07, proof of Theorem 5.9].

Since the second assertion requires some longish but elementary calculations we only give a rough sketch. At first we use (13.9) with $\beta := \frac{1}{2}$ and the representation $x^\dagger = \vartheta(A^*A)w$ to obtain

$$\begin{aligned} d_\psi(r)^2 = D_{\psi,\beta}(r) &= \sup_{x\in X}\left(\tfrac{1}{4}\|x - x^\dagger\|^2 - \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\|\right) \\ &= \sup_{x\in X}\left(\langle -x^\dagger, x - x^\dagger\rangle - \tfrac{1}{4}\|x - x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\|\right) \\ &\le \sup_{x\in X}\left(\|\vartheta(A^*A)(x - x^\dagger)\| - r\|\psi(A^*A)(x - x^\dagger)\| - \tfrac{1}{4}\|x - x^\dagger\|^2\right). \end{aligned}$$

Applying an interpolation inequality (see, e.g., [MP03a, Theorem 4]), which requires concavity of $\vartheta^2 \circ \left(\psi^2\right)^{-1}$, we further estimate

$$\begin{aligned} d_\psi(r)^2 \le \sup_{x\in X\setminus\{x^\dagger\}}\Bigg(&\|\vartheta(A^*A)(x - x^\dagger)\| \\ &- r\|x - x^\dagger\|\left(\psi \circ \vartheta^{-1}\right)\left(\frac{\|\vartheta(A^*A)(x - x^\dagger)\|}{\|x - x^\dagger\|}\right) - \tfrac{1}{4}\|x - x^\dagger\|^2\Bigg). \end{aligned}$$

Thus,

$$d_\psi(r)^2 \le \sup_{s>0,t>0}\left(t - rs\left(\psi \circ \vartheta^{-1}\right)\left(\frac{t}{s}\right) - \frac{1}{4}s^2\right).$$

If we use the fact $\frac{\psi(t)}{\vartheta(t)} \to 0$ if $t \to 0$ (see [HM07, Lemma 5.8]) and if we carry out several elementary calculations, then we see

$$\sup_{s>0,t>0} \left(t - rs\big(\psi \circ \vartheta^{-1}\big)\left(\frac{t}{s}\right) - \frac{1}{4}s^2 \right) = \big(-\vartheta \circ \psi^{-1}\big)^*(-r),$$

which completes the proof. □

Using the estimates for d_ψ given in the theorem one easily shows that the two convergence rate expressions in Proposition 13.3 yield the convergence rate $\mathcal{O}(\vartheta(\alpha))$. This is exactly the rate we also obtain directly from the source condition $x^\dagger \in \mathcal{R}(\vartheta(A^*A))$ via Proposition 13.2. In other words, passing from source conditions to approximate source conditions with high benchmark (higher than the satisfied source condition) we do not lose convergence rates.

13.4. Lower bounds for the regularization error

In [MH08, HMvW09] it was shown that if A is injective then for each element $w \in X$ there is an index function $\tilde{\vartheta}$ and some $v \in X$ such that $w = \tilde{\vartheta}(A^*A)v$. Consequently, if $x^\dagger = \vartheta(A^*A)w$ with an index function ϑ and $w \in X$, then there are $\tilde{\vartheta}$ and $v \in X$ such that $x^\dagger = (\vartheta\tilde{\vartheta})(A^*A)v$. Since $(\vartheta\tilde{\vartheta})(t) = \vartheta(t)\tilde{\vartheta}(t)$ goes faster to zero than $\vartheta(t)$ if $t \to 0$, the convergence rate $\|x_\alpha - x^\dagger\| = \mathcal{O}(\vartheta(\alpha))$ obtained from $x^\dagger \in \mathcal{R}(\vartheta(A^*A))$ cannot be optimal, at least as long as $\vartheta\tilde{\vartheta}$ is concave. In other words, convergence rates based on general source conditions can always be improved somewhat.

In the present section we show that under suitable assumptions approximate source conditions yield optimal rates. Here 'optimal' means, that also lower bounds for the regularization error $\|x_\alpha - x^\dagger\|$ in terms of distance functions can be shown and that these lower bounds coincide up to a constant with the upper bound from Proposition 13.3 (the first of the two rate expressions there).

The results of this section are joint work with Bernd Hofmann (Chemnitz) and Peter Mathé (Berlin) and have been published in [FHM11]. Here we only consider Tikhonov regularization, but in [FHM11] it is shown that lower bounds for the regularization error in terms of distance functions are also available for more general linear regularization methods.

Theorem 13.11. *Let $x^\dagger$ satisfy an approximate source condition with benchmark function ψ such that $d_\psi > 0$ on $[0,\infty)$, and define Φ by $\Phi(r) := \frac{d_\psi(r)}{r}$ on $(0,\infty)$.*

- *If $t \mapsto \frac{\psi(t)}{\sqrt{t}}$ is monotonically increasing on $(0,\infty)$, then*

$$\tfrac{1}{2} d_\psi\big(\tfrac{3}{2}\Phi^{-1}(\psi(\alpha))\big) \le \|x_\alpha - x^\dagger\| \le 2 d_\psi\big(\Phi^{-1}(\psi(\alpha))\big)$$

  for all $\alpha > 0$.

- *If $t \mapsto \frac{\psi(t)}{\sqrt{t}}$ is monotonically decreasing on $(0,\infty)$, then*

$$d_\psi\big(2\Phi^{-1}(\psi(\alpha))\big) \le \|x_\alpha - x^\dagger\| \le 2 d_\psi\big(\Phi^{-1}(\psi(\alpha))\big)$$

  for all $\alpha > 0$.

Proof. We start with the case that $t \mapsto \frac{\psi(t)}{\sqrt{t}}$ is monotonically increasing. Consider the element $v_\alpha := \big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}\psi(A^*A)x^\dagger$ for $\alpha > 0$. By the definitions of d_ψ and v_α and by (13.2) we have

$$\begin{aligned}
d_\psi(\|v_\alpha\|) &\le \|x^\dagger - \psi(A^*A)v_\alpha\| \\
&= \big\|\big(I - \big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}\psi^2(A^*A)\big)x^\dagger\big\| \\
&= \psi^2(\alpha)\big\|\big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}x^\dagger\big\| \\
&= \frac{\psi^2(\alpha)}{\alpha}\big\|\big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}(A^*A + \alpha I)\big(I - (A^*A + \alpha I)^{-1}A^*A\big)x^\dagger\big\| \\
&\le \frac{\psi^2(\alpha)}{\alpha}\left(\sup_{t\in(0,\|A^*A\|]}\frac{t+\alpha}{\psi^2(t)+\psi^2(\alpha)}\right)\|x_\alpha - x^\dagger\| \\
&= \left(\sup_{t\in(0,\|A^*A\|]}\frac{\frac{t}{\alpha}+1}{\frac{\psi^2(t)}{\psi^2(\alpha)}+1}\right)\|x_\alpha - x^\dagger\|.
\end{aligned}$$

For $t \le \alpha$ we immediately see

$$\frac{\frac{t}{\alpha}+1}{\frac{\psi^2(t)}{\psi^2(\alpha)}+1} \le \frac{1+1}{0+1} = 2.$$

If $t \ge \alpha$, then $\frac{\psi^2(t)}{t} \ge \frac{\psi^2(\alpha)}{\alpha}$ or equivalently $\frac{\psi^2(t)}{\psi^2(\alpha)} \ge \frac{t}{\alpha}$. The last inequality yields

$$\frac{\frac{t}{\alpha}+1}{\frac{\psi^2(t)}{\psi^2(\alpha)}+1} \le \frac{\frac{t}{\alpha}+1}{\frac{t}{\alpha}+1} = 1 \le 2.$$

Thus, $d_\psi(\|v_\alpha\|) \le 2\|x_\alpha - x^\dagger\|$ for all $\alpha > 0$.

We now derive an upper bound for $\|v_\alpha\|$, which by the monotonicity of d_ψ yields a lower bound for $d_\psi(\|v_\alpha\|)$. For each $r \ge 0$ and each $w \in X$ with $\|w\| \le r$ we have

$$\begin{aligned}
\|v_\alpha\| &= \big\|\big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}\psi(A^*A)x^\dagger\big\| \\
&\le \big\|\big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}\psi(A^*A)(x^\dagger - \psi(A^*A)w)\big\| \\
&\qquad + \big\|\big(\psi^2(A^*A) + \psi^2(\alpha)I\big)^{-1}\psi^2(A^*A)w\big\| \\
&\le \left(\sup_{t\in(0,\|A^*A\|]}\frac{\psi(t)}{\psi^2(t)+\psi^2(\alpha)}\right)\|x^\dagger - \psi(A^*A)w\| \\
&\qquad + \left(\sup_{t\in(0,\|A^*A\|]}\frac{\psi^2(t)}{\psi^2(t)+\psi^2(\alpha)}\right)\|w\| \\
&\le \left(\sup_{t\in(0,\|A^*A\|]}\frac{\psi(t)}{\psi^2(t)+\psi^2(\alpha)}\right)\|x^\dagger - \psi(A^*A)w\| + r
\end{aligned}$$

and thus

$$\|v_\alpha\| \le \left(\sup_{t\in(0,\|A^*A\|]}\frac{\psi(t)}{\psi^2(t)+\psi^2(\alpha)}\right)d_\psi(r) + r.$$

Further

$$\frac{\psi(t)}{\psi^2(t)+\psi^2(\alpha)} = \frac{\sqrt{\psi^2(t)}\sqrt{\psi^2(\alpha)}}{\psi^2(t)+\psi^2(\alpha)}\frac{1}{\psi(\alpha)} \le \frac{1}{2}\frac{1}{\psi(\alpha)}$$

and therefore

$$\|v_\alpha\| \le \frac{d_\psi(r)}{2\psi(\alpha)} + r \quad \text{for all } r \ge 0,\ \alpha > 0.$$

Choosing $r = \Phi^{-1}(\psi(\alpha))$ we obtain $\|v_\alpha\| \le \frac{3}{2}\Phi^{-1}(\psi(\alpha))$, yielding

$$d_\psi\big(\tfrac{3}{2}\Phi^{-1}(\psi(\alpha))\big) \le d_\psi(\|v_\alpha\|) \le 2\|x_\alpha - x^\dagger\|$$

for all $\alpha > 0$.

The estimate $\|x_\alpha - x^\dagger\| \le 2d_\psi\big(\Phi^{-1}(\psi(\alpha))\big)$ was derived in the proof of Proposition 13.3.

Now we come to the case that $t \mapsto \frac{\psi(t)}{\sqrt{t}}$ is monotonically decreasing. Consider the element $w_\alpha := \psi(A^*A)^{-1}A^*A\big(A^*A+\alpha I\big)^{-1}x^\dagger$ for $\alpha > 0$, which is well-defined because

$$\sup_{t\in(0,\|A^*A\|]} \frac{t}{(t+\alpha)\psi(t)} \le \frac{1}{\psi(\alpha)} < \infty$$

as we show now. Indeed, for $t \le \alpha$ we have

$$\frac{t}{(t+\alpha)\psi(t)} = \frac{\sqrt{t}}{\psi(t)}\frac{\sqrt{t}\sqrt{\alpha}}{t+\alpha}\frac{1}{\sqrt{\alpha}} \le \frac{\sqrt{\alpha}}{\psi(\alpha)}\frac{1}{2}\frac{1}{\sqrt{\alpha}} \le \frac{1}{\psi(\alpha)}$$

by the monotonicity of $t \mapsto \frac{\psi(t)}{\sqrt{t}}$ and for $t \ge \alpha$ we have

$$\frac{t}{(t+\alpha)\psi(t)} = \frac{t}{t+\alpha}\frac{1}{\psi(t)} \le 1 \cdot \frac{1}{\psi(\alpha)}$$

by the monotonicity of ψ. Using the definitions of d_ψ and w_α and taking into account (13.2) we obtain

$$d_\psi(\|w_\alpha\|) \le \|x^\dagger - \psi(A^*A)w_\alpha\| = \big\|x^\dagger - A^*A(A^*A-\alpha I)^{-1}x^\dagger\big\| = \|x_\alpha - x^\dagger\|.$$

For each $r \ge 0$ and each $w \in X$ with $\|w\| \le r$ an upper bound for $\|w_\alpha\|$ is given by

$$\begin{aligned}
\|w_\alpha\| &= \big\|\psi(A^*A)^{-1}A^*A\big(A^*A+\alpha I\big)^{-1}x^\dagger\big\| \\
&\le \big\|\psi(A^*A)^{-1}A^*A\big(A^*A+\alpha I\big)^{-1}(x^\dagger - \psi(A^*A)w)\big\| + \big\|A^*A\big(A^*A+\alpha I\big)^{-1}w\big\| \\
&\le \left(\sup_{t\in(0,\|A^*A\|]} \frac{t}{(t+\alpha)\psi(t)}\right)\|x^\dagger - \psi(A^*A)w\| + \left(\sup_{t\in(0,\|A^*A\|]} \frac{t}{t+\alpha}\right)\|w\| \\
&\le \frac{1}{\psi(\alpha)}\|x^\dagger - \psi(A^*A)w\| + r.
\end{aligned}$$

Thus, $\|w_\alpha\| \le \frac{d_\psi(r)}{\psi(\alpha)} + r$ for all $r \ge 0$ and all $\alpha > 0$. If we choose $r = \Phi^{-1}(\psi(\alpha))$, then $\|w_\alpha\| \le 2\Phi^{-1}(\psi(\alpha))$ and therefore

$$d_\psi\big(2\Phi^{-1}(\psi(\alpha))\big) \le d_\psi(\|w_\alpha\|) \le \|x_\alpha - x^\dagger\|.$$

□

If the distance function d_ψ decays not too fast, that is, if there is a constant $\tilde{c} > 0$ such that

$$\frac{d_\psi(cr)}{d_\psi(r)} \geq \tilde{c} \quad \text{for all } r \geq r_0 \tag{13.13}$$

with $r_0 > 0$ and $c \in \{\frac{3}{2}, 2\}$, then the theorem yields

$$\bar{c} d_\psi\big(\Phi^{-1}(\psi(\alpha))\big) \leq \|x_\alpha - x^\dagger\| \leq 2 d_\psi\big(\Phi^{-1}(\psi(\alpha))\big)$$

for sufficiently small $\alpha > 0$. The constant $\bar{c}$ either equals $\tilde{c}$ or $\frac{1}{2}\tilde{c}$ depending on the monotonicity of $t \mapsto \frac{\psi(t)}{\sqrt{t}}$. In other words, the behavior of the distance function d_ψ at infinity completely determines the behavior of the regularization error $\|x_\alpha - x^\dagger\|$ for small α. Thus the concept of approximate source conditions is superior to general source conditions, since the latter in general do not provide optimal convergence rates as discussed in the first paragraph of the present section.

A distance function d_ψ satisfies condition (13.13) for instance if

$$c_1 r^{-a} \leq d_\psi(r) \leq c_2 r^{-a} \quad \text{for all } r \geq r_0 > 0$$

with $c_1, c_2, a > 0$ or if

$$c_1 (\ln r)^{-a} \leq d_\psi(r) \leq c_2 (\ln r)^{-a} \quad \text{for all } r \geq r_0 > 1$$

with $c_1, c_2, a > 0$. Condition (13.13) is not satisfied if

$$c_1 \exp(-r) \leq d_\psi(r) \leq c_2 \exp(-r) \quad \text{for all } r \geq r_0 \geq 0$$

with $c_1, c_2 > 0$, which represents a very fast decay of d_ψ at infinity.

The more general version of Theorem 13.11 which is proven in [FHM11] allows to draw further conclusions concerning general linear regularization methods. The technique especially allows to generalize a well-known converse result presented in [Neu97]. For details we refer to [FHM11].

13.5. Examples of alternative expressions for source conditions

The aim of this section is to provide concrete examples of approximate source conditions, variational inequalities, and approximate variational inequalities. We derive distance functions (for both approximate concepts) and modifier functions (for variational inequalities) starting from a source condition. This way we see how the concept of source conditions, with which most readers are well acquainted, carries over to the other smoothness concepts.

13.5.1. Power-type source conditions

Assume that $x^\dagger$ satisfies a general source condition with index function $\vartheta(t) = t^\mu$, where $\mu \in (0, \frac{1}{2})$. That is,

$$x^\dagger \in \mathcal{R}\big((A^*A)^\mu\big).$$

We first consider the concept of approximate source conditions with benchmark function $\psi(t) = t^{\frac{1}{2}}$ and distance function d_ψ. To apply Theorem 13.10 we write $x^\dagger$ as $x^\dagger = c(A^*A)^\mu w$ with $\|w\| = 1$. Since $\frac{\psi(t)}{c\vartheta(t)} = \frac{1}{c}t^{\frac{1}{2}-\mu}$ is an index function the theorem yields

$$d_\psi(r) \le c^{\frac{1}{1-2\mu}} r^{\frac{-2\mu}{1-2\mu}} \quad \text{for all } r \ge 0.$$

By Corollary 13.8 we immediately obtain

$$D_{\psi,\beta}(r) \le \tfrac{1}{2(1-\beta)} c^{\frac{2}{1-2\mu}} r^{\frac{-4\mu}{1-2\mu}} \quad \text{for all } r \ge 0,$$

if $\beta \in (0,1)$.

From Corollary 13.7 we see that a variational inequality with modifier function $\varphi = -D^*_{\psi,\beta}(-\bullet)$ is satisfied. With $\tilde{c} := \frac{1}{2(1-\beta)} c^{\frac{2}{1-2\mu}}$ we have

$$-D^*_{\psi,\beta}(-t) = \inf_{r\ge 0}\bigl(tr + D_{\psi,\beta}(r)\bigr) \le \inf_{r\ge 0}\bigl(tr + \tilde{c} r^{\frac{-4\mu}{1-2\mu}}\bigr)$$

and the infimum is attained at $r = \bigl(\frac{4\mu\tilde{c}}{1-2\mu}\bigr)^{\frac{1-2\mu}{1+2\mu}} t^{\frac{1-2\mu}{-1-2\mu}}$. That is,

$$-D^*_{\psi,\beta}(-t) \le \bigl(\tfrac{1-2\mu}{4\mu}\bigr)^{\frac{4\mu}{1+2\mu}} \tilde{c}^{\frac{1-2\mu}{1+2\mu}} t^{\frac{4\mu}{1+2\mu}}.$$

Thus, we obtain a variational inequality with modifier function

$$\varphi(t) = \bigl(\tfrac{1-2\mu}{4\mu}\bigr)^{\frac{4\mu}{1+2\mu}} \tilde{c}^{\frac{1-2\mu}{1+2\mu}} t^{\frac{4\mu}{1+2\mu}}.$$

13.5.2. Logarithmic source conditions

Next to power-type source conditions also logarithmic source conditions were considered in the literature before general source conditions appeared. In [Hoh97, Hoh00] it is shown that power-type source conditions are too strong in some applications, but logarithmic ones are likely to be satisfied. Logarithmic source conditions are general source conditions with index function $\vartheta(t) = (-\ln t)^{-\mu}$ for small $t > 0$, where $\mu > 0$ controls the strength of the source condition. The function $t \mapsto (-\ln t)^{-\mu}$ has a pole at $t = 1$. To obtain an index function (which is defined on $[0,\infty)$) we define ϑ by

$$\vartheta(t) := \begin{cases} 0, & t = 0, \\ (-\ln t)^{-\mu}, & t \in (0, \mathrm{e}^{-2\mu-1}), \\ (2\mu+1)^{-\mu-\frac{1}{2}} \sqrt{2\mu \mathrm{e}^{2\mu+1} t + 1}, & t \ge \mathrm{e}^{-2\mu-1}. \end{cases}$$

This function is twice continuously differentiable and concave.

In [Hoh97] and [Hoh00] convergence rates (for an iteratively regularized Gauss–Newton method and general linear regularization methods, respectively) of the type

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\bigl((-\ln\alpha)^{-\mu}\bigr) \quad \text{if } \alpha \to 0$$

and

$$\bigl\|x^{y^\delta}_{\alpha(\delta)} - x^\dagger\bigr\| = \mathcal{O}\bigl((-\ln\delta)^{-\mu}\bigr) \quad \text{if } \delta \to 0$$

with a priori parameter choice $\alpha(\delta) \sim \delta$ were obtained from a logarithmic source condition. Here, '$\sim$' means that there are $c_1, c_2, \bar{\delta} > 0$ such that $c_1\delta \le \alpha(\delta) \le c_2\delta$ for all $\delta \in (0, \bar{\delta}]$.

In case of Tikhonov regularization the estimate for $\|x_\alpha - x^\dagger\|$ is also a consequence of Proposition 13.2 since ϑ is concave. The estimate for $\|x_{\alpha(\delta)}^{y^\delta} - x^\dagger\|$ at first glance does not coincide with the one proposed by (13.1). Inequality (13.1) with $f = \vartheta$ yields

$$\|x_{\alpha(\delta)}^{y^\delta} - x^\dagger\| \le \tfrac{3}{2}\vartheta\big(g^{-1}(\delta)\big),$$

where $g(t) := \sqrt{t}\vartheta(t)$ and $\alpha(\delta) = g^{-1}(\delta)$. For $\delta \in \big(0, \mathrm{e}^{-\mu-\frac{1}{2}}(2\mu+1)^{-\mu}\big)$ the value $s := g^{-1}(\delta) \in (0, \mathrm{e}^{-2\mu-1})$ can be computed as follows:

$$\begin{aligned} s = g^{-1}(\delta) \quad &\Leftrightarrow \quad \sqrt{s}(-\ln s)^{-\mu} = \delta \quad \Leftrightarrow \quad -\tfrac{1}{2\mu}(\ln s)\mathrm{e}^{-\frac{1}{2\mu}\ln s} = \tfrac{1}{2\mu}\delta^{-\frac{1}{\mu}} \\ &\Leftrightarrow \quad -\tfrac{1}{2\mu}\ln s = W\Big(\tfrac{1}{2\mu}\delta^{-\frac{1}{\mu}}\Big) \quad \Leftrightarrow \quad s = \mathrm{e}^{-2\mu W\left(\frac{1}{2\mu}\delta^{-\frac{1}{\mu}}\right)}. \end{aligned}$$

The function W is called *Lambert W function* and described in Chapter D. Having the inverse function g^{-1} at hand we further obtain

$$\vartheta\big(g^{-1}(\delta)\big) = \left(-\ln \mathrm{e}^{-2\mu W\left(\frac{1}{2\mu}\delta^{-\frac{1}{\mu}}\right)}\right)^{-\mu} = \Big(2\mu W\Big(\tfrac{1}{2\mu}\delta^{-\frac{1}{\mu}}\Big)\Big)^{-\mu}$$

and the asymptotic behavior of W (cf. (D.2)) yields

$$\vartheta\big(g^{-1}(\delta)\big) \sim \Big(2\mu \ln\Big(\tfrac{1}{2\mu}\delta^{-\frac{1}{\mu}}\Big)\Big)^{-\mu} = \left(-2\ln\left(\left(\tfrac{1}{2\mu}\right)^{-\mu}\delta\right)\right)^{-\mu} \sim (-\ln\delta)^{-\mu}.$$

Thus, also from (13.1) we obtain the convergence rate

$$\|x_{\alpha(\delta)}^{y^\delta} - x^\dagger\| = \mathcal{O}\big((-\ln\delta)^{-\mu}\big) \quad \text{if } \delta \to 0.$$

Now we come to the main purpose of this subsection, the reformulation of a logarithmic source condition as approximate source condition and (approximate) variational inequality. We start with approximate source conditions.

Let $\psi(t) = t^{\frac{1}{2}}$ be the benchmark index function and choose $c > 0$ such that $x^\dagger = c\vartheta(A^*A^*)w$ for some $w \in X$ with $\|w\| = 1$ (which is always possible if $x^\dagger \in \mathcal{R}(\vartheta(A^*A))$). One easily verifies that the function $\frac{\psi}{c\vartheta}$ is an index function (with $\frac{\psi}{c\vartheta}(0) := 0$). Thus by Theorem 13.10 we have

$$d_\psi(r) \le r\psi\left(\left(\frac{\psi}{c\vartheta}\right)^{-1}\left(\frac{1}{r}\right)\right) \quad \text{for all } r \ge 0.$$

For $r > c\mathrm{e}^{\mu+\frac{1}{2}}(2\mu+1)^{-\mu}$ the value $s := \big(\frac{\psi}{c\vartheta}\big)^{-1}\big(\frac{1}{r}\big) \in (0, \mathrm{e}^{-2\mu-1})$ can be calculated as follows:

$$\begin{aligned} s = \left(\frac{\psi}{c\vartheta}\right)^{-1}\left(\frac{1}{r}\right) \quad &\Leftrightarrow \quad \frac{1}{c}\sqrt{s}(-\ln s)^{\mu} = \frac{1}{r} \quad \Leftrightarrow \quad \frac{1}{2\mu}(\ln s)\mathrm{e}^{\frac{1}{2\mu}\ln s} = -\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}} \\ &\Leftrightarrow \quad \frac{1}{2\mu}\ln s = W_{-1}\left(-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right) \quad \Leftrightarrow \quad s = \mathrm{e}^{2\mu W_{-1}\left(-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right)}; \end{aligned}$$

here W_{-1} denotes a branch a the Lambert W function (see Chapter D). Thus,

$$d_\psi(r) \le r\mathrm{e}^{\mu W_{-1}\left(-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right)} \quad \text{for all } r > c\mathrm{e}^{\mu+\frac{1}{2}}(2\mu+1)^{-\mu}.$$

By the definition of W_{-1} the equality $\mathrm{e}^{W_{-1}(t)} = \frac{t}{W_{-1}(t)}$ is true and therefore

$$d_\psi(r) \le r\left(\frac{-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}}{W_{-1}\left(-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right)}\right)^{\mu} = (2\mu)^{-\mu}c\left(-W_{-1}\left(-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right)\right)^{\mu}.$$

The asymptotic behavior of W_{-1} near zero (cf. (D.3)) implies

$$(2\mu)^{-\mu}c\left(-W_{-1}\left(-\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right)\right)^{\mu} \sim (2\mu)^{-\mu}c\left(-\ln\left(\frac{1}{2\mu}\left(\frac{c}{r}\right)^{\frac{1}{\mu}}\right)\right)^{\mu} \sim (\ln r)^{-\mu}.$$

In other words, there are $\tilde{c}, r_0 > 0$ such that

$$d_\psi(r) \le \tilde{c}(\ln r)^{-\mu} \quad \text{for all } r \ge r_0.$$

For approximate variational inequalities we have the estimate

$$D_{\psi,\beta}(r) \le \tfrac{\tilde{c}^2}{2(1-\beta)}(\ln r)^{-2\mu} \quad \text{for all } r \ge r_0$$

by Corollary 13.8.

It remains to derive a variational inequality. Corollary 13.7 yields a variational inequality with $\varphi = -D^*_{\psi,\beta}(-\bullet)$, which is our starting point. For $t \ge 0$ and with $\bar{c} := \frac{\tilde{c}^2}{2(1-\beta)}$ we first observe

$$-D^*_{\psi,\beta}(-t) = \inf_{r\ge 0}\big(tr + D_{\psi,\beta}(r)\big) \le \inf_{r\ge r_0}\big(tr + D_{\psi,\beta}(r)\big) \le \inf_{r\ge r_0}\big(tr + \bar{c}(\ln r)^{-2\mu}\big).$$

For small $t > 0$ we may choose $r = r_t$ in the infimum with r_t defined by

$$tr_t = \bar{c}(\ln r_t)^{-2\mu}.$$

This definition can be reformulated as follows:

$$\begin{aligned} tr_t = \bar{c}(\ln r_t)^{-2\mu} \quad &\Leftrightarrow \quad \frac{1}{2\mu}(\ln r_t)\mathrm{e}^{\frac{1}{2\mu}\ln r_t} = \frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}} \\ &\Leftrightarrow \quad \frac{1}{2\mu}\ln r_t = W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right) \quad \Leftrightarrow \quad r_t = \mathrm{e}^{2\mu W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)}. \end{aligned}$$

Therefore

$$-D^*_{\psi,\beta}(-t) \le tr_t + \bar{c}(\ln r_t)^{-2\mu} = 2tr_t = 2t\mathrm{e}^{2\mu W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)}$$

for small $t > 0$. Using again the relation $\mathrm{e}^{W(t)} = \frac{t}{W(t)}$ and taking into account the asymptotic behavior of W at infinity (cf. (D.2)) we see

$$\begin{aligned} 2t\mathrm{e}^{2\mu W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)} &= 2t\left(\mathrm{e}^{W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)}\right)^{2\mu} = 2t\left(\frac{\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}}{W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)}\right)^{2\mu} \\ &= 2(2\mu)^{-2\mu}\bar{c}W\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)^{-2\mu} \\ &\sim 2(2\mu)^{-2\mu}\bar{c}\left(\ln\left(\frac{1}{2\mu}\bar{c}^{\frac{1}{2\mu}}t^{-\frac{1}{2\mu}}\right)\right)^{-2\mu} \sim (-\ln t)^{-2\mu}. \end{aligned}$$

Thus, there are $\bar{t}, \hat{c} > 0$ such that

$$-D^*_{\psi,\beta}(-t) \leq \hat{c}(-\ln t)^{-2\mu} \quad \text{for all } t \in [0, \bar{t}].$$

Consequently we find a concave function φ with

$$\varphi(t) = \hat{c}(-\ln t)^{-2\mu}$$

for small $t > 0$ and $-D^*_{\psi,\beta}(-t) \leq \varphi(t)$ for all $t \geq 0$. For such a function φ a variational inequality is fulfilled (since this is the case if $\varphi = -D_{\psi,\beta}(-\bullet)$). The corresponding convergence rate in case of noisy data is

$$\left\|x^{y^\delta}_{\alpha(\delta)} - x^\dagger\right\| = \mathcal{O}\big((-\ln\delta)^{-\mu}\big) \quad \text{if } \delta \to 0$$

with a suitable a priori parameter choice $\delta \mapsto \alpha(\delta)$ (cf. (13.7)). In other words, the obtained variational inequality provides the same convergence rate as the logarithmic source condition.

13.6. Concrete examples of distance functions

In this last section on solution smoothness in Hilbert spaces we calculate and plot distance functions for two concrete operators A and one fixed exact solution $x^\dagger$. At first we present the general approach and then we come to the examples.

13.6.1. A general approach for calculating and plotting distance functions

For calculating distance functions we use the following simple observation, which also appears in [FHM11, Lemma 4] and for a special case also in [HSvW07].

Proposition 13.12. *Let $x^\dagger$ satisfy an approximate source condition with benchmark ψ and let d_ψ be the corresponding distance function. Then for all $\lambda > 0$ the equality*

$$d_\psi\big(\|(\psi^2(A^*A) + \lambda I)^{-1}\psi(A^*A)x^\dagger\|\big) = \lambda\|(\psi^2(A^*A) + \lambda I)^{-1}x^\dagger\| \tag{13.14}$$

holds true.

Proof. Setting $w_\lambda := (\psi^2(A^*A) + \lambda I)^{-1}\psi(A^*A)x^\dagger$ we see $\psi(A^*A)(\psi(A^*A)w_\lambda - x^\dagger) = -\lambda w_\lambda$, that is, w_λ is a minimizer of $\|\psi(A^*A)w - x^\dagger\|^2$ with constraint $\|w\|^2 - r^2 \leq 0$ and λ is the corresponding Lagrange multiplier. Because $\lambda > 0$, we have $\|w_\lambda\| = r$ and therefore

$$d_\psi(\|w_\lambda\|) = \|x^\dagger - \psi(A^*A)w_\lambda\| = \lambda\|(\psi^2(A^*A) + \lambda I)^{-1}x^\dagger\|.$$

□

We define functions $g : (0, \infty) \to [0, \infty)$ and $h : (0, \infty) \to [0, \infty)$ by

$$g(\lambda) := \|(\psi^2(A^*A) + \lambda I)^{-1}\psi(A^*A)x^\dagger\| \tag{13.15}$$

and

$$h(\lambda) := \lambda\|(\psi^2(A^*A) + \lambda I)^{-1}x^\dagger\| \tag{13.16}$$

for $\lambda > 0$. With these functions equation (13.14) reads as

$$d_\psi(g(\lambda)) = h(\lambda) \quad \text{for all } \lambda > 0.$$

The two functions have the following useful properties.

Proposition 13.13. *Let the functions g and h be defined by* (13.15) *and* (13.16)*, respectively, and assume $x^\dagger \neq 0$. Then*

- *g is strictly monotonically decreasing and h is strictly monotonically increasing;*
- *g and h are continuous;*
- $\lim\limits_{\lambda\to+0} g(\lambda) = \begin{cases} \|w\| & \text{if } x^\dagger = \psi(A^*A)w, \\ \infty, & \text{if } x^\dagger \notin \mathcal{R}(\psi(A^*A)) \end{cases}$ *and* $\lim\limits_{\lambda\to\infty} g(\lambda) = 0$*;*
- $\lim\limits_{\lambda\to+0} h(\lambda) = 0$ *and* $\lim\limits_{\lambda\to\infty} h(\lambda) = \|x^\dagger\|$.

Proof. All assertions can be easily verified by writing the functions g and h as an integral over the spectrum of A^*A (see [Yos95] for details on spectral calculus) and applying Lebesgue's dominant convergence theorem. □

By the proposition the function g is invertible (if $x^\dagger \neq 0$) and thus equation (13.14) is equivalent to

$$d_\psi(r) = h\big(g^{-1}(r)\big) \quad \text{for all } r \in \mathcal{R}(g),$$

where

$$\mathcal{R}(g) = \begin{cases} (0, \|w\|) & \text{if } x^\dagger = \psi(A^*A)w, \\ (0, \infty), & \text{if } x^\dagger \notin \mathcal{R}(\psi(A^*A)). \end{cases}$$

In case $x^\dagger = \psi(A^*A)w$ we obviously have $d_\psi(r) = 0$ for $r \geq \|w\|$.

These observations allow to calculate distance functions d_ψ for concrete operators A and concrete exact solutions $x^\dagger$ if g and h can be calculated. If the derivation of an explicit expression for g^{-1} is too complicated then for plotting d_ψ it suffices to invert g numerically.

13.6.2. Example 1: integration operator

Set $X := Y := L^2(0,1)$ and let $A : L^2(0,1) \to L^2(0,1)$ be the integration operator defined by

$$(Ax)(t) := \int_0^t x(s)\,\mathrm{d}s, \quad t \in [0,1].$$

Integration by parts yields the adjoint

$$(A^*y)(t) = \int_t^1 y(s)\,\mathrm{d}s, \quad t \in [0,1],$$

and therefore

$$(A^*Ax)(t) = \int_t^1 \int_0^s x(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}s, \quad t \in [0,1].$$

We choose the benchmark function

$$\psi(t) := t^{\frac{1}{2}}, \quad t \in [0,\infty),$$

and the exact solution

$$x^\dagger(t) := 1, \quad t \in [0,1].$$

Since

$$\mathcal{R}(\psi(A^*A)) = \mathcal{R}\big((A^*A)^{\frac{1}{2}}\big) = \mathcal{R}(A^*) = \{x \in L^2(0,1) : x \in H^1(0,1),\ x(1) = 0\},$$

we see that $x^\dagger \notin \mathcal{R}(\psi(A^*A))$.

The present example is also considered in [HSvW07], where the authors bound the distance function d_ψ by

$$\frac{\sqrt{3}}{4} r^{-1} \le d_\psi(r) \le \frac{1}{\sqrt{2}} r^{-1} \tag{13.17}$$

for sufficiently large r. Our aim is to derive better constants and to plot the graph of d_ψ.

For calculating the functions g and h defined by (13.15) and (13.16), respectively, we first evaluate the expression $(\psi^2(A^*A) + \lambda I)^{-1}x^\dagger$ with $\lambda > 0$, that is, we solve

$$(\psi^2(A^*A) + \lambda I)x = x^\dagger$$

for x. This last equality is equivalent to

$$\int_t^1 \int_0^s x(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}s + \lambda x(t) = 1, \quad t \in [0,1],$$

and simple calculations yield the equivalent formulation

$$x = \lambda x'', \quad x(1) = \frac{1}{\lambda}, \quad x'(0) = 0.$$

The solution of the differential equation is

$$\big((\psi^2(A^*A) + \lambda I)^{-1}x^\dagger\big)(t) = x(t) = \frac{1}{\lambda \cosh\left(\frac{1}{\sqrt{\lambda}}\right)} \cosh\left(\frac{1}{\sqrt{\lambda}}t\right), \quad t \in [0,1].$$

As the next step for obtaining the function g we observe

$$\begin{aligned}\|(\psi^2(A^*A)+\lambda I)^{-1}\psi(A^*A)x^\dagger\| &= \big\|(A^*A)^{\frac{1}{2}}(A^*A+\lambda I)^{-1}x^\dagger\big\| \\ &= \|A(A^*A+\lambda I)^{-1}x^\dagger\|.\end{aligned}$$

Thus, with

$$\begin{aligned}\big(A(A^*A+\lambda I)^{-1}x^\dagger\big)(t) &= \int_0^t \frac{1}{\lambda\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\cosh\left(\frac{1}{\sqrt{\lambda}}t\right)\,\mathrm{d}t \\ &= \frac{1}{\sqrt{\lambda}\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\sinh\left(\frac{1}{\sqrt{\lambda}}t\right)\end{aligned}$$

for $t\in[0,1]$ we see

$$g(\lambda)=\sqrt{\int_0^1\big(\big(A(A^*A+\lambda I)^{-1}x^\dagger\big)(t)\big)^2\,\mathrm{d}t}=\frac{1}{2\sqrt{\lambda}\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\sqrt{\sqrt{\lambda}\sinh\left(\frac{2}{\sqrt{\lambda}}\right)-2}$$

for $\lambda>0$. The function h is given by

$$\begin{aligned}h(\lambda)=\lambda\|(\psi^2(A^*A)+\lambda I)^{-1}x^\dagger\| &= \lambda\sqrt{\int_0^1\left(\frac{1}{\lambda\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\cosh\left(\frac{1}{\sqrt{\lambda}}t\right)\right)^2\,\mathrm{d}t} \\ &= \frac{1}{2\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\sqrt{\sqrt{\lambda}\sinh\left(\frac{2}{\sqrt{\lambda}}\right)+2}\end{aligned}$$

for $\lambda>0$. Equation (13.14) thus reads as

$$d_\psi\left(\frac{1}{2\sqrt{\lambda}\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\sqrt{\sqrt{\lambda}\sinh\left(\frac{2}{\sqrt{\lambda}}\right)-2}\right)=\frac{1}{2\cosh\left(\frac{1}{\sqrt{\lambda}}\right)}\sqrt{\sqrt{\lambda}\sinh\left(\frac{2}{\sqrt{\lambda}}\right)+2}$$

for all $\lambda>0$.

To obtain better constants for bounding d_ψ than in (13.17) we calculate the limit of $\frac{d_\psi(r)}{r^{-1}}$ for $r\to\infty$. We have

$$\begin{aligned}\lim_{r\to\infty}\frac{d_\psi(r)}{r^{-1}} &= \lim_{r\to\infty} r d_\psi(r) = \lim_{\lambda\to 0} g(\lambda)d_\psi(g(\lambda)) = \lim_{\lambda\to 0} g(\lambda)h(\lambda) \\ &= \lim_{\lambda\to 0}\frac{\sqrt{\lambda\sinh^2\left(\frac{2}{\sqrt{\lambda}}\right)-4}}{4\sqrt{\lambda}\cosh^2\left(\frac{1}{\sqrt{\lambda}}\right)} = \lim_{t\to\infty}\frac{\sqrt{\sinh^2(2t)-4t^2}}{4\cosh^2 t}\end{aligned}$$

and using the relation $2\cosh^2 t=\cosh(2t)+1$ we further obtain

$$\lim_{r\to\infty}\frac{d_\psi(r)}{r^{-1}}=\frac{1}{2}\lim_{t\to\infty}\frac{\sqrt{\sinh^2(2t)-4t^2}}{\cosh(2t)+1}=\frac{1}{2}\lim_{s\to\infty}\frac{\sqrt{\sinh^2 s-s^2}}{\cosh s+1}.$$

Replacing sinh and cosh by corresponding sums of exponential functions we see that the last limit is one. Therfore, $\lim_{r\to\infty} \frac{d_\psi(r)}{r^{-1}} = \frac{1}{2}$ and thus for each $\varepsilon > 0$ we find $r_\varepsilon > 0$ such that

$$\left(\frac{1}{2} - \varepsilon\right) r^{-1} \leq d_\psi(r) \leq \left(\frac{1}{2} + \varepsilon\right) r^{-1} \quad \text{for all } r \geq r_\varepsilon.$$

We have no explicit formula for g^{-1} but for plotting d_ψ we can invert g numerically using, e.g., the bisection method. The plot obtained this way is shown in Figure 13.1.

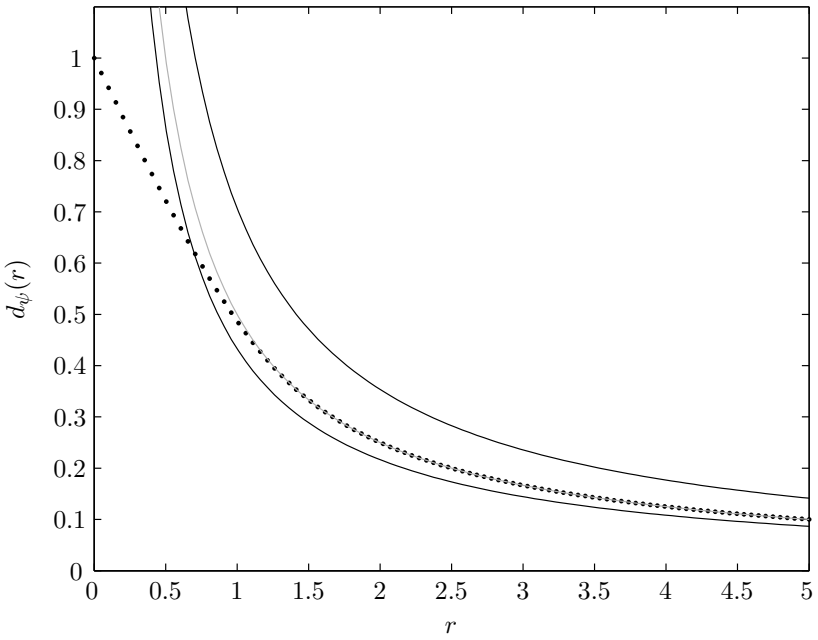

Figure 13.1.: Distance function d_ψ (black dots), function $r \mapsto \frac{1}{2}r^{-1}$ (solid gray line), and lower and upper bound from [HSvW07] (solid black lines).

13.6.3. Example 2: multiplication operator

Set $X := Y := L^2(0,1)$ and let $A : L^2(0,1) \to L^2(0,1)$ be the multiplication operator defined by

$$(Ax)(t) := m(t)x(t), \quad t \in [0,1],$$

with $m \in L^\infty(0,1)$. Obviously the adjoint is given by

$$(A^*y)(t) = m(t)y(t), \quad t \in [0,1],$$

and therefore

$$(A^*Ax)(t) = m^2(t)x(t), \quad t \in [0,1].$$

We choose the benchmark function

$$\psi(t) := t^{\frac{1}{2}}, \quad t \in [0,\infty),$$

and the exact solution

$$x^\dagger(t) := 1, \quad t \in [0,1].$$

Then $x^\dagger \in \mathcal{R}(\psi(A^*A)) = \mathcal{R}(A^*)$ if and only if $\frac{1}{m} \in L^2(0,1)$.

The results on distance functions in case of the multiplication operators under consideration were already published in [FHM11]. In this thesis we give some more details.

For calculating the functions g and h defined by (13.15) and (13.16), respectively, we observe

$$\big((A^*A + \lambda I)^{-1} x^\dagger\big)(t) = \frac{1}{m^2(t) + \lambda}, \quad t \in [0,1].$$

From this relation we easily obtain

$$g(\lambda) = \|A(A^*A + \lambda I)^{-1} x^\dagger\| = \sqrt{\int_0^1 \frac{m^2(t)}{(m^2(t)+\lambda)^2}\,\mathrm{d}t}$$

and

$$h(\lambda) = \lambda \|(A^*A + \lambda I)^{-1} x^\dagger\| = \sqrt{\int_0^1 \frac{\lambda^2}{(m^2(t)+\lambda)^2}\,\mathrm{d}t}$$

for all $\lambda > 0$.

From now on we only consider the multiplier function

$$m(t) := \sqrt{t}, \quad t \in [0,1].$$

Since $\frac{1}{m} \notin L^2(0,1)$ we conclude $x^\dagger \notin \mathcal{R}(\psi(A^*A))$. With this specific m we have

$$g(\lambda) = \sqrt{\ln \frac{\lambda+1}{\lambda} - \frac{1}{\lambda+1}}$$

and

$$h(\lambda) = \sqrt{\frac{\lambda}{\lambda+1}}$$

for all $\lambda > 0$. Thus,

$$d_\psi\left(\sqrt{\ln \frac{\lambda+1}{\lambda} - \frac{1}{\lambda+1}}\right) = \sqrt{\frac{\lambda}{\lambda+1}}$$

for all $\lambda > 0$.

In the remaining part of this section we first derive lower and upper bounds for d_ψ which hold for all $r > 0$. Then we improve the constants in the bounds and finally we plot the distance function.

To obtain lower and upper bounds for d_ψ we transform equation (13.14), which is equivalent to $d_\psi(g(\lambda)) = h(\lambda)$ for all $\lambda > 0$, as follows:

$$\begin{aligned}
d_\psi(r) = h(g^{-1}(r)) \quad &\Leftrightarrow \quad r = g\big(h^{-1}(d_\psi(r))\big) \quad \Leftrightarrow \quad r = g\left(\frac{d_\psi^2(r)}{1 - d_\psi^2(r)}\right) \\
&\Leftrightarrow \quad r^2 = \ln \frac{1}{d_\psi^2(r)} - 1 + d_\psi^2(r) \quad \Leftrightarrow \quad \mathrm{e}^{r^2} = \frac{1}{d_\psi^2(r)} \mathrm{e}^{-1} \mathrm{e}^{d_\psi^2(r)} \\
&\Leftrightarrow \quad d_\psi(r) = \mathrm{e}^{-\frac{1}{2}(r^2+1)} \mathrm{e}^{\frac{1}{2} d_\psi^2(r)}.
\end{aligned}$$

Since

$$1 \leq \mathrm{e}^{\frac{1}{2}d_\psi^2(r)} \leq \mathrm{e}^{\frac{1}{2}d_\psi^2(0)} = \mathrm{e}^{\frac{1}{2}\|x^\dagger\|} = \mathrm{e}^{\frac{1}{2}} \quad \text{for all } r > 0$$

we obtain

$$\mathrm{e}^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}r^2} \leq d_\psi(r) \leq \mathrm{e}^{-\frac{1}{2}r^2} \quad \text{for all } r > 0. \tag{13.18}$$

To improve the constants in the bounds we calculate

$$\begin{aligned}\lim_{r\to\infty} \frac{d_\psi(r)}{\mathrm{e}^{-\frac{1}{2}r^2}} &= \lim_{r\to\infty} d_\psi(r)\mathrm{e}^{\frac{1}{2}r^2} = \lim_{\lambda\to 0} d_\psi(g(\lambda))\mathrm{e}^{\frac{1}{2}g(\lambda)^2} = \lim_{\lambda\to 0} h(\lambda)\mathrm{e}^{\frac{1}{2}g(\lambda)^2} \\ &= \lim_{\lambda\to 0} \sqrt{\frac{\lambda}{\lambda+1}}\mathrm{e}^{\frac{1}{2}\left(\ln\frac{\lambda+1}{\lambda} - \frac{1}{\lambda+1}\right)} = \lim_{\lambda\to 0} \mathrm{e}^{-\frac{1}{2(\lambda+1)}} = \mathrm{e}^{-\frac{1}{2}}.\end{aligned}$$

Thus, for each $\varepsilon > 0$ there is $r_\varepsilon > 0$ such that

$$\left(\mathrm{e}^{-\frac{1}{2}} - \varepsilon\right)\mathrm{e}^{-\frac{1}{2}r^2} \leq d_\psi(r) \leq \left(\mathrm{e}^{-\frac{1}{2}} + \varepsilon\right)\mathrm{e}^{-\frac{1}{2}r^2} \quad \text{for all } r \geq r_\varepsilon.$$

If we invert the function g numerically we can plot the distance function d_ψ. The result is depicted in Figure 13.2.

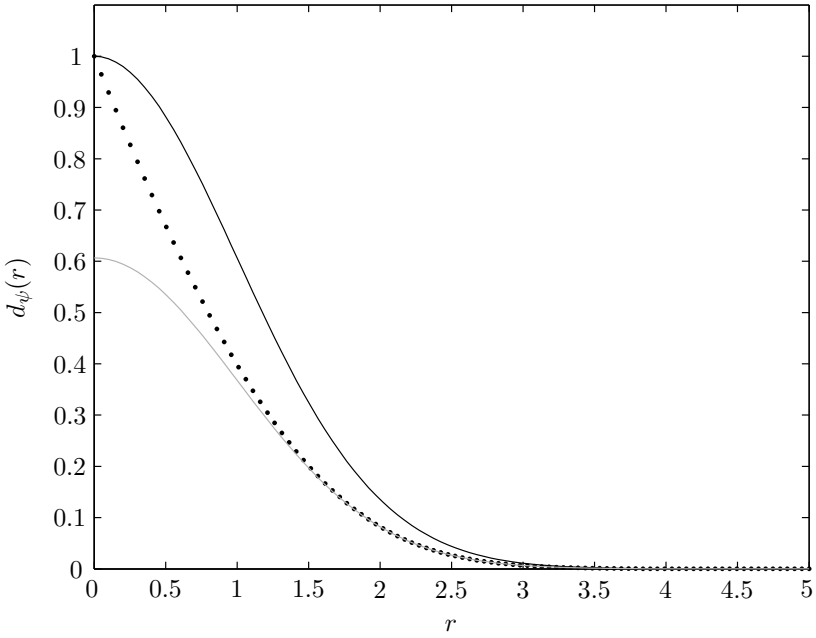

Figure 13.2.: Distance function d_ψ (black dots), function $r \mapsto \mathrm{e}^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}r^2}$ (solid gray line), and upper bound from (13.18) (solid black line). Note that the lower bound in (13.18) coincides with the gray line.

Note that in the specific example under consideration also analytic inversion of g is possible. Using the formula $d_\psi(r) = h(g^{-1}(r))$ then we obtain

$$d_\psi(r) = \sqrt{-W\left(-\mathrm{e}^{-(r^2+1)}\right)} \quad \text{for all } r \geq 0,$$

where W denotes the Lambert W function (see Chapter D).

Appendix

A. General topology

In this chapter we collect some basic definitions and theorems from general topology. Only the things necessary for reading the first part of this thesis are given here. Since most of the material can be found in every book on topology, we omit the proofs.

A.1. Basic notions

By $\mathcal{P}(X)$ we denote the power set of a set X.

Definition A.1. A *topological space* is a pair (X, τ) consisting of a nonempty set X and a nonempty family $\tau \subseteq \mathcal{P}(X)$ of subsets of X, where τ has to satisfy the following properties:

- $\emptyset \in \tau$ and $X \in \tau$,
- $G_i \in \tau$ for $i \in I$ (arbitrary index set) implies $\bigcup_{i \in I} G_i \in \tau$,
- $G_1, G_2 \in \tau$ implies $G_1 \cap G_2 \in \tau$.

The family τ is called *topology* on X.

Definition A.2. Let (X, τ) be a topological space. A set $A \subseteq X$ is called *open* if $A \in \tau$. It is called *closed* if $X \setminus A \in \tau$. The union of all open sets contained in a set $A \subseteq X$ is called the *interior* of A and is denoted by $\operatorname{int} A$. The intersection of all closed sets covering A is the *closure* of A and it is denoted by $\overline{A}$.

Definition A.3. Let X be a nonempty set and let $\tau_1, \tau_2 \subseteq \mathcal{P}(X)$ be two topologies on X. The topology τ_1 is *weaker* (or *coarser*) than τ_2 if $\tau_1 \subseteq \tau_2$. In this case τ_2 is *stronger* (or *finer*) than τ_1.

Definition A.4. Let (X, τ) be a topological space and let $x \in X$. A set $N \subseteq X$ is called *neighborhood* of x if there is an open set $G \in \tau$ with $w \in G \subseteq N$. The family of all neighborhoods of x is denoted by $\mathcal{N}(x)$.

Definition A.5. Let (X, τ) be a topological space and let $\tilde{X} \subseteq X$. The set $\tilde{\tau} := \{G \cap \tilde{X} : G \in \tau\} \subseteq \mathcal{P}(\tilde{X})$ is the *topology induced by* τ and the topological space $(\tilde{X}, \tilde{\tau})$ is called *topological subspace* of (X, τ).

Given two topological spaces (X, τ_X) and (Y, τ_Y) there is a natural topology on $X \times Y$, the *product topology* $\tau_X \otimes \tau_Y$. We do not want to go into the details of its definition here, since this requires some effort and the interested reader finds the topic in each book on general topology.

A.2. Convergence

The natural notion of convergence in general topological spaces is the convergence of nets.

Definition A.6. A nonempty index set I is *directed* if there is a relation $\preceq$ on I such that

- $i \preceq i$ for all $i \in I$,
- $i_1 \preceq i_2$, $i_2 \preceq i_3$ implies $i_1 \preceq i_3$ for all $i_1, i_2, i_3 \in I$,
- for all $i_1, i_2 \in I$ there is some $i_3 \in I$ with $i_1 \preceq i_3$ and $i_2 \preceq i_3$.

Definition A.7. Let X be a nonempty set and let I be a directed index set. A *net* (or a *Moore–Smith sequence*) in X is a mapping $\phi : I \to X$. Instead of ϕ we usually write $(x_i)_{i\in I}$ with $x_i := \phi(i)$.

Definition A.8. Let (X, τ) be a topological space. A net $(x_i)_{i\in I}$ *converges* to $x \in X$ if for each neighborhood $N \in \mathcal{N}(x)$ there is an index $i_0 \in I$ such that $x_i \in N$ for all $i \succeq i_0$. In this case we write $x_i \to x$.

Note that a net may converge to more than one element. Only additional assumptions guarantee the uniqueness of limiting elements.

Definition A.9. A topological space (X, τ) is a *Hausdorff space* if for arbitrary $x_1, x_2 \in X$ with $x_1 \neq x_2$ there are open sets $G_1, G_2 \in \tau$ with $x_1 \in G_1$, $x_2 \in G_2$, and $G_1 \cap G_2 = \emptyset$.

Proposition A.10. *A topological space is a Hausdorff space if and only if each convergent net converges to exactly one element.*

An example for nets are sequences: each sequence $(x_k)_{k\in\mathbb{N}}$ is a net with index set $I = \mathbb{N}$ and with the usual ordering $\leq$ on $\mathbb{N}$.

In metric spaces topological properties like continuity, closedness, and compactness can be characterized by convergence of sequences. The same is possible in general topological spaces if sequences are replaced by nets. Under additional assumptions it suffices to consider sequences as we will see in the subsequent sections.

Definition A.11. A topological space is an A_1*-space* if for each $x \in X$ there is a countable family $\mathcal{B}(x) \subseteq \mathcal{N}(x)$ of neighborhoods such that for each $N \in \mathcal{N}(x)$ one finds $B \in \mathcal{B}(x)$ with $B \subseteq N$.

We close this section with a remark on convergence with respect to a product topology. Let (X, τ_X) and (Y, τ_Y) be two topological spaces and let $(X \times Y, \tau_X \otimes \tau_Y)$ be the corresponding product space. Then a net $\big((x_i, y_i)\big)_{i\in I}$ in $X \times Y$ converges to $(x, y) \in X \times Y$ if and only if $x_i \to x$ and $y_i \to y$.

A.3. Continuity

Let (X, τ_X) and (Y, τ_Y) be topological spaces.

Definition A.12. A mapping $f : X \to Y$ is *continuous at the point* $x \in X$ if for each neighborhood $M \in \mathcal{N}(f(x))$ there is a neighborhood $N \in \mathcal{N}(x)$ with $f(N) \subseteq M$. The mapping f is *continuous* if it is continuous at each point $x \in X$.

Proposition A.13. *A mapping $f : X \to Y$ is continuous at $x \in X$ if and only if for each net $(x_i)_{i \in I}$ with $x_i \to x$ we also have $f(x_i) \to f(x)$.*

Definition A.14. A mapping $f : X \to Y$ is *sequentially continuous at the point* $x \in X$ if each sequence $(x_k)_{k \in \mathbb{N}}$ with $x_k \to x$ also satisfies $f(x_k) \to f(x)$. The mapping f is *sequentially continuous* if it is sequentially continuous at every point $x \in X$.

Proposition A.15. *Let X be an A_1-space. Then a mapping $f : X \to Y$ is continuous if and only if it is sequentially continuous.*

Definition A.16. A mapping $f : X \to (-\infty, \infty]$ is *sequentially lower semi-continuous* if for each sequence $(x_k)_{k \in \mathbb{N}}$ in X converging to some $x \in X$ we have

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

A mapping $\tilde{f} : X \to [-\infty, \infty)$ is *sequentially upper semi-continuous* if $-\tilde{f}$ is sequentially lower semi-continuous.

A.4. Closedness and compactness

We first characterize closed sets in terms of nets and sequences.

Proposition A.17. *A set $A \subseteq X$ in a topological space (X, τ) is closed if and only if all limits of each convergent net contained in A belong to A.*

Definition A.18. A set $A \subseteq X$ in a topological space (X, τ) is *sequentially closed* if the limits of each convergent sequence contained in A belong to A. The *sequential closure* of a set $A \subseteq X$ is the intersection of all sequentially closed sets covering A.

Since the intersection of sequentially closed sets is sequentially closed, the sequential closure of a set is well-defined.

Proposition A.19. *Let (X, τ) be an A_1-space. A set in (X, τ) is closed if and only if it is sequentially closed.*

The notion of sequential lower semi-continuity can be characterized by the sequential closedness of certain sets.

Proposition A.20. *Let (X, τ) be a topological space. A mapping $f : X \to (-\infty, \infty]$ is sequentially lower semi-continuous if and only if the sublevel sets $M_f(c) := \{x \in X : f(x) \leq c\}$ are sequentially closed for all $c \in \mathbb{R}$.*

In general topological spaces there are several notions of compactness. We restrict our attention to the ones of interest in this thesis.

Definition A.21. Let (X, τ) be a topological space.

- The space (X, τ) is *compact* if every open covering of X contains a finite open covering of X.
- The space (X, τ) is *sequentially compact* if every sequence in X contains a convergent subsequence.
- A set $A \subseteq X$ is *(sequentially) compact* if it is (sequentially) compact as a subspace of (X, τ).

Proposition A.22. *Let (X, τ) be an A_1-space. Then every compact set $A \subseteq X$ is sequentially compact.*

To characterize compactness of a set by convergence of nets one has to introduce the notion of subnets. But this is beyond the scope of this chapter.

Proposition A.23. *Let (X, τ) be sequentially compact. Then each sequentially closed set in X is sequentially compact.*

Next, we introduce a slightly weaker notion of sequential compactness. The definition can also be formulated for non-sequential compactness, but we do not need the non-sequential version in the thesis. The assumption that the underlying space is a Hausdorff space is necessary to establish a strong connection between the weakend notion of sequential compactness and the original sequential compactness (see the corollary below).

Definition A.24. A set in a Hausdorff space is called *relatively sequentially compact* if its sequential closure is sequentially compact.

Proposition A.25. *Each sequentially compact set in a Hausdorff space is sequentially closed.*

Corollary A.26. *A set in a Hausdorff space is sequentially compact if and only if it is sequentially closed and relatively sequentially compact*

Finally we prove a simple result on sequences in compact sets, which we used in Part I.

Proposition A.27. *Let (X, τ) be a topological space and let $(x_k)_{k\in\mathbb{N}}$ be a sequence contained in a sequentially compact set $A \subseteq X$. If all convergent subsequences have the same unique limit $x \in X$, then also the whole sequence (x_k) converges to x and x is the only limit of (x_k).*

Proof. Assume $x_k \nrightarrow x$. Then there would be a neighborhood $N \in \mathcal{N}(x)$ and a subsequence $(x_{k_l})_{l\in\mathbb{N}}$ such that $x_{k_l} \notin N$ for all $l \in \mathbb{N}$. By the compactness of A this sequence would have a convergent subsequence and by assumption its limit would be x. But this is a contradiction. □

B. Convex analysis

We briefly summarize some definitions from convex analysis which we use throughout the thesis.

Let X be a real topological vector space, that is, a real vector space endowed with a topology such that addition and multiplication by scalars are continuous with respect to this topology. By X^* we denote the set of all continuous linear functionals $\xi : X \to \mathbb{R}$ on X. If X is a normed space then X^* is typically equipped with the usual operator norm $\|\xi\| := \sup\{\xi(x) : x \in X,\ \|x\| = 1\}$. Instead of $\xi(x)$ we often write $\langle \xi, x \rangle$.

In convex analysis it is useful to consider functionals on X which may attain the value $+\infty$ or $-\infty$; but not both. Addition and multiplication by scalars is extended to such functionals whenever this extension is intuitive, e.g. $1 + \infty = +\infty$ or $1 \cdot (+\infty) = +\infty$.

Definition B.1. A functional $\Gamma : X \to (-\infty, \infty]$ is *convex* if $\Gamma(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\Gamma(x_1) + (1-\lambda)\Gamma(x_2)$ for all $x_1, x_2 \in X$ and all $\lambda \in [0,1]$. A functional $\tilde{\Gamma} : X \to [-\infty, \infty)$ is *concave* if $-\tilde{\Gamma}$ is convex.

For convex functionals the notion of derivative can be generalized to non-smooth convex functionals.

Definition B.2. Let $\Gamma : X \to (-\infty, \infty]$ be convex and let $x_0 \in X$. An element $\xi \in X^*$ is called a *subgradient* of Γ at x_0 if

$$\Gamma(x) \geq \Gamma(x_0) + \langle \xi, x - x_0 \rangle \quad \text{for all } x \in X.$$

The set of all subgradients at x_0 is called *subdifferential* of Γ at x_0 and is denoted by $\partial\Gamma(x_0)$.

Note that if $\Gamma(x_0) = \infty$ and if there is some $x \in X$ with $\Gamma(x) < \infty$, then $\partial\Gamma(x_0) = \emptyset$. But also in case $\Gamma(x_0) < \infty$ it may happen that $\partial\Gamma(x_0) = \emptyset$.

Based on a convex functional Γ one can define another convex functional which expresses the distance between Γ and one of its linearizations at a fixed point x_0.

Definition B.3. Let $\Gamma : X \to (-\infty, \infty]$ be convex and let $x_0 \in X$ and $\xi_0 \in \partial\Gamma(x_0)$. The functional $B^{\Gamma}_{\xi_0}(\bullet, x_0) : X \to [0, \infty]$ defined by

$$B^{\Gamma}_{\xi_0}(x, x_0) := \Gamma(x) - \Gamma(x_0) - \langle \xi_0, x - x_0 \rangle \quad \text{for } x \in X$$

is called *Bregman distance* with respect to Γ, x_0, and ξ_0.

The Bregman distance can only be defined for $x_0 \in X$ with $\partial\Gamma(x_0) \neq \emptyset$. Since Γ is assumed to be convex, the Bregman distance is also convex. The nonnegativity of $B^{\Gamma}_{\xi_0}(\bullet, x_0)$ follows from $\xi_0 \in \partial\Gamma(x_0)$.

If X is a Hilbert space and $\Gamma = \frac{1}{2}\|\bullet\|^2$, then $\partial\Gamma(x_0) = \{x_0\}$ and the corresponding Bregman distance is given by $B^{\Gamma}_{x_0}(\bullet, x_0) = \frac{1}{2}\|x - x_0\|^2$. Thus, Bregman distances can be regarded as a generalization of Hilbert space norms.

Next to subdifferentiability we also use another concept from convex analysis, conjugate functions. Both concepts are closely related and we refer to [ABM06] for details and proofs.

Definition B.4. Let $f : X \to (-\infty, \infty]$ be a functional on X which is finite at least at one point. The functional $f^* : X^* \to (-\infty, \infty]$ defined by

$$f^*(\xi) := \sup_{x \in X} \big(\langle \xi, x\rangle - f(x)\big) \quad \text{for } \xi \in X^*$$

is the *conjugate function* of f.

Conjugate functions are always convex and lower semi-continuous since they are the supremum over $x \in X$ of the affine functions $\langle \bullet, x\rangle - f(x)$.

We frequently consider conjugate functions of convex functions defined only on $[0, \infty)$ instead of $X = \mathbb{R}$. In such cases we set the function to $+\infty$ on $(-\infty, 0)$. This extension preserves convexity and $\sup_{x \in \mathbb{R}}$ can be replaced by $\sup_{x \geq 0}$ in the definition of the conjugate function.

When working with infima and suprema of functions it is sometimes sensible to use so called indicator functions:

Definition B.5. Let $A \subseteq X$. The function $\delta_A : X \to [0, \infty]$ defined by

$$\delta_A(x) := \begin{cases} 0, & \text{if } x \in A, \\ \infty, & \text{if } x \notin A \end{cases}$$

is called *indicator function* of the set A.

In Subsection 12.1.6 we used two further definitions.

Definition B.6. A Banach space X is *strictly convex* if for all $x_1, x_2 \in X$ with $x_1 \neq x_2$ and $\|x_1\| = \|x_2\| = 1$ and for all $\lambda \in (0, 1)$ the strict inequality $\|\lambda x_1 + (1 - \lambda)x_2\| < 1$ is satisfied.

Definition B.7. A Banach space X is *smooth* if for each $x \in X$ with $\|x\| = 1$ there is exactly one linear functional $\xi \in X^*$ such that $\langle \xi, x\rangle = \|\xi\|$.

C. Conditional probability densities

Since conditional probability densities and Bayes' formula are the main ingredient of the MAP approach described in Chapter 7, it is of high importance to understand the concept behind them. If we do not know how to interpret conditional densities we cannot be sure that the mathematical formulation of the MAP approach coincides with our intuitive notion of 'maximizing conditional probabilities'.

The core, that is, all definitions, propositions, and theorems, of this chapter is taken from the highly recommendable textbook [FG97]. Since all the proofs are given there we do not repeat them here. The core material is enriched by comprehensive interpretive remarks, because we do not aim solely at presenting the technical handling of conditional densities but we intend to bring light into the relations between 'intuitive conditional probabilities' and the mathematical concept of conditional densities.

C.1. Statement of the problem

Let $(\Theta, \mathcal{A}, P)$ be a probability space, that is, $P(A)$ is the probability that the outcome $\theta \in \Theta$ of a realization of the underlying experiment lies in $A \in \mathcal{A}$. This interpretation is only true if no information about θ is available. But if we would have additional knowledge then the probability for $\theta \in A$, in general, would be different from $P(A)$. So, two questions have to be answered: How to formulate 'additional information' in mathematical terms, and how to express the probability of $\theta \in A$ with respect to additional knowledge?

The well-known 'simple' conditional probability is usually formulated as follows: Given the information $\theta \in B \in \mathcal{A}$ and assuming $P(B) > 0$ the probability of $\theta \in A$ is $\frac{P(A \cap B)}{P(B)}$. This definition is very intuitive, but it does not include the case $P(B) = 0$. Thus, a more general, though less intuitive, definition is necessary.

First, we introduce a generalized concept of 'additional information': The idea is to look at all events whose occurrence is completely determined by the information available about θ. More precisely, let $\mathcal{B} \subseteq \mathcal{A}$ be the family of all events $B \in \mathcal{A}$ for which we can decide whether $\theta \in B$ or $\theta \notin B$. In the above case of knowing that $\theta \in B$ we would get $\mathcal{B} = \{\emptyset, \Theta, B, \Theta \setminus B\}$, which is a σ-algebra, as the family of decidable events. Intuition suggests that the informally defined family $\mathcal{B}$ is always a σ-algebra (if we know whether $\theta \in B$ or not then we also know whether $\theta \in \Theta \setminus B$ and so on). Thus, sub-σ-algebras $\mathcal{B} \subseteq \mathcal{A}$ of decidable events turn out to be a suitable tool for expressing additional knowledge about an outcome θ.

A more serious question is the second one: How to express the probability of $\theta \in A$ with respect to additional knowledge, that is, with respect to a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{A}$? If $P(B) > 0$ would be true for all $B \in \mathcal{B}$ the appropriate answer would be the mapping $P(A|\bullet) : \mathcal{B} \to [0, 1]$ defined by $P(A|B) := \frac{P(A \cap B)}{P(B)}$. For $P(B) = 0$ this definition does not

work (division by zero), but we have an intuitive notion of 'conditional probability' even with respect to events of probability zero. This deficiency is not due to an inappropriate mathematical definition of this intuitive notion, but it is an intrinsic property of events of probability zero. A bit sloppy one could say that events of probability zero are events the probability measure does not care about, that is, it provides no means of working with them. Thus, the only chance to get (maybe very weak) information about the relation between $A \cap B$ and B is by approximating B. In other words, we try to extend the mapping $P(A|\bullet)$ to events of probability zero.

Defining $P(A|B)$ for $P(B) = 0$ to be the limit of $P(A|B_i)$ if $i \to \infty$ for some sequence $(B_i)_{i\in\mathbb{N}}$ satisfying $P(B_i) > 0$ and $B_1 \supseteq B_2 \supseteq \cdots \supseteq B$ seems to be a good idea, at least at the first look. But simple examples show that, assuming it exists, the limit depends on the sequence $(B_i)_{i\in\mathbb{N}}$ and the sequence can be chosen in such a way that the limit coincides with any given real number. Thus, to grasp the approximation idea in a precise mathematical way we need another concept. As we will see below, considering densities of measures instead of measures themselves solves the problem.

C.2. Interpretation of densities

Let $(\Theta, \mathcal{A})$ be a measurable space and let μ and ν be two measures on $(\Theta, \mathcal{A})$. A measurable function $f : \Theta \to [0, \infty)$ is called *density of ν with respect to μ* if

$$\nu(A) = \int_A f \, \mathrm{d}\mu \quad \text{for all } A \in \mathcal{A}.$$

Obviously, if ν has a density with respect to μ then ν is absolutely continuous with respect to μ, that is $\mu(A) = 0$ for some $A \in \mathcal{A}$ implies $\nu(A) = 0$. The following theorem tells us that for σ-finite measures also the converse direction is true.

Theorem C.1 (Radon–Nikodym)**.** *Let μ and ν be σ-finite measures on a common measurable space. If ν is absolutely continuous with respect to μ then ν has a density with respect to μ. If f_1 and f_2 are two such densities then $f_1 = f_2$ almost everywhere with respect to μ.*

Why do we need densities? At first glance, they reveal exactly the same information as the measure ν itself. But since, in general, densities are not uniquely determined we could choose a 'nice' one to work with. In our sense, a nice density should realize the approximation idea from the end of the previous section. More precisely, on sets of ν-measure zero it should exhibit the same behavior as on a 'neighborhood' of this set.

Such a connection between a density's behavior on a set and its behavior on a neighborhood of this set can be described by continuity with respect to a topology on Θ and a topology on $[0, \infty]$. The former topology should be as weak as possible and the latter one should be as strong as possible to make the connection a strong one. In brief, we can state the following:

- Without additional knowledge about a density the density's values on sets of μ-measure zero do not contain any information.

- Knowing that a density is continuous with respect to a (hopefully weak) topology on Θ and a (hopefully strong) topology on $[0,\infty]$, its behavior on sets of μ-measure zero provides information about the measure of sets in a neighborhood of sets of μ-measure zero.

Care has to be taken of what measure is used when talking about sets of measure zero. Actually we are interested in ν-null sets, but the above considerations concentrate on sets with μ-measure zero. On sets for which the above-mentioned does not apply, that is, on sets $A \in \mathcal{A}$ with $\nu(A) = 0$ but $\mu(A) > 0$ each density is zero μ-almost everywhere. Thus, in the sense of densities such sets are 'strong' null sets (with respect to μ).

C.3. Definition of conditional probabilities

The previous section suggests that densities of measures are a suitable tool to work with sets of measure zero. To motivate the use of densities for representing conditional probabilities also with respect to sets of probability zero we first look at 'simple' conditional probabilities in conjunction with the law of total probability.

Let $(\Theta, \mathcal{A}, P)$ be a probability space, let $A \in \mathcal{A}$, and let $\mathcal{B} \subseteq \mathcal{A}$ be the σ-algebra of decidable events. For fixed $B \in \mathcal{B}$ assume that $\{B_1, \ldots, B_n\} \subseteq \mathcal{B}$ is a partition of B, that is, the sets B_i are mutually disjoint and $B = \bigcup_{i=1}^n B_i$. If $P(B_i) > 0$ for $i = 1, \ldots, n$ the *law of total probability* states that

$$P(A \cap B) = \sum_{i=1}^{n} P(B_i) P(A|B_i).$$

Writing the right-hand side as an integral gives the equivalent expression

$$P(A \cap B) = \int_B \sum_{i=1}^{n} P(A|B_i) \chi_{B_i} \, \mathrm{d}P, \tag{C.1}$$

where χ_{B_i} is one on B_i and zero on $\Theta \setminus B_i$. The integrand is a $\mathcal{B}$-measurable step function on Θ. Noticing that relation (C.1) holds for arbitrarily fine partitions of B into sets of positive measure and for all $B \in \mathcal{B}$ with $P(B) > 0$, one could ask whether there is a limiting function $\omega_A : \Theta \to [0,1]$ satisfying the same two properties as each of the step functions:

- ω_A is $\mathcal{B}$ measurable,
- $P(A \cap B) = \int_B \omega_A \, \mathrm{d}P$ for all $B \in \mathcal{B}$.

Applying Theorem C.1 to the restriction $\mu := P|_{\mathcal{B}}$ of P to $\mathcal{B}$ and to $\nu := P(A \cap \bullet)$, which is a finite measure on $\mathcal{B}$, we obtain a nonnegative $\mathcal{B}$-measurable function $f : \Theta \to [0,\infty]$ satisfying $P(A \cap B) = \int_B f \, \mathrm{d}P$ for all $B \in \mathcal{B}$, and one easily shows that $f \leq 1$ almost everywhere with respect to $P|_{\mathcal{B}}$. Thus, the following definition is correct.

Definition C.2. Let $(\Theta, \mathcal{A}, P)$ be a probability space, let $A \in \mathcal{A}$, and let $\mathcal{B} \subseteq \mathcal{A}$ be a sub-σ-algebra of $\mathcal{A}$. A random variable $\omega_{A|\mathcal{B}} : \Theta \to \mathbb{R}$ is called *conditional probability of A given $\mathcal{B}$* if $\omega_{A|\mathcal{B}}$ is $\mathcal{B}$-measurable and

$$P(A \cap B) = \int_B \omega_{A|\mathcal{B}} \, \mathrm{d}P \quad \text{for all } B \in \mathcal{B}.$$

Before we go on introducing conditional densities we want to clarify the connection between 'simple' conditional probabilities and the new concept. Obviously, for $B \in \mathcal{B}$ with $P(B) > 0$ we have

$$P(A|B) = \frac{1}{P(B)} \int_B \omega_{A|\mathcal{B}} \, \mathrm{d}P.$$

Now let $B \in \mathcal{B}$ be an atomic set of $\mathcal{B}$, that is, for each other set $C \in \mathcal{B}$ either $B \subseteq C$ or $B \subseteq \Theta \setminus C$ is true. Then the measurability of $\omega_{A|\mathcal{B}}$ implies that $\omega_{A|\mathcal{B}}$ is constant on B; denote the value of $\omega_{A|\mathcal{B}}$ on B by $\omega_{A|\mathcal{B}}(B)$. If $P(B) > 0$ then

$$P(A|B) = \frac{1}{P(B)} \int_B \omega_{A|\mathcal{B}} \, \mathrm{d}P = \omega_{A|\mathcal{B}}(B).$$

To give a more tangible interpretation of $\omega_{A|\mathcal{B}}$ assume that $\mathcal{B}$ is generated by a countable partition $\{B_1, B_2, \ldots\} \subseteq \mathcal{A}$ of Θ. Then the above statements on atomic sets imply

$$\omega_{A|\mathcal{B}} = \sum_{i \in \mathbb{N} : P(B_i) > 0} P(A|B_i) \chi_{B_i} \quad \text{a.e. on } \Theta,$$

which is quite similar to the integrand in (C.1).

As described in the previous section, by choosing a 'nice', that is, an in some sense continuous conditional probability for A given $\mathcal{B}$ we can also give meaning to conditional probabilities with respect to sets of probability zero.

C.4. Conditional distributions and conditional densities

Now, that we know how to express the conditional probability of some event $A \in \mathcal{A}$ with respect to a σ-algebra $\mathcal{B} \subseteq \mathcal{A}$, we want to extend the concept to families of events. In more detail, we want to pool conditional probabilities of all events in a σ-algebra $\mathcal{C} \subseteq \mathcal{A}$ by introducing a mapping on Θ taking values in the set of probability measures on $(\Theta, \mathcal{C})$.

Definition C.3. Let $(\Theta, \mathcal{A}, P)$ be a probability space and, let $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$ be sub-σ-algebras of $\mathcal{A}$. A mapping $W_{\mathcal{C}|\mathcal{B}} : \Theta \times \mathcal{C} \to [0, 1]$ is called *conditional distribution of $\mathcal{C}$ given $\mathcal{B}$* if for each $\theta \in \Theta$ the mapping $W_{\mathcal{C}|\mathcal{B}}(\theta, \bullet)$ is a probability measure on $\mathcal{C}$ and for each $C \in \mathcal{C}$ the mapping $W_{\mathcal{C}|\mathcal{B}}(\bullet, C)$ is a conditional probability of C given $\mathcal{B}$.

We do not need this definition in full generality. Thus, we give a slightly modified formulation adapted to σ-algebras $\mathcal{C}$ generated by some random variable.

Definition C.4. Let $(\Theta, \mathcal{A}, P)$ be a probability space and let $\mathcal{B} \subseteq \mathcal{A}$ be a sub-σ-algebra of $\mathcal{A}$. Further let $(X, \mathcal{A}_X)$ be a measurable space and let $\xi : \Theta \to X$ be a

random variable. A mapping $W_{\xi|\mathcal{B}} : \Theta \times \mathcal{A}_X \to [0,1]$ is called *conditional distribution of ξ given $\mathcal{B}$* if for each $\theta \in \Theta$ the mapping $W_{\xi|\mathcal{B}}(\theta, \bullet)$ is a probability measure on $\mathcal{A}_X$ and for each $C \in \mathcal{A}_X$ the mapping $W_{\xi|\mathcal{B}}(\bullet, C)$ is a conditional probability of $\xi^{-1}(C)$ given $\mathcal{B}$.

The requirement that $W_{\xi|\mathcal{B}}(\bullet, C)$ is a conditional probability of $\xi^{-1}(C)$ for all $C \in \mathcal{A}_X$ can be easily fulfilled because conditional probabilities always exist. The main point lies in demanding that $W_{\xi|\mathcal{B}}(\theta, \bullet)$ is a probability measure for each $\theta \in \Theta$. To state a theorem on existence of conditional distributions we need some preparation.

Definition C.5. Two measurable spaces are called *isomorphic* if there exists a bijective mapping φ between them such that both φ and φ^{-1} are measurable.

Definition C.6. A measurable space is called *Borel space* if it is isomorphic to some measurable space $(A, \mathcal{B}(A))$, where A is a Borel set in $[0,1]$ and $\mathcal{B}(A)$ is the σ-algebra of Borel subsets of A.

Proposition C.7. *Every Polish space, that is, every separable complete metric space, is a Borel space. A product of a countable number of Borel spaces is a Borel space. Every measurable subset A of a Borel space $(B, \mathcal{B})$ equipped with the σ-algebra of all subsets of A lying in $\mathcal{B}$ is a Borel space.*

The following theorem gives a sufficient condition for the existence of conditional distributions.

Theorem C.8. *Let $(\Theta, \mathcal{A}, P)$ be a probability space and let $\mathcal{B} \subseteq \mathcal{A}$ be a sub-σ-algebra of $\mathcal{A}$. Further let $(X, \mathcal{A}_X)$ be a Borel space and let $\xi : \Theta \to X$ be a random variable. Then ξ has a conditional distribution given $\mathcal{B}$. If W_1 and W_2 are two conditional distributions of ξ given $\mathcal{B}$ then $W_1(\theta, \bullet) = W_2(\theta, \bullet)$ for almost all $\theta \in \Theta$.*

As described in Section C.2 looking at probability measures via densities allows more comfortable handling of events of probability zero. Thus, we introduce conditional densities.

Definition C.9. Let $(\Theta, \mathcal{A}, P)$ be a probability space and let $\mathcal{B} \subseteq \mathcal{A}$ be a sub-σ-algebra of $\mathcal{A}$. Further, let $(X, \mathcal{A}_X, \mu_X)$ be a σ-finite measure space and let $\xi : \Theta \to X$ be a random variable. A measurable function $p_{\xi|\mathcal{B}} : \Theta \times X \to [0, \infty)$ on the product space $(\Theta \times X, \mathcal{A} \otimes \mathcal{A}_X)$ is called *conditional density of ξ with respect to μ_X given $\mathcal{B}$* if $W_{\xi|\mathcal{B}} : \Theta \times \mathcal{A}_X \to [0,1]$ defined by

$$W_{\xi|\mathcal{B}}(\theta, C) := \int_C p_{\xi|\mathcal{B}}(\theta, \bullet)\, \mathrm{d}\mu_X \quad \text{for } \theta \in \Theta \text{ and } C \in \mathcal{A}_X$$

is a conditional distribution of ξ given $\mathcal{B}$.

Considering two random variables over a common probability space we can give an explicit formula for a conditional density of one random variable given the σ-algebra generated by the other one.

Proposition C.10. *Let $(\Theta, \mathcal{A}, P)$ be a probability space, assume that $(X, \mathcal{A}_X, \mu_X)$ and $(Y, \mathcal{A}_Y, \mu_Y)$ are σ-finite measure spaces, and let $\xi : \Theta \to X$ and $\eta : \Theta \to Y$ be random variables. Further, assume that the random variable (ξ, η) taking values in the product space $(X \times Y, \mathcal{A}_X \otimes \mathcal{A}_Y)$ has a density $p_{(\xi,\eta)} : X \times Y \to [0, \infty)$ with respect to the product measure $\mu_X \otimes \mu_Y$ on $\mathcal{A}_X \otimes \mathcal{A}_Y$. Then, setting*

$$p_\xi(x) := \int_Y p_{(\xi,\eta)}(x, \bullet)\, \mathrm{d}\mu_Y \quad \text{and} \quad p_\eta(y) := \int_X p_{(\xi,\eta)}(\bullet, y)\, \mathrm{d}\mu_X$$

for $x \in X$ and $y \in Y$, the function $p_{\xi|\eta} : \Theta \times X \to [0, \infty)$ defined by

$$p_{\xi|\eta}(\theta, x) := \begin{cases} \dfrac{p_{(\xi,\eta)}(x, \eta(\theta))}{p_\eta(\eta(\theta))} & \text{if } p_\eta(\eta(\theta)) > 0, \\ p_\xi(x) & \text{if } p_\eta(\eta(\theta)) = 0 \end{cases}$$

is a conditional density of ξ with respect to μ_X given $\sigma(\eta)$.

Reversing the roles of ξ and η in this proposition we get

$$p_{\eta|\xi}(\theta, y) := \begin{cases} \dfrac{p_{(\xi,\eta)}(\xi(\theta), y)}{p_\xi(\xi(\theta))} & \text{if } p_\xi(\xi(\theta)) > 0, \\ p_\eta(y) & \text{if } p_\xi(\xi(\theta)) = 0 \end{cases}$$

as a conditional density of η with respect to μ_Y given $\sigma(\xi)$. Combining the formulas for $p_{\xi|\eta}$ and $p_{\eta|\xi}$ we arrive at *Bayes' formula* for densities: For all $\theta \in \Theta$ fulfilling $p_\xi(\xi(\theta)) > 0$ the equation

$$p_{\xi|\eta}(\theta, \xi(\theta)) = \begin{cases} \dfrac{p_{\eta|\xi}(\theta, \eta(\theta)) p_\xi(\xi(\theta))}{p_\eta(\eta(\theta))} & \text{if } p_\eta(\eta(\theta)) > 0, \\ p_\xi(\xi(\theta)) & \text{if } p_\eta(\eta(\theta)) = 0 \end{cases}$$

is satisfied.

Although θ appears explicitly as an argument of $p_{\xi|\eta}$ the above proposition implies that $p_{\xi|\eta}$ depends only on $\eta(\theta)$ instead of θ itself. Thus, assuming that η is surjective, the definition

$$p_{\xi|\eta=y}(x) := p_{\xi|\eta}\big(\eta^{-1}(y), x\big) \quad \text{for } x \in X \text{ and } y \in Y$$

is viable. The same reasoning justifies the analogue definition

$$p_{\eta|\xi=x}(y) := p_{\eta|\xi}\big(\xi^{-1}(x), y\big) \quad \text{for } x \in X \text{ and } y \in Y,$$

if ξ is surjective. Using this notation Bayes' formula reads as

$$p_{\xi|\eta=y}(x) = \begin{cases} \dfrac{p_{\eta|\xi=x}(y) p_\xi(x)}{p_\eta(y)} & \text{if } p_\eta(y) > 0, \\ p_\xi(x) & \text{if } p_\eta(y) = 0 \end{cases} \tag{C.2}$$

for all $x \in X$ satisfying $p_\xi(x) > 0$ and for all $y \in Y$.

Eventually, we state a last relation, which is of use, too: The definition of $p_{\eta|\xi}$ implies

$$p_{(\xi,\eta)}(x, y) = p_\xi(x) p_{\eta|\xi=x}(y) \tag{C.3}$$

for all $x \in X$ satisfying $p_\xi(x) > 0$ and for all $y \in Y$.

D. The Lambert W function

In this chapter we briefly summarize some properties of the so called *Lambert W function* which occurs in Subsections 13.5.2 and 13.6.3. The material presented here can be found in [CGH+96].

The Lambert W function at a point $t \geq -\frac{1}{\mathrm{e}}$ is defined as the solution of

$$s\mathrm{e}^s = t. \tag{D.1}$$

We only consider real solutions s. For $t \geq 0$ and $t = -\frac{1}{\mathrm{e}}$ there is exactly one real solution. For $t \in (-\frac{1}{\mathrm{e}}, 0)$ there are two solutions. Thus we define two different W functions. By $W(t)$ for $t \in [-\frac{1}{\mathrm{e}}, \infty)$ we denote the solution of (D.1) which satisfies $s \geq -1$ and by $W_{-1}(t)$ for $t \in [-\frac{1}{\mathrm{e}}, 0)$ we denote the solution of (D.1) which satisfies $s \leq -1$. The functions W and W_{-1} are depicted in Figure D.1.

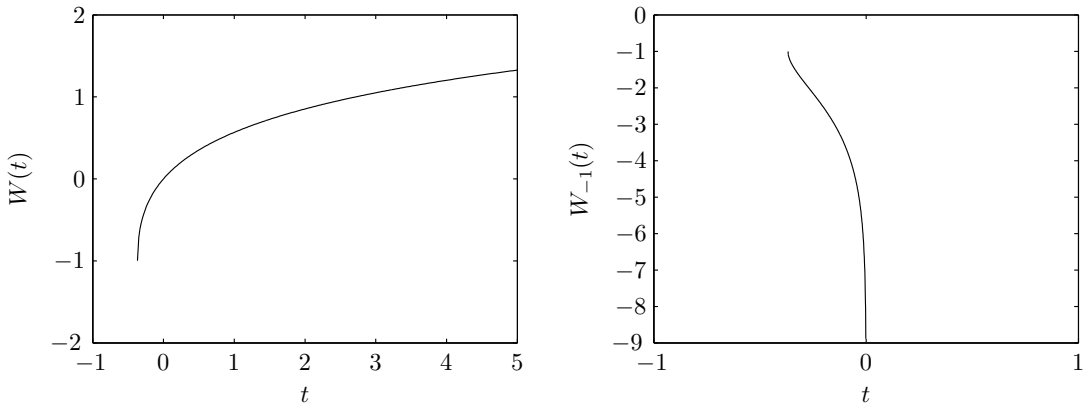

Figure D.1.: Branches of the Lambert W function; function W (left) and function W_{-1} (right).

Some specific values of W and W_{-1} are

$$W(-\tfrac{1}{\mathrm{e}}) = W_{-1}(-\tfrac{1}{\mathrm{e}}) = -1 \quad \text{and} \quad W(0) = 0.$$

Both functions are continuous and

$$\lim_{t\to\infty} W(t) = +\infty \quad \text{and} \quad \lim_{t\to-0} W_{-1} = -\infty.$$

The derivatives of W and W_{-1} are given by

$$W'(t) = \frac{W(t)}{t(1+W(t))} \quad \text{and} \quad W'_{-1}(t) = \frac{W_{-1}(t)}{t(1+W_{-1}(t))}$$

for $t \notin \{-\frac{1}{\mathrm{e}}, 0\}$ and $W'(0) = 1$.

The asymptotic behavior of W and W_{-1} can be expressed in terms of logarithms: For each $\varepsilon > 0$ there are $\bar{t} > 1$ and $\bar{t}_{-1} < 0$ such that

$$(1-\varepsilon)\ln t \leq W(t) \leq (1+\varepsilon)\ln t \quad \text{for all } t \geq \bar{t} \tag{D.2}$$

and

$$(1+\varepsilon)\ln(-t) \leq W_{-1}(t) \leq (1-\varepsilon)\ln(-t) \quad \text{for all } t \in [\bar{t}_{-1}, 0). \tag{D.3}$$

These estimates are a direct consequence of

$$\lim_{t\to\infty} \frac{W(t)}{\ln t} = 1 \quad \text{and} \quad \lim_{t\to -0} \frac{W_{-1}(t)}{\ln(-t)} = 1,$$

which can be seen with the help of l'Hôpital's rule.

Theses

1. Many problems from natural sciences, engineering, and finance can be formulated as an equation
$$F(x) = y, \quad x \in X, \quad y \in Y,$$
in topological spaces X and Y. If this equation is ill-posed then one has to apply regularization methods to obtain a stable approximation to the exact solution. One variant of regularization are Tikhonov-type methods
$$T_\alpha^z(x) := S(F(x), z) + \alpha\Omega(x) \to \min_{x \in X}$$
with fitting functional $S : Y \times Z \to [0, \infty]$, stabilizing functional $\Omega : X \to (-\infty, \infty]$, and regularization parameter $\alpha > 0$. The element $z \in Z$ represents a measurement of the exact right-hand side $y \in Y$. Typcially $Y \neq Z$ since Y is infinite dimensional but the data space Z is of finite dimension in practice. Due to the ill-posedness detailed handling of data errors (noise) by the model is indispensable. The very general Tikhonov-type approach investigated in the thesis allows for improved models of practical problems.

2. Typically the analysis of Tikhonov-type methods in Banach spaces with norm based fitting functionals relies on the triangle inequality. But also for non-metric fitting functionals, that is, fitting functionals which do not satisfy a triangle inequality, important analytic results can be obtained. In particular, it is possible to prove convergence rates for the convergence of regularized solutions $x_\alpha^z \in \operatorname{argmin}_{x \in X} T_\alpha^z(x)$ to exact solutions $x^\dagger$ of $F(x) = y$ if the noise becomes small and if the regularization parameter α is chosen appropriately. Next to a priori parameter choices also the discrepancy principle can be applied for choosing the regularization parameter.

3. The convergence rates result is based on a special variational inequality, which combines all assumptions on the exact solution $x^\dagger$, on the mapping F, on the functionals S and Ω, as well as on the spaces X, Y, and Z in one inequality. In principle the approach is known in the literature, but the new and extremely general form presented in this thesis covers also non-metric fitting functionals and arbitrary measures for expressing the rates.

4. One example which benefits from the general form of Tikhonov-type methods under consideration is regularization with Poisson distributed data. Typical applications where the measured data follows a Poisson distribution can be found in the fields of astronomical and medical imaging. Statistical methods motivate the minimization of a Tikhonov-type functional with non-metric fitting term for approximately solving such problems. The general theory developed in the thesis

applies to this specific Tikhonov-type functional and thus convergence rates can be proven.

5. The minimization of Tikhonov-type functionals involving non-metric fitting terms is challenging but not impossible. Algorithms for solving the Tikhonov-type minimization problem with Poisson distributed data in the case that Ω represents a sparsity constraint with respect to a Haar base can be formulated and implemented. A comparison of the obtained minimizers with the solutions of a norm based Tikhonov-type approach shows that the non-metric fitting functional has advantageous influence on the minimizers.

6. The sufficient condition for obtaining convergence rates in the general setting reduces to a form already known in the literature as variational inequality if Banach spaces and norm based fitting functionals are considered. There exist several other conditions yielding convergence rates. Next to variational inequalities these are source conditions, projected source conditions, approximate source conditions, and approximate variational inequalities. Only few relations between these conditions are given in the literature. Most results only state that a source condition implies one of the other conditions. But extensive investigations reveal much stronger connections between the different concepts.

7. Variational inequalities and approximate variational inequalities are equivalent concepts in Banach spaces. Approximate source conditions contain almost the same information as (approximate) variational inequalities and in Hilbert spaces approximate source conditions are equivalent to (approximate) variational inequalities. Further, source conditions and projected source conditions are an equivalent formulation of certain variational inequalities in Banach spaces.

8. Approximate source conditions in Hilbert spaces (and thus also the equivalent concepts) yield not only upper bounds for the regularization error but also lower bounds. Under suitable assumptions both bounds coincide up to a constant. Thus, approximate source conditions provide better estimates for the regularization error than source conditions. This result is also true for more general linear regularization methods in Hilbert spaces.

Symbols and notations

All occurring integrals are Lebesgue integrals. Usually we write

$$\int_A f \, \mathrm{d}\mu$$

for the integral of the function f over the set A with respect to the measure μ. If necessary, we indicate the variable to which the integral sign refers:

$$\int_A f(t) \, \mathrm{d}\mu(t).$$

In case that μ is the Lebesgue measure on $\mathbb{R}$ we also write

$$\int_A f(t) \, \mathrm{d}t.$$

We frequently work with convergent sequences $(x_k)_{k\in\mathbb{N}}$ with limit x. If confusion can be excluded we leave out the 'for $k \to \infty$' and write only $x_k \to x$. Further, we do not indicate the topology which stands behind the convergence. Only if it is not clear from the context we use '$\xrightarrow{\tau}$' instead of '$\to$' (with τ denoting the topology).

Throughout the thesis we use the following symbols:

X, Y, Z	topological spaces, Banach spaces, Hilbert spaces
x, y, z	elements of the spaces X, Y, and Z, respectively
τ_X, τ_Y, τ_Z	topologies on the spaces X, Y, and Z, respectively
Θ	sampling set of a probability space
θ	element of Θ
X^*, Y^*	topological duals of X and Y
ξ, η, ζ	elements of X^*, Y^*, and Z^*, respectively, or random variables taking values in X, Y, and Z, respectively
$\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$	σ-algebras
P, μ, ν	measures
F, A	mapping or operator of the equation to be solved
$\mathcal{R}(f)$	range of a mapping f
$D(f)$	domain or essential domain of a mapping f
$\overline{A}$	closure of a set A
f^*	conjugate function of a function f

In the context of Tikhonov regularization the following symbols frequently occur:

S	fitting functional
Ω	stabilizing functional
α	regularization parameter
T_α^z	Tikhonov functional with data element z
x_α^z	regularized solution with data element z
δ	noise level
y^δ, z^δ	noisy data

When discussing convergence rates we further use:

M_Ω	sublevel set of Ω
$\partial\Omega(x)$	subdifferential of Ω at $x \in X$
$B_\xi^\Omega(\tilde{x}, x)$	Bregman distance of $\tilde{x}, x \in X$ with respect to $\xi \in \partial\Omega(x)$
D_{y^0}	measure for data error
$E_{x^\dagger}$	measure for solution error
d, D_β, d_ψ, $D_{\psi,\beta}$	distance functions
φ, ψ, θ	(index) functions
M	domain of a variational inequality

Occurring standard symbols are:

$L^1(T, \mu)$	space of μ-integrable functions over the set T
$L^2(0,1)$	space of Lebesgue integrable functions over $(0,1)$ for which the squared function has finite integral
$L^\infty(0,1)$	space of Lebesgue integrable functions over $(0,1)$ which are essentially bounded
$H^1(0,1)$	space of functions from $L^2(0,1)$ which have a generalized derivative in $L^2(0,1)$
$l^2(\mathbb{N})$	space of sequences for which the series of squares converges
$\mathbb{N}$	natural numbers without zero
$\mathbb{N}_0$	natural numbers including zero
$\mathbb{Z}$	integers
$\mathbb{R}$	real numbers
$B_r(x)$	open ball with center x and radius r
$\overline{B}_r(x)$	closed ball with center x and radius r

Bibliography

[ABM06] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*. MPS–SIAM series on optimization. Society for Industrial and Applied Mathematics and the Mathematical Programming Society, Philadelphia, 2006.

[AR11] S. W. Anzengruber and R. Ramlau. Convergence rates for Morozov's Discrepancy Principle using Variational Inequalities. RICAM report 2011-06, Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria, 2011.

[Bar08] J. M. Bardsley. An efficient computational method for total variation-penalized poisson likelihood estimation. *Inverse Problems and Imaging*, 2(2):167–185, 2008.

[BB88] J. Barzilai and J. M. Borwein. Two-Point Step Size Gradient Methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988.

[BB09] M. Benning and M. Burger. Error estimation for variational models with non-Gaussian noise. Technical report, WWU Münster, Münster, Germany, 2009. http://wwwmath.uni-muenster.de/num/publications/2009/BB09.

[BCW08] R. I. Boţ, E. R. Csetnek, and G. Wanka. Regularity Conditions via Quasi-Relative Interior in Convex Programming. *SIAM Journal on Optimization*, 19(1):217–233, 2008.

[BGW09] R. I. Boţ, S.-M. Grad, and G. Wanka. *Duality in Vector Opimization*. Vector Optimization. Springer, Berlin, Heidelberg, 2009.

[BH09] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for nonlinear ill-posed problems. *arXiv.org*, arXiv:0909.5093v1 [math.NA], September 2009. http://arxiv.org/abs/0909.5093.

[BH10] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22(3):369–392, 2010.

[BL08] K. Bredies and D. A. Lorenz. Linear Convergence of Iterative Soft-Thresholding. *Journal of Fourier Analysis and Applications*, 14(5–6):813–837, 2008.

[BL09] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011 (14pp), 2009.

[BO04] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

[Bon09] T. Bonesky. Morozov's discrepancy principle and Tikhonov-type functionals. *Inverse Problems*, 25(1):015015 (11pp), 2009.

[BP86] V. Barbu and Th. Precupanu. *Convexity and Optimization in Banach Spaces*. Mathematics and its Applications. East European Series. Editura Academiei, D. Reidel Publishing Company, Bucharest, Dordrecht, 2nd English edition, 1986.

[BZZ09] S. Bonettini, R. Zanella, and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse Problems*, 25(1):015002 (23pp), 2009.

[CDB05] F. Cammaroto and B. Di Bella. Separation Theorem Based on the Quasirelative Interior and Application to Duality Theory. *Journal of Optimization Theory and Applications*, 125(1):223–229, 2005.

[CGH+96] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996.

[CK94] G. Chavent and K. Kunisch. Convergence of Tikhonov regularization for constrained ill-posed inverse problems. *Inverse Problems*, 10(1):63–76, 1994.

[Dau92] I. Daubechies. *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[DDDM04] I. Daubechies, M. Defrise, and C. De Mol. An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[Egg93] P. P. B. Eggermont. Maximum Entropy Regularization for Fredholm Integral Equations of the First Kind. *SIAM Journal on Mathematical Analysis*, 24(6):1557–1576, 1993.

[EHN96] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[Eps08] C. L. Epstein. *Introduction to the Mathematics of Medical Imaging*. Society for Industrial and Applied Mathematics, Philadelphia, 2nd edition, 2008.

[FG97] B. Fristedt and L. Gray. *A Modern Approach to Probability Theory*. Probability and its Applications. Birkhäuser, Boston, 1997.

[FH10] J. Flemming and B. Hofmann. A New Approach to Source Conditions in Regularization with General Residual Term. *Numerical Functional Analysis and Optimization*, 31(3):254–284, 2010.

[FH11] J. Flemming and B. Hofmann. Convergence rates in constrained Tikhonov regularization: equivalence of projected source conditions and variational inequalities. *Inverse Problems*, 27(8):085001 (11pp), 2011.

[FHM11] J. Flemming, B. Hofmann, and P. Mathé. Sharp converse results for the regularization error using distance functions. *Inverse Problems*, 27(2):025006 (18pp), 2011.

[Fle10a] J. Flemming. Theory and examples of variational regularization with non-metric fitting functionals. *Journal of Inverse and Ill-posed Problems*, 18(6):677–699, 2010.

[Fle10b] J. Flemming. Theory and examples of variational regularization with non-metric fitting functionals. Preprint series of the Department of Mathematics 2010-14, Chemnitz University of Technology, Chemnitz, Germany, July 2010.

[Fle11] J. Flemming. Solution smoothness of ill-posed equations in Hilbert spaces: four concepts and their cross connections. *Applicable Analysis*, 2011. DOI 10.1080/00036811.2011.563736. To appear.

[FSBM10] M. J. Fadili, J.-L. Starck, J. Bobin, and Y. Moudden. Image Decomposition and Separation Using Sparse Representations: An Overview. *Proceedings of the IEEE*, 98(6):983–994, 2010.

[GC99] J. C. Goswami and A. K. Chan. *Fundamentals of Wavelets*. Wiley Series in Microwave and Optical Engineering. John Wiley & Sons, New York, 1999.

[Gei09] J. Geißler. Studies on Convergence Rates for the Tikhonov Regularization with General Residual Functionals. Diploma thesis, Chemnitz University of Technology, Chemnitz, Germany, July 2009. In German.

[GHS09] M. Grasmair, M. Haltmeier, and O. Scherzer. The Residual Method for Regularizing Ill-Posed Problems. *arXiv.org*, arXiv:cs/0905.1187v1 [math.OC], May 2009. http://arxiv.org/abs/0905.1187.

[Gil10] G. L. Gilardoni. On Pinsker's and Vajda's Type Inequalities and Csiszár's f-Divergences. *IEEE Transactions on Information Theory*, 56(11):5377–5386, 2010.

[Gra10a] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems*, 26(11):115014 (16pp), 2010.

[Gra10b] M. Grasmair. Generalized Bregman Distances and Convergence Rates for Non-convex Regularization Methods. Industrial Geometry report 104, University of Vienna, Vienna, Austria, May 2010.

[Gra10c] M. Grasmair. Non-convex sparse regularisation. *Journal of Mathematical Analysis and Applications*, 365(1):19–28, 2010.

[Gun06] A. Gundel. *Robust Utility Maximization, f-Projections, and Risk Constraints*. PhD thesis, Humboldt-Universität Berlin, Berlin, Germany, February 2006.

[Hei08a] T. Hein. Convergence rates for multi-parameter regularization in Banach spaces. *International Journal of Pure and Applied Mathematics*, 43(4):593–614, 2008.

[Hei08b] T. Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Problems*, 24(4):045007 (10pp), 2008.

[Hei09] T. Hein. Tikhonov regularization in Banach spaces—improved convergence rates results. *Inverse Problems*, 25(3):035002 (18pp), 2009.

[HH09] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035033 (16pp), 2009.

[HK05] B. Hofmann and R. Krämer. On maximum entropy regularization for a specific inverse problem of option pricing. *Journal of Inverse and Ill-posed Problems*, 13(1):41–63, 2005.

[HKPS07] 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.

[HM07] B. Hofmann and P. Mathé. Analysis of Profile Functions for General Linear Regularization Methods. *SIAM Journal on Numerical Analysis*, 45(3):1122–1141, 2007.

[HMvW09] B. Hofmann, P. Mathé, and H. von Weizsäcker. Regularization in Hilbert space under unbounded operators and general source conditions. *Inverse Problems*, 25(11):115013 (15pp), 2009.

[Hof06] B. Hofmann. Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators. *Mathematical Methods in the Applied Sciences*, 29(3):351–371, 2006.

[Hoh97] T. Hohage. Logarithmic convergence rates of the iteratively regularized Gauss–Newton method for an inverse potential and an inverse scattering problem. *Inverse Problems*, 13(5):1279–1299, 1997.

[Hoh00] T. Hohage. Regularization of exponentially ill-posed problems. *Numerical Functional Analysis and Optimization*, 21(3&4):439–464, 2000.

[How06] S. B. Howell. *Handbook of CCD Astronomy*. Number 5 in Cambridge Observing Handbooks for Research Astronomers. Cambridge University Press, Cambridge, 2nd edition, 2006.

[HSvW07] B. Hofmann, M. Schieck, and L. von Wolfersdorf. On the analysis of distance functions for linear ill-posed problems with an application to the integration operator in L^2. *Journal of Inverse and Ill-posed Problems*, 15(1):83–98, 2007.

[HW11] T. Hohage and F. Werner. Iteratively regularized Newton methods with general data misfit functionals and applications to Poisson data. *arXiv.org*, arXiv:1105.2690v1 [math.NA], May 2011. http://arxiv.org/abs/1105.2690.

[HY10] B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Applicable Analysis*, 89(11):1705–1727, 2010.

[IJT10] K. Ito, B. Jin, and T. Takeuchi. A Regularization Parameter for Nonsmooth Tikhonov Regularization. Preprint 2010-03, Graduate School of Mathematical Sciences, University of Tokyo, Tokyo, Japan, March 2010.

[JZ10] B. Jin and J. Zou. Iterative Parameter Choice by Discrepancy Principle. Technical report 2010-02 (369), Department of Mathematics, Chinese University of Hong Kong, Shatin, Hong Kong, February 2010.

[Kan03] S. Kantorovitz. *Introduction to Modern Analysis*. Number 8 in Oxford Graduate Texts in Mathematics. Oxford University Press, Oxford, New York, 2003.

[KH10] B. Kaltenbacher and B. Hofmann. Convergence rates for the iteratively regularized Gauss–Newton method in Banach spaces. *Inverse Problems*, 26(3):035007 (21pp), 2010.

[Kie02] B. T. Kien. The Normalized Duality Mapping and Two Related Characteristic Properties of a Uniformly Convex Banch Space. *Acta Mathematica Vietnamica*, 27(1):53–67, 2002.

[KNS08] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Number 6 in Radon Series on Computational and Applied Mathematics. Walter de Gruyter, Berlin, 2008.

[KS05] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Number 160 in Applied Mathematical Sciences. Springer, New York, 2005.

[Mat08] P. Mathé. Lecture series on inverse problems. Personal lecture notes, 2008. Chemnitz University of Technology.

[MH08] P. Mathé and B. Hofmann. How general are general source conditions? *Inverse Problems*, 24(1):015009 (5pp), 2008.

[MP03a] P. Mathé and S. V. Pereverzev. Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(6):1263–1277, 2003.

[MP03b] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[Neu88] A. Neubauer. Tikhonov-regularization of ill-posed linear operator equations on closed convex sets. *Journal of Approximation Theory*, 53(3):304–320, 1988.

[Neu97] A. Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM Journal on Numerical Analysis*, 34(2):517–527, 1997.

[Neu09] A. Neubauer. On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces. *Inverse Problems*, 25(6):065009 (10pp), 2009.

[NHH+10] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn. Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces. *Applicable Analysis*, 89(11):1729–1743, 2010.

[NW06] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2nd edition, 2006.

[OQ04] C. Oliver and S. Quegan. *Understanding snythetic aperture radar images*. SciTech Publishing, Raleigh, 2004.

[Pös08] C. Pöschl. *Tikhonov Regularization with General Residual Term*. PhD thesis, University of Innsbruck, Innsbruck, Austria, October 2008. Corrected version.

[PRS05] C. Pöschl, E. Resmerita, and O. Scherzer. Discretization of variational regularization in Banach spaces. *Inverse Problems*, 26(10):105017 (18pp), 2005.

[RA07] E. Resmerita and R. S. Anderssen. Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems. *Mathematical Methods in the Applied Sciences*, 30(13):1527–1544, 2007.

[Res05] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[SGG+09] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Number 167 in Applied Mathematical Sciences. Springer, New York, 2009.

[TA76] A. N. Tikhonov and V. Arsénine. *Méthodes de résolution de problèmes mal posés*. Editions Mir, Moscou, 1976. In French.

[TLY98] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola. *Nonlinear Ill-posed Problems*. Number 14 in Applied Mathematics and Mathematical Computation. Chapman & Hall, London, 1998.

[Wil] S. Wilhelm. Confocal Laser Scanning Microscopy. Technical report, Carl Zeiss MicroImaging GmbH, Jena, Germany.

[Yos95] K. Yosida. *Functional Analysis*. Springer, Berlin, Heidelberg, New York, 6th edition, 1995.

[ZBZB09] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero. Efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems*, 25(4):045010 (24pp), 2009.

[Zei85] E. Zeidler. *Nonlinear Functional Analysis and its Applications III. Variational Methods and Optimization*. Springer, New York, Berlin, Heidelberg, 1985.

[Zor04] A. V. Zorich. *Mathematical Analysis I*. Springer, Berlin, Heidelberg, 2004.