Karlsruher Institut für Technologie

On Inexact Newton Methods for Inverse Problems in Banach Spaces

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von
M.Sc. Fábio J. Margotti
aus Nova Veneza

Tag der mündlichen Prüfung: 15. Juli 2015

Referent: Prof. Dr. Andreas Rieder
Korreferent: Prof. Dr. Andreas Kirsch

I dedicate this work to my beloved wife PATRÍCIA

Contents

Acknowledgements

I am very glad to express the most sincere gratitude to my advisor Prof. Dr. Andreas Rieder for trusting in my work since its beginning. His excellent mathematical skills have strongly contributed to the improvement of my knowledge and his constant explanations about German language and culture greatly facilitated my everyday life. I would like to thank Prof. Dr. Andreas Kirsch too, for being a referee of this work, for kindly writing recommendations to support the extension of my scholarship in two occasions and for helping to complement my mathematical formation with his useful lectures as well as with his fruitful discussions in our work group on inverse problems.

This work would have been impossible without the important financial support provided by the *Deutscher Akademischer Austauschdienst.* I am very thankful to DAAD for introducing me to the interesting German's culture too. In particular, I would like to say thank you to Mrs. Maria Salgado for being so supportive every time I needed. I am much obliged to my previous advisor Prof. Dr. Antonio Leitão, who introduced me to this fascinating area of inverse problems and provided a strong support for my coming to Germany.

It is very necessary to thank all my colleagues of my research group of scientific computing IANM3 for creating a very pleasant and productive work environment. Special thanks to Sonja Becker, Nathalie Sonnefeld, Johannes Ernesti, Julian Krämmer, Daniel Weiß, Marcel Mikl, Ramin S. Nejad, Lydia Wagner, professors Tobias Jahnke and Christian Wieners and my ex-colleagues Tudor Udrescu and Tim Kreutzmann.

My gratefulness is also directed to my colleague Robert Winkler for his always useful discussions concerning the inverse problem of Electrical Impedance Tomography as well as for providing me with his very organized and well-written code of EIT-CEM. Further, I want to thank Ekkachai Thawinan and Andreas Schulz for being good friends in these last four years.

Last but not least, I am deeply thankful to my dear wife Patrícia, who has been always very considerate and comprehensive and with whom I have shared each little moment in this amazing experience in Germany. Muito obrigado!

Abstract

This monograph is concerned with the investigation of an inexact-Newton algorithm and its adaptation to the Kaczmarz methods. The new iterative methods aim the construction of regularized solutions for nonlinear inverse ill-posed problems in Banach spaces.

To perform an iteration, the Newton-type algorithm linearizes the original problem around the current iterate and then applies a regularization technique in order to stably solve the resulting linear system. Various methods are proposed to the task of solving the generated linear system and a complete convergence analysis is carried out. Furthermore, stability and regularization properties of the computed solutions are established.

The new methods are tested in the nonlinear and highly ill-posed inverse problem of Electrical Impedance Tomography in some numerical implementations realized at the end of the work. The experiments simultaneously provide the necessary support to the theoretical results and contribute to a relevant improvement to the solutions of the referred inverse problem.

Key words: Inverse problems, Ill-posed problems, Regularization theory, Iterative methods, Inexact-Newton methods, Electrical Impedance Tomography, Banach spaces.

Chapter 1

Introduction

This work is dedicated to the study of regularization methods for obtaining stable solutions of nonlinear inverse ill-posed problems in Banach spaces. It is focused on iterative methods and places particular emphasis on Newton-type algorithms. Further, a convergence analysis of the investigated methods is provided and some numerical implementations support the theoretical results.

In order to properly introduce the classical definition of ill-posedness due to Hadamard, we start defining a *forward problem* as a function which associates the *cause* of a specific phenomenon with the *effect* produced by it. The effect of this phenomenon is usually called the *solution of the forward problem*. To each forward problem belongs a correlated inverse problem.

The *inverse problems* consist of a set of mathematical problems where one aims to determine the cause of a particular phenomenon (solution of the inverse problem) by means of observing the effect produced by it (data). This kind of procedure encompasses an extensive amount of real problems described by mathematical models in the most diversified fields such as medicine, astronomy, engineering, geology, meteorology and a large sort of physical problems as well as image restoration and several other applications [30, 45, 48]. The biggest challenge in the resolution of such problems is related to the fact that they are frequently ill-posed in the sense of Hadamard.

The French mathematician Jacques Hadamard [19] has defined a problem as *well-posed* if for this problem:

1. A solution exists (existence);
2. There exists at most one solution (uniqueness);
3. The solution depends continuously on the data (stability).

A problem which is not well-posed is called *ill-posed*.

If the data space is defined as the set of the solutions of the forward problem, existence of a solution of the inverse problem is clear. However, a solution corresponding to a measured data may fail to exist if this data is corrupted by noise. In this case, the solution space should be enlarged in order to guarantee the existence of a solution. Typical examples of enlargement of the solution space are constructed by using the Moore-Penrose generalized inverse for linear inverse problems, and by accepting distributional solutions of partial differential equations in Sobolev spaces.

In contrast, a restriction on the set of the admissible solutions can be imposed with the purpose of ensuring uniqueness of a solution (the use of a minimal norm solution is a common example). In this case, some a priori information about the solution must be available. Non-uniqueness of solutions might also mean that insufficient data has been gathered and consequently more data needs to be observed.

The third item in the definition of Hadamard is certainly the most delicate to deal with. The non-fulfillment of the stability statement implies that unavoidable measurement and round-off errors can be amplified by an arbitrarily large factor, severely compromising the reliability of the computed solutions. Moreover, in contrast to the two first items of Hadamard's definition, a reformulation of the problem in order to achieve stability is not a trivial issue. The stability property depends on the topology of the involved spaces and an alteration of these topologies often modifies the original features of the problem and deprives it of its fundamental characteristics, or in other words, the reformulated problem becomes meaningless.

The observance of item 3 of Hadamard's definition on the other hand, signifies that the solution of the inverse problem corresponding to data corrupted by a low-level noise cannot be distant from the searched-for solution. Hadamard believed, just as many of his contemporaries, that a mathematical model could only correctly represent a natural phenomenon if the associated inverse problem was well-posed (*natura non facit saltum*). If this was not the case, the model was considered incorrectly formulated and the related problem was called ill-posed. This was a controversy perception until the beginning of last century, when was finally realized that an enormous quantity of real problems are actually ill-posed if translated in any reasonable mathematical model. This conclusion initiated in the second half of last century a huge amount of research targeting methods capable of stably solving inverse ill-posed problems. At this time, the *regularization theory* was born.

In practical situations one usually has access only to a perturbed version of the data, which is invariably contaminated by noise. As a consequence, the exact reconstruction of a solution is unattainable and therefore, the best it can be done is finding an approximate solution. However, unless a *regularization method*[1] is engaged, the ill-posedness phenomenon combined with even very low *noise levels* can easily ruin the process of computing a solution.

Due to both, its technological relevance and the challenging difficulties involving the development of regularization methods, inverse problems are still nowadays an active area of research. This fact is reflected in the large number of Journals, books and monographs devoted to this subject. A particular interesting group, widely applied in the resolution of large scale inverse ill-posed problems, is the iterative regularization methods. The literature in this area is vast and the books [30, 15, 45] can be duly included in the most complete references concerning the subject in Hilbert spaces. The work [29] deserves to be mentioned too. For more recent progress, including the regularization theory in Banach spaces, consult [48] and the references therein.

Known for their fast convergence properties, Newton-type methods are often regarded as robust algorithms for solving nonlinear inverse problems. Among all the members of this large group of remarkable iterative methods, we highlight the `REGINN` method. Introduced in 1999 by A. Rieder [43], the `REG`ularization based on `IN`exact `N`ewton iteration is a class of inexact-Newton methods, which solves nonlinear inverse ill-posed problems by means of solving a sequence of linearizations of the original problem. This algorithm linearizes the original problem around the current iterate and then applies an iterative regularization technique in the so-called *inner iteration* to find and approximate solution of the resulting linearized system. Afterwards, this approximate solution is added to the current iterate in the *outer iteration* in order to generate an update. Finally, the discrepancy principle is employed to terminate the outer iteration.

The properties of `REGINN` in Hilbert spaces are well-known. Convergence results, regularization property and convergence rates have been established under standard requirements and with different iterative regularization methods in the inner iteration [43, 44, 46, 25]. A

[1] If for each sequence of positive noise levels converging to zero, the correspondent sequence of approximate solutions converges to an exact solution of the problem, then the related method is called a *regularization method* and it is accordingly said to satisfy the *regularization property*.

few examples of possible inner iterations are given by gradient methods like the Conjugate Gradient, Steepest Descent and Landweber methods and by Tikhonov-like methods such as the Iterated-Tikhonov and Tikhonov-Phillips methods.

Perhaps the biggest disadvantage of Newton-like methods is the necessity of calculating the derivative of the forward problem. This is normally an expensive process and it is generally the bottleneck in numerical implementations. A wise approach to overcome this obstacle is the utilization of a *loping* strategy. This procedure, proposed by S. Kaczmarz in [27], splits the problem into a finite number of sub-problems, which are cyclically processed, using one sub-problem each time, in order to find a solution of the original problem. The Kaczmarz's technique reduces drastically the computational effort necessary to perform a single iteration. The expected effect is the acceleration of the original method with a consequent gain in the convergence speed. Kaczmarz methods have an additional advantage: they describe more appropriately the problems whose data is gathered by a set of individual measurements, which is the case in Electrical Impedance Tomography for example (see the numerical experiments in Chapter 5).

There are many papers concerning Kaczmarz methods in the context of ill-posed problems, but similar to `REGINN`, most results are proven in the Hilbert spaces framework [31, 4, 20, 7]. Several inverse problems however, are naturally described in more general contexts than Hilbert spaces. The Lebesgue spaces $L^p(\Omega)$, as well as the Sobolev spaces $W^{n,p}(\Omega)$ and the space of the $p-$summable sequences $\ell^p(\mathbb{N})$ for different values of $p \in [1, \infty]$ are classical examples of Banach spaces, which model more appropriately numerous inverse problems. Further, describing an inverse problem in more suitable Banach spaces contributes with supplementary benefits such as the non-destruction of sparsity characteristics of the searched-for solution and the increase of data accuracy when dealing with impulsive noise[2]. These improvements are commonly obtained changing in the solution and data spaces, the parameter $p = 2$, commonly used to characterize Hilbert spaces, with $p \approx 1$, see Daubechies et al [12].

Figure 1.1 illustrates the advantage of using more general Banach spaces in the description of inverse problems. The searched-for (sparse) electrical conductivity is reconstructed in the Electrical Impedance Tomography problem using 1% of impulsive noise and compares two different frameworks: in the first case, the solution space X and the data space Y are both the Hilbert space L^2, while in the second one, the conductivity is reconstructed in the Banach spaces $X = Y = L^{1.01}$.

For more examples of inverse problems modelled in Banach spaces, see the numerical experiments in Chapter 5 and the book of Schuster et al [48].

The difficulties of carrying out a convergence analysis in Banach spaces grow massively if the solution and data spaces have poor smoothness/convexity properties. It is not straightforward modifying classical methods from Hilbert spaces in order to adjust them to more complicated Banach spaces. As long as we know, the first version of `REGINN` in Banach spaces has been published by Q. Jin in [24], where a weak convergence result was proven. The first strong convergence result has been firstly successfully accomplished for the combination Kaczmarz/inexact-Newton/Banach spaces in our previous work [40], where the Landweber method was employed as inner iteration. The most recent progress has been made using the Iterated-Tikhonov method as inner iteration [39].

The present thesis contributes to expand the results above mentioned, providing a relative general convergence analysis of a Kaczmarz version of `REGINN` (in short `K-REGINN`) in Banach spaces, valid at the same time for different methods in the inner iteration. This work is structured as following:

- In Chapter 2, the necessary preliminaries concerning the geometry of Banach spaces

[2]Impulsive noise is a kind of sparsely distributed noise, common in many practical applications, see e.g. [11].

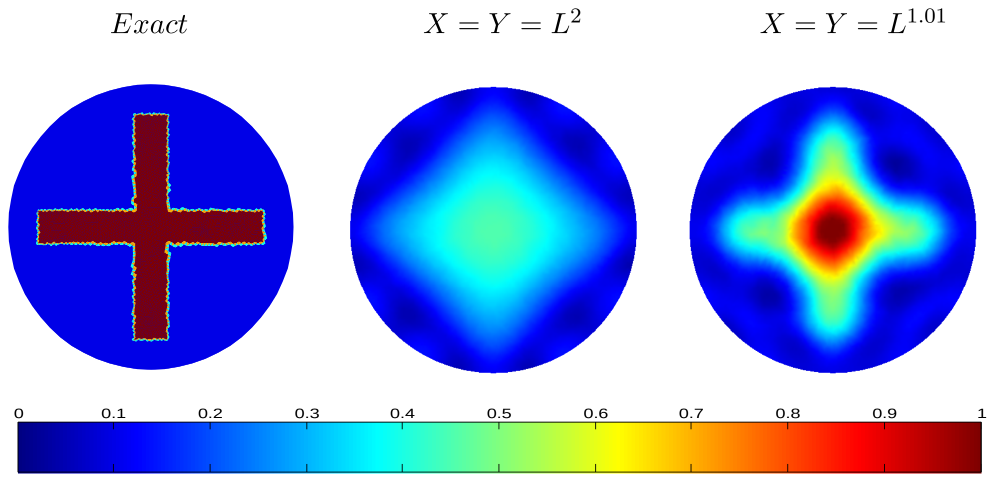


Figure 1.1: Sparse conductivity (picture on the left) reconstructed in the Electrical Impedance Tomography problem with 1% of impulsive noise. The Hilbert spaces $X = Y = L^2$ and the Banach spaces $X = Y = L^{1.01}$ have been used to reconstruct the pictures displayed on the middle and on the right respectively.

are presented. First, the concepts of convexity and smoothness are discussed. These concepts determine "how favorable" the geometry of a Banach space is. Hilbert spaces are usually regarded as the Banach spaces with the most favorable geometrical features, being at the same time the most convex and the smoothest Banach spaces. After a discussion about the connections between convexity and smoothness, the definition of *duality mapping* is introduced. Inspired by the Riesz Representation Theorem, this mapping associates each element x in the solution space with a subset $J_p(x)$ of its dual space so that the resulting duality pairing $\langle x^*, y \rangle$, with $x^* \in J_p(x)$, mimics the properties of the inner product $\langle x, y \rangle$ in Hilbert spaces. At the end of this chapter, the *Bregman distance* is presented. It serves to simulate in Banach spaces the polarization identity and replaces the norm in some specific situations. Finally, using the Xu-Roach famous Theorems [53], some connections between the Bregman distance and the standard norm are proven.

- The Kaczmarz version of `REGINN` is introduced in Chapter 3. In order to solve the inner iteration in Banach spaces, we carry out in Section 3.1 an adaptation of different regularization techniques from Hilbert to more general Banach spaces. The gradient methods are first considered. In Hilbert spaces, these algorithms update the current iterate adding a multiple of the gradient of the residual-squared-norm functional. The primal and dual gradient methods, presented respectively in Subsections 3.1.1 and 3.1.2, are the Banach space versions. The fundamental difference between these two kinds of gradient methods is the space where the iteration occurs: while the primal gradient methods have an iteration occurring in the original space, the dual version performs this iteration in the dual space. Among all the dual gradient methods of Subsection 3.1.2, we highlight the important Landweber, Modified Steepest Descent and Decreasing Error methods, with the last one being a new strategy, firstly presented in this work. In Subsection 3.1.3 the Tikhonov-like methods are introduced and special attention is put on to the study of the relevant Iterated-Tikhonov and Tikhonov-Phillips methods. It is well-known that Tikhonov-type methods provide further stability to the computations, however, the necessity of minimizing functionals in order to perform the inner iteration, substantially increases the effort of these methods. To simplify this task, a mixed version of Tikhonov and dual gradient methods is suggested in Subsection 3.1.4. This novel algorithm yields extra stability to the computations of the inner iteration without requiring the solution of an optimization problem.

- Most methods presented in Section 3.1 have a similar nature and share similar properties. This particular aspect makes possible a general convergence analysis, which is implemented in Chapter 4. The convergence analysis of `K-REGINN` presented in Chapter 4 is carried out without considering any specific method in the inner iteration. It is made assuming only specific properties of the sequence generated in the inner iteration, which makes this convergence analysis valid for every method having the required properties. Since many of the requested properties have already been proven for the methods presented in Section 3.1, the results of Chapter 4 hold true in particular for these methods. We highlight the main results: In Section 4.1 it is proven that `REGINN` terminates and has a decreasing residual behavior whenever a primal gradient method or a Tikhonov-Phillips-like method is used as inner iteration. For the remaining methods, we further provide in the subsequent sections the proofs of strong convergence of `REGINN` in the noiseless situation and of the regularization property, see Theorems 43 and 47. Moreover, a decreasing error behavior and a weak convergence result for the Kaczmarz version are provided in Theorem 38 and Corollary 41 respectively. Additionally, for the dual gradient Landweber method, Iterated-Tikhonov and Tikhonov-Phillips methods, strong convergence in the noiseless situation and the regularization property are proven for the Kaczmarz versions too.

- In Chapter 5, the performance of `K-REGINN` is tested for solving the inverse problem of the Electrical Impedance Tomography. Section 5.1 presents the Continuum Model [8] and the Complete Electrode Model [50] is presented in Section 5.2. At the beginning of each section, a brief explanation of the mathematical model is given and in the subsequent subsection, the evaluation of the derivatives is discussed. Further, some numerical experiments are performed in Subsections 5.1.2 and 5.2.2 and the respective results discussed.

Chapter 2

Geometry of Banach Spaces

This first chapter lists the most relevant concepts and facts concerning the geometry of Banach spaces. The main goals are to point out and explain the main results about convexity and smoothness of Banach spaces, duality mappings, Bregman distances and the properties connecting these notions. We only prove some few results and suggest some references where the remaining proofs can be found.

We consider the book of Cioranescu [10] a good reference. It presents in a very clear and understandable way the main ideas concerning the uniform convexity and uniform smoothness of Banach spaces and their connections with duality mappings. However, it lacks the important concepts of convexity and smoothness of power type. Moreover, the Cioranescu's book had been written before the very famous paper of Xu and Roach [53] was published, which means that the main relations between convexity/smoothness and the Bregman distances is not present. To understand this significant issue and learn its main results, we suggest beyond the paper of Xu and Roach itself, the article [47], the books [48, 9] and the references therein.

The theoretical results of this work cover simultaneously real and complex Banach spaces. However, as we will discuss in a moment, their dual spaces have analogous properties, which permits to assume without restriction of the generality that the Banach space in question is always a real Banach space. To make it clear, we give the following definition.

Definition 1 *Let X be a normed space defined over the field $\Bbbk$ (either $\mathbb{R}$ or $\mathbb{C}$). We call the set*

$$X^* := \{x^* \colon X \to \Bbbk : \ x^* \text{ linear and continuous}\}$$

the DUAL SPACE *of X. We write $X_{\mathbb{R}}$ to represent the vector space X regarded as a real vector space. Accordingly,*

$$\begin{aligned} X^*_{\mathbb{R}} &= \{x^* \colon X_{\mathbb{R}} \to \mathbb{R} : \ x^* \text{ linear and continuous}\} \\ &= \{x^* \colon X \to \mathbb{R} : \ x^* \ \mathbb{R}-\text{linear and continuous}\} \end{aligned}$$

denotes the dual space of $X_{\mathbb{R}}$. An element of X^* or $X^*_{\mathbb{R}}$ is called a FUNCTIONAL.

As $\mathbb{R}$ and $\mathbb{C}$ are Banach spaces, so is X^*, independent of X being a Banach space itself. Let X be a complex Banach space. It is easy to see that if $x^* \in X^*$, then the function $\operatorname{Re} x^* : X \to \mathbb{R}$ defined by

$$\langle \operatorname{Re} x^*, x\rangle := \operatorname{Re}\langle x^*, x\rangle \,, \ x \in X,$$

belongs to $X^*_{\mathbb{R}}$. Further, the operator $T : X^* \to X^*_{\mathbb{R}}$ defined by $Tx^* = \operatorname{Re} x^*$ satisfies

$$T(x^* + \lambda y^*) = Tx^* + \lambda T y^* \text{ for all } x^*, y^* \in X^* \text{ and } \lambda \in \mathbb{R},$$

which means that T is $\mathbb{R}-$linear.

Let $x^* \in X^*$ be fixed. We claim that for each $x \in X$, there exists a vector $\widehat{x} \in X$ such that

$$\|\widehat{x}\| = \|x\| \text{ and } \langle \operatorname{Re} x^*, \widehat{x}\rangle = |\langle x^*, x\rangle| . \tag{2.1}$$

In fact, this is obviously true if $\langle x^*, x\rangle = 0$. Suppose that $\langle x^*, x\rangle \neq 0$ and define the vector $\widehat{x} := \overline{\operatorname{sgn}(\langle x^*, x\rangle)}x$, with $\operatorname{sgn}(z) := z/|z|$ being the signal function. Thus $\|\widehat{x}\| = \|x\|$ and

$$\langle x^*, \widehat{x}\rangle = \frac{\overline{\langle x^*, x\rangle}\,\langle x^*, x\rangle}{|\langle x^*, x\rangle|} = |\langle x^*, x\rangle| \in \mathbb{R},$$

which proves the claim. Using this property, it is not so hard to prove that (see e.g. [6, Lemma 6.39])

$$\|\operatorname{Re} x^*\|_{\mathcal{L}(X,\mathbb{R})} = \|x^*\|_{\mathcal{L}(X,\mathbb{C})}$$

for all $x^* \in X^*$ and consequently T is an isometric $\mathbb{R}-$linear operator. In particular T is injective. Writing now $\langle x^*, x\rangle = a + ib$, with $a, b \in \mathbb{R}$, we see that $a = \operatorname{Re}\langle x^*, x\rangle$ and

$$\operatorname{Re}\langle x^*, ix\rangle = \operatorname{Re}(i\,\langle x^*, x\rangle) = \operatorname{Re}(i\,(a + ib)) = -b.$$

Hence

$$\langle x^*, x\rangle = \langle \operatorname{Re} x^*, x\rangle - i\,\langle \operatorname{Re} x^*, ix\rangle\,,\ x \in X$$

which implies that T is surjective. Indeed, if $\widehat{x}^* \in X^*_{\mathbb{R}}$, then the operator $x^*\colon X \to \mathbb{C}$ defined by

$$\langle x^*, x\rangle = \langle \widehat{x}^*, x\rangle - i\,\langle \widehat{x}^*, ix\rangle\,,\ x \in X$$

is $\mathbb{C}-$linear and continuous, which means that $x^* \in X^*$. Further, it is clear that

$$\langle \operatorname{Re} x^*, x\rangle = \operatorname{Re}\langle x^*, x\rangle = \langle \widehat{x}^*, x\rangle \text{ for all } x \in X$$

and consequently $\operatorname{Re} x^* = \widehat{x}^*$. This leads to the conclusion that T is a $\mathbb{R}-$linear isometric isomorphismus.

Remark 2 *The above result implies in particular that*

- *the spaces $(X^*)_{\mathbb{R}}$ and $X^*_{\mathbb{R}}$ are isometric isomorphic;*
- *if X and Y are complex Banach spaces and $A : X \to Y$ is a bounded linear operator, then the operator $A_{\mathbb{R}} : X_{\mathbb{R}} \to Y_{\mathbb{R}}$ defined by $A_{\mathbb{R}}x = Ax$ is linear and bounded with the same operator norm and its Banach adjoint operator satisfies $A^*_{\mathbb{R}} = TA^*T^{-1}$;*
- *the complex pre-Hilbert space H is a Hilbert space if and only if $H_{\mathbb{R}}$ is a Hilbert space with the inner product $\langle x, y\rangle_{H_{\mathbb{R}}} := \operatorname{Re}\langle x, y\rangle_H$;*

All these considerations show that it suffices to consider only real Banach spaces (resp. real Hilbert spaces). For this reason, we assume without restriction of the generality, that X is always a real Banach space for the rest of this work. Accordingly, the Hilbert space H is always considered a real Hilbert space.

2.1 The subdifferential

Definition 3 *Let X be a vector space and $\overline{\mathbb{R}} := (-\infty, \infty]$. The function $f\colon X \to \overline{\mathbb{R}}$ is said to be* CONVEX *if*

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

for all $x, y \in X$ and $\lambda \in (0,1)$. If this inequality is strict whenever $x \neq y$, then f is said to be STRICTLY CONVEX. *The set $D(f) := \{x \in X : f(x) < \infty\}$ is called the* EFFECTIVE DOMAIN *of f. Finally, the function f is said to be* PROPER *if $D(f) \neq \varnothing$.*

A convex function f always has a convex effective domain, i.e., $\lambda x + (1-\lambda) y \in$D$(f)$ whenever $x, y \in$D(f) and $\lambda \in (0,1)$. Further, a proper convex function f is continuous in the interior of its effective domain int(D (f)) if X is finite dimensional. If X is infinite dimensional, the same result holds whenever f is bounded from above on a neighborhood of an interior point $x_0 \in$int(D (f)) [10, Theo. 1.10 and Cor. 1.11, Ch. I].

Definition 4 *Let X and Y be normed vector spaces, $C \subset X$ open, $F\colon C \to Y$ a function and $x_0 \in C$. The* DIRECTIONAL DERIVATIVE *of F at x_0 in the direction $x \in X$ is defined by the limit*

$$F_+ (x_0, x) := \lim_{t \to 0^+} \frac{F(x_0 + tx) - F(x_0)}{t},$$

if it exists.

If there exists a linear and bounded operator $A\colon X \to Y$ satisfying

$$Ax = \lim_{t \to 0} \frac{F(x_0 + tx) - F(x_0)}{t}$$

for each $x \in X$, then the function F is said to be GATEAUX-DIFFERENTIABLE *(in short G-differentiable) at x_0 and we denote by $F'(x_0) := A$ the G-derivative of F at x_0.*

We say that F is FRECHET-DIFFERENTIABLE *(in short F-differentiable) at x_0, if it is G-differentiable at this point and*

$$\lim_{t \to 0} \sup_{\|x\|=1} \left\| \frac{F(x_0 + tx) - F(x_0)}{t} - F'(x_0) \right\| = 0.$$

In this case, the G-derivative of F at x_0, $F'(x_0)$, is also called the F-derivative. Finally, F is said to be CONTINUOUSLY F-DIFFERENTIABLE *(resp. continuously* G-DIFFERENTIABLE*) in C if the F-derivative (resp. G-derivative) $F'\colon C \to \mathcal{L}(X, Y)$ is a continuous function.*

It is clear from definition that F F-differentiable at x_0 implies F G-differentiable at x_0. The second condition in turn, implies that the directional derivative of F at x_0 exists in all directions $x \in X$ and

$$F_+ (x_0, x) = F'(x_0) x = F_- (x_0, x),$$

where

$$F_- (x_0, x) := \lim_{t \to 0^-} \frac{F(x_0 + tx) - F(x_0)}{t} = -F_+ (x_0, -x).$$

Under appropriate assumptions, the important *chain rule* holds true for F-differentiable functions $F\colon X \to Y$ and $G\colon Y \to Z$ (see [10, Theo. 1.14, Ch. I]): $G(F(x))' = G'(F(x)) F'(x)$ for all $x \in X$.

If $f\colon X \to \overline{\mathbb{R}}$ is a proper convex function, then for any $x_0 \in$int(D (f)), the directional derivatives $f'_+(x_0, x)$ and $f'_-(x_0, x)$ exist in all directions $x \in X$ and $f'_-(x_0, x) \leq f'_+(x_0, x)$. It is clear that this inequality becomes an equality if f is G-differentiable at x_0. Reciprocally, $f'_-(x_0, x) = f'_+(x_0, x)$ for all $x \in X$ implies the G-differentiability of f and in this case, $f'_-(x_0, x) = f'_+(x_0, x) = f'(x_0) x$.

Definition 5 *Let X be a Banach space, $f\colon X \to \overline{\mathbb{R}}$ be a function and let 2^X denote the set of all subsets of X. We say that f is* SUBDIFFERENTIABLE *at a point $x_0 \in X$ if there exists a functional $x_0^* \in X^*$, called* SUBGRADIENT *of f at x_0, such that*

$$f(x) - f(x_0) \geq \langle x_0^*, x - x_0 \rangle \text{ for all } x \in X.$$

The set of all subgradients of f at x_0 is denoted by $\partial f(x_0)$ and the mapping $\partial f\colon X \to 2^{X^*}$ is called the SUBDIFFERENTIAL *of f.*

The equivalence

$$\alpha f(x) - \alpha f(x_0) \geq \langle x_0^*, x - x_0 \rangle \Longleftrightarrow f(x) - f(x_0) \geq \left\langle \frac{1}{\alpha} x_0^*, x - x_0 \right\rangle$$

for any $\alpha > 0$, leads to the conclusion: $x_0^* \in \partial(\alpha f)(x_0) \Longleftrightarrow \frac{1}{\alpha} x_0^* \in \partial f(x_0)$, this is,

$$\partial(\alpha f)(x_0) = \alpha \partial f(x_0) \text{ for each } \alpha > 0.$$

A subdifferentiable function f is convex and lower semi-continuous in any open convex set $C \subset \mathrm{D}(f)$. Reciprocally, a proper convex and lower semi-continuous function f is always subdifferentiable on $\mathrm{int}(\mathrm{D}(f))$.

The *optimality condition* $0 \in \partial f(x_0)$ is equivalent to $f(x_0) \leq f(x)$ for all $x \in X$, which means that x_0 is a minimizer of f.

Let f be proper and convex and let $x_0 \in \mathrm{int}(\mathrm{D}(f))$. Then $x_0^* \in \partial f(x_0)$ if and only if

$$f'_-(x_0, x) \leq \langle x_0^*, x \rangle \leq f'_+(x_0, x), \text{ for all } x \in X,$$

see [10, Prop. 2.5, Ch. I]. From it, we conclude that f has a unique subgradient at $x_0 \in \mathrm{int}(\mathrm{D}(f))$ if and only if f is G-differentiable at x_0. In this case, $f'(x_0)x = \langle x_0^*, x \rangle$ for all $x \in X$, this is, $\partial f(x_0) = \{f'(x_0)\}$.

If f_1 and f_2 are two convex functions defined on X such that there is a point $x_0 \in \mathrm{D}(f_1) \cap \mathrm{D}(f_2)$ where f_1 is continuous, then [10, Theo. 2.8, Ch. I]

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \text{ for all } x \in X.$$

If $A\colon X \to Y$ is a bounded linear operator between the Banach spaces X and Y and $f\colon Y \to \overline{\mathbb{R}}$ is convex and continuous at some point of the range of A, then

$$\partial(f \circ A)(x) = A^*(\partial f(Ax)) \text{ for all } x \in X,$$

where $A^*\colon Y^* \to X^*$ represents the Banach adjoint of A (cf. [49]). Finally, for $f\colon Y \to \overline{\mathbb{R}}$ convex and $b \in Y$ fixed, see [48, Theo. 2.24],

$$\partial(f(\cdot - b))(y) = (\partial f)(y - b) \text{ for all } y \in Y.$$

Example 6 *Let X and Y be Banach spaces. Define the Tikhonov functional*

$$T_\alpha(x) := \frac{1}{r}\|Ax - b\|^r + \alpha \frac{1}{p}\|x\|^p,$$

with $p, r > 1$, $\alpha > 0$, $A\colon X \to Y$ linear and bounded and $b \in Y$. Using the optimality condition, we find that $x_0 \in X$ minimizes T_α if and only if

$$0 \in \partial T_\alpha(x_0) = A^*\left(\partial\left(\frac{1}{r}\|\cdot\|^r\right)\right)(Ax_0 - b) + \alpha \partial\left(\frac{1}{p}\|\cdot\|^p\right)(x_0). \tag{2.2}$$

We concentrate in the particular case $p = r = 2$. In a Hilbert space, making use of the definition of G-derivative and the polarization identity (2.5)*, it is not too difficult to prove that the convex function $f(x) = \frac{1}{2}\|x\|^2$ satisfies $f'(x_0)x = \langle x_0, x \rangle$ for all $x \in X$. Consequently $\partial f(x_0)$ is single valued and $\partial f(x_0) = f'(x_0) = x_0$, this is, $\partial\left(\frac{1}{2}\|\cdot\|^2\right)$ is just the identity operator. Equality* (2.2) *assumes in this case the form (for $p = r = 2$)*

$$0 = A^*(Ax_0 - b) + \alpha x_0 \Longrightarrow x_0 = (A^*A + \alpha I)^{-1} A^* b.$$

Using the definition of subdifferential, one can prove that for all $x \in X \setminus \{0\}$ [10, Prop. 3.4, Ch. I],

$$\partial\|x\| = \{x^* \in X^* : \langle x^*, x \rangle = \|x\| \text{ and } \|x^*\| = 1\}. \tag{2.3}$$

This result motivates the definition of a smooth Banach space.

2.2 Smoothness of Banach spaces

Definition 7 *A Banach space X is called* SMOOTH *if for every $x \neq 0$, there is a unique $x^* \in X^*$ such that $\langle x^*, x\rangle = \|x\|$ and $\|x^*\| = 1$.*

The existence of x^* in the above definition is guaranteed by the Theorem of Hahn-Banach. Thus, smoothness of Banach spaces is only a matter of uniqueness of x^*.

Remark 8 *We want to mention the fact that from the isomorphismus T constructed at the beginning of this chapter it follows that a complex Banach space is smooth if and only if its corresponding real Banach space is smooth. This is a direct consequence of the equivalence*

$$\langle x^*, x\rangle = \|x\| \text{ and } \|x^*\| = 1 \Longleftrightarrow \langle Tx^*, x\rangle = \|x\| \text{ and } \|Tx^*\| = 1, \tag{2.4}$$

which we prove now. Indeed, first remember that $\|Tx^*\| = \|x^*\|$. Now, if $\langle x^*, x\rangle = \|x\| \in \mathbb{R}$, then

$$\|x\| = \langle x^*, x\rangle = \operatorname{Re}\langle x^*, x\rangle = \langle Tx^*, x\rangle .$$

Reciprocally, assuming $\langle Tx^*, x\rangle = \|x\|$ we define the vector $\widehat{x} := \overline{\operatorname{sgn}(\langle x^*, x\rangle)}x$ and obtain from (2.1),

$$\begin{aligned}\|x\| = \langle Tx^*, x\rangle &= \operatorname{Re}\langle x^*, x\rangle \leq |\langle x^*, x\rangle| = \operatorname{Re}\langle x^*, \widehat{x}\rangle \\ &= \langle Tx^*, \widehat{x}\rangle \leq \|Tx^*\| \, \|\widehat{x}\| = \|\widehat{x}\| = \|x\| .\end{aligned}$$

Consequently, $\operatorname{Re}\langle x^*, x\rangle = |\langle x^*, x\rangle|$ which implies that $\langle x^*, x\rangle = \operatorname{Re}\langle x^*, x\rangle = \|x\|$ and the proof is complete.

As the subdifferential of $\|\cdot\| : X \to \mathbb{R}$ is always a subset of $X^*_{\mathbb{R}}$, we see that in a complex Banach space X, the set in (2.3) is actually described by

$$\partial \|x\| = \{x^* \in X^*_{\mathbb{R}} : \langle x^*, x\rangle = \|x\| \text{ and } \|x^*\| = 1\} .$$

But as these two sets can be identified to each other using the isomorphismus T (see (2.4) above), (2.3) can be used, in a slight abuse of notation, even in complex Banach spaces.

Using the last definition and (2.3), we conclude that a Banach space is smooth if and only if the subdifferential of the convex function $f(x) = \|x\|$ is single valued for all $x \in X\setminus\{0\}$. This is in turn, an equivalent condition to the G-differentiability of f in $X\setminus\{0\}$, i.e., the Banach space X is smooth if and only if the norm-function is G-differentiable in $X\setminus\{0\}$.

Definition 9 *Let X be a Banach space. The function $\rho_X : \mathbb{R}^+ \to \mathbb{R}$ defined by*

$$\rho_X(\tau) := \frac{1}{2}\sup\{\|x + \tau y\| + \|x - \tau y\| - 2 : \|x\| = \|y\| = 1\},$$

is called the MODULUS OF SMOOTHNESS *of X. The space X is said to be* UNIFORMLY SMOOTH *if $\lim_{\tau\to 0^+} \rho_X(\tau)/\tau = 0$ and it is called $p-$*SMOOTH*, for $p > 1$ fixed, if $\rho_X(\tau) \leq C_p\tau^p$, where $C_p > 0$ is a constant independent on τ.*

X $p-$smooth implies that $\rho_X(\tau)/\tau \leq C_p\tau^{p-1} \to 0$ as $\tau \to 0$, which in turn implies that X is uniformly smooth.

Define the function $f(x) = \|x\|$ and observe that

$$\frac{\rho_X(\tau)}{\tau} = \frac{1}{2}\sup\left\{\frac{f(x+\tau y) - f(x)}{\tau} + \frac{f(x+\tau(-y)) - f(x)}{\tau} : \|x\| = \|y\| = 1\right\}.$$

Assume now that X is uniformly smooth, then for all $x, y \in X$ with $\|x\| = \|y\| = 1$,

$$0 = \lim_{\tau \to 0^+} \left(\frac{f(x+\tau y) - f(x)}{\tau} + \frac{f(x+\tau(-y)) - f(x)}{\tau} \right) = f'_+(x,y) - f'_-(x,y).$$

It is not so difficult to extend this equality for all $y \in X$, which implies that f is G-differentiable at x, for all $x \in X$ satisfying $\|x\| = 1$. Finally, one can prove that the result actually holds for all $x \in X\backslash\{0\}$, which means that X is smooth. Hence, the uniform smoothness of X implies the smoothness of this space. The converse is however not true, as can be seen in [10, Theo. 3.12, Ch. I], where a proof of the equivalence between uniform smoothness and uniform F-differentiability of the norm-function on the unit sphere is given. In particular, the norm-function is F-differentiable in $X\backslash\{0\}$ provided X is uniformly smooth.

Example 10 *With help of the* POLARIZATION IDENTITY*:*

$$\frac{1}{2}\|x-y\|^2 = \frac{1}{2}\|x\|^2 - \langle x,y\rangle + \frac{1}{2}\|y\|^2, \tag{2.5}$$

which holds true for all vectors x and y in an arbitrary Hilbert space H, one can prove that

$$(\|x+\tau y\| + \|x-\tau y\|)^2 \leq 2\left(\|x+\tau y\|^2 + \|x-\tau y\|^2\right) = 4\left(\|x\|^2 + \tau^2\|y\|^2\right), \tag{2.6}$$

which implies that

$$\frac{1}{2}(\|x+\tau y\| + \|x-\tau y\|) \leq \sqrt{1+\tau^2}$$

for $\|x\| = \|y\| = 1$. Consequently, $\rho_H(\tau) \leq \sqrt{1+\tau^2} - 1 < \frac{1}{2}\tau^2$, which proves that a Hilbert space is always $2-$smooth. Further, as $\tau > 0$, the inequality in (2.6) holds if and only if $\|x+\tau y\| = \|x-\tau y\|$, i.e., $\langle x,y\rangle = 0$. This is possible if and only if $\dim H \geq 2$, and in this case we obtain $\rho_H(\tau) = \sqrt{1+\tau^2} - 1$.

2.3 Convexity of Banach spaces

Definition 11 *The Banach space X is said to be* STRICTLY CONVEX *if for all $x, y \in X$ with $x \neq y$ and $\|x\| = \|y\| = 1$ it holds $\|\lambda x + (1-\lambda) y\| < 1$ for all $\lambda \in (0,1)$.*

The above definition states that in a strictly convex Banach space, the line segment connecting two points in the unit sphere has only points lying inside this sphere, except for the extremal points themselves. It is possible to prove (see [10, Prop. 1.2, Ch. II]) that a Banach space is strictly convex if and only if the unit sphere has no line segments. It is also an equivalent condition to $\left\|\frac{1}{2}(x+y)\right\| < 1$ for all $x, y \in X$ with $x \neq y$ and $\|x\| = \|y\| = 1$ (the midpoint of a line segment with extremal points lying in the unit sphere does not belong to this sphere).

The strict convexity of the Banach space X is also equivalent to the strict convexity of the function $h(x) = \|x\|^p$, for any $p > 1$ fixed. In fact, let $p > 1$ and $x, y \in X$ be fixed. As the function $f\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$, $t \longmapsto t^p$ is strictly convex, we get from triangle inequality,

$$\begin{aligned} h(\lambda x + (1-\lambda) y) &= \|\lambda x + (1-\lambda) y\|^p \leq (\lambda\|x\| + (1-\lambda)\|y\|)^p \\ &= f(\lambda\|x\| + (1-\lambda)\|y\|) \leq \lambda f(\|x\|) + (1-\lambda) f(\|y\|) \\ &= \lambda h(x) + (1-\lambda) h(y), \end{aligned} \tag{2.7}$$

for all $\lambda \in (0,1)$. Hence h is convex. Suppose now that h is not strictly convex. Then there exist $x, y \in X$ with $x \neq y$ and $\lambda_0 \in (0,1)$ such that $h(\lambda_0 x + (1-\lambda_0) y) = \lambda_0 h(x) + (1-\lambda_0) h(y)$. From (2.7),

$$f(\lambda_0\|x\| + (1-\lambda_0)\|y\|) = \lambda_0 f(\|x\|) + (1-\lambda_0) f(\|y\|),$$

which implies that $\|x\| = \|y\| \neq 0$ because f is strictly convex. Again from (2.7),

$$\|\lambda_0 x + (1-\lambda_0)\, y\| = \lambda_0 \|x\| + (1-\lambda_0)\|y\| = \|x\| = \|y\|,$$

this is, $\left\|\lambda_0 \frac{x}{\|x\|} + (1-\lambda_0)\frac{y}{\|y\|}\right\| = 1$, which implies that X is not strictly convex. Conversely, suppose that X is not strictly convex. Then there exist $\lambda \in (0,1)$ and $x, y \in X$ with $x \neq y$ and $\|x\| = \|y\| = 1$ such that

$$\begin{aligned} h(\lambda x + (1-\lambda)\, y) &= \|\lambda x + (1-\lambda)\, y\|^p = 1 \\ &= \lambda \|x\|^p + (1-\lambda)\|y\|^p = \lambda h(x) + (1-\lambda)\, h(y), \end{aligned}$$

which implies that h is not strictly convex.

Definition 12 *Let X be a Banach space. The function $\delta_X\colon\ (0,2] \to \mathbb{R}$ defined by*

$$\delta_X(\epsilon) := \inf\left\{1 - \left\|\frac{1}{2}(x+y)\right\| : \|x\| = \|y\| = 1 \text{ and } \|x-y\| \geq \epsilon\right\},$$

is called the MODULUS OF CONVEXITY *of X. The space X is said to be* UNIFORMLY CONVEX *if $\delta_X(\epsilon) > 0$ for all $0 < \epsilon \leq 2$ and it is called $p-$*CONVEX, *for $p > 1$ fixed, if $\delta_X(\epsilon) \geq K_p \epsilon^p$, where $K_p > 0$ is a constant independent on τ.*

It is easy to prove that $p-$convexity implies uniform convexity.

Fix $x, y \in X$ with $x \neq y$ and $\|x\| = \|y\| = 1$. If X is uniformly convex, then for $\epsilon := \|x-y\| \in (0,2]$ it holds $\delta(\epsilon) > 0$, which implies in view of definition of δ_X that $1 - \left\|\frac{1}{2}(x+y)\right\| > 0$, which in turn implies that X is strictly convex.

In [9, Example 1.7], the author presents an interesting example of two different Banach spaces, which have equivalent norms, and at the same time the first one is strictly convex, the second one is not. Further, the strictly convex space is not uniformly convex, which proves that these are not equivalent concepts.

An interesting property of uniformly convex Banach spaces is the following: if $x_n \rightharpoonup x$ and $\|x_n\| \to \|x\|$ as $n \to \infty$, then $x_n \to x$ as $n \to \infty$ [10, Prop. 2.8, Ch. II].

For an arbitrary Hilbert space H and a Banach space X, the inequalities $\delta_H(\epsilon) \geq \delta_X(\epsilon)$ for all $0 < \epsilon \leq 2$ and $\rho_H(\tau) \leq \rho_X(\tau)$ for all $\tau > 0$ always hold. Moreover X $p-$convex implies $p \geq 2$ and X $p-$smooth implies $p \leq 2$. Since Hilbert spaces are $2-$smooth and $2-$convex (see Example 10 and Example 13 below), we conclude that they are the "most convex" and "smoothest" Banach spaces.

As shown in Example 10, an explicit formula for $\rho_H(\tau)$ is known in case of $\dim H \geq 2$. The same is true for $\delta_H(\epsilon)$ as the next example shows.

Example 13 *In a Hilbert space H, the polarization identity* (2.5) *shows that*

$$\|x+y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) - \|x-y\|^2.$$

With $\|x\| = \|y\| = 1$, it results in

$$\left\|\frac{1}{2}(x+y)\right\|^2 = 1 - \left(\frac{\|x-y\|}{2}\right)^2$$

and with $\|x-y\| \geq \epsilon$ we find $\delta_H(\epsilon) \geq 1 - \sqrt{1 - \left(\frac{\epsilon}{2}\right)^2} > \frac{1}{8}\epsilon^2$, which proves that Hilbert spaces are $2-$convex. If $\dim H \geq 2$, then choosing two sequences in the unit sphere satisfying $\|x_n - y_n\| \to \epsilon$ as $n \to \infty$, we obtain the equality $\delta_H(\epsilon) = 1 - \sqrt{1 - \left(\frac{\epsilon}{2}\right)^2}$.

2.4 Relations between smoothness and convexity

In this section we always consider X a Banach space and X^* its dual space. The numbers $p, p^* > 1$ represent *conjugate numbers*, i.e., $1/p + 1/p^* = 1$.

The strict convexity of X is equivalent to the smoothness of its dual space and vice-versa in case of X being reflexive, as shown in [10, Cor. 1.4, Ch. II]. X uniformly smooth or uniformly convex implies X reflexive [10, Theo. 2.9 and 2.15, Ch. II].

A very important result is the so called Lindenstrauss' duality formulas, which connects the modulus of smoothness of X and the modulus of convexity of its dual space X^* and vice-versa (see [10, Prop. 2.12, Ch. II]): For each $\tau > 0$ it holds

$$\rho_X(\tau) = \sup\left\{\frac{\tau\epsilon}{2} - \delta_{X^*}(\epsilon) : 0 < \epsilon \leq 2\right\}$$

and

$$\rho_{X^*}(\tau) = \sup\left\{\frac{\tau\epsilon}{2} - \delta_X(\epsilon) : 0 < \epsilon \leq 2\right\}.$$

A consequence of Lindenstrauss' formulas is the following important result: X is uniformly convex (resp. $p-$convex) if and only if X^* is uniformly smooth (resp. p^*-smooth) and vice-versa (cf [38, 143, Vol II 1.e]).

Example 14 *The Lebesgue space $L^p(\Omega)$ satisfies:*

$$\delta_{L_p}(\epsilon) = \begin{cases} \frac{p-1}{8}\epsilon^2 + o\left(\epsilon^2\right) & > \frac{p-1}{8}\epsilon^2 \quad : 1 < p < 2 \\ 1 - \left(1 - \left(\frac{\epsilon}{2}\right)^p\right)^{\frac{1}{p}} & > \frac{1}{p}\left(\frac{\epsilon}{2}\right)^p \quad : 2 \leq p < \infty \end{cases}$$

and

$$\rho_{L_p}(\tau) = \begin{cases} (1+\tau^p)^{\frac{1}{p}} - 1 & < \frac{1}{p}\tau^p \quad : 1 < p < 2 \\ \frac{p-1}{2}\tau^2 + o\left(\tau^2\right) & < \frac{p-1}{2}\tau^2 \quad : 2 \leq p < \infty \end{cases},$$

which means that this space is[1] *$p \vee 2-$convex and $p \wedge 2-$smooth. In particular, it is uniformly smooth and uniformly convex. The space of the summable sequences $\ell^p(\mathbb{N})$ and the Sobolev spaces $W^{n,p}(\Omega)$, $n \in \mathbb{N}$, are also $p \vee 2-$convex and $p \wedge 2-$smooth for $1 < p < \infty$. As these spaces are not reflexive for $p = 1$ and $p = \infty$, we conclude that they cannot be uniformly smooth nor uniformly convex. One can actually prove that they are not even strictly convex or smooth Banach spaces.*

2.5 Duality mapping

The Riesz Representation Theorem states that in an arbitrary Hilbert space H, for each element $x \in H$, there exists a unique linear and continuous functional $x^* : H \to \mathbb{R}$ such that

$$\langle x^*, y\rangle_{H^* \times H} = \langle x, y\rangle_H\,, \text{ for all } y \in H$$

and $\|x^*\|_{\mathcal{L}(H,\mathbb{R})} = \|x\|_H$. This implies that

$$\langle x^*, y\rangle \leq \|x\| \, . \, \|y\| \text{ and } \langle x^*, x\rangle = \|x\|^2 . \tag{2.8}$$

This reasoning suggests we could cover the lack of an inner product in a general Banach space X associating each element $x \in X$ to a functional $x^* \in X^*$ and then replacing the inner product $\langle x, y\rangle$ with $\langle x^*, y\rangle_{X^* \times X}$. In the ideal case, the dual pairing $\langle \cdot, \cdot\rangle_{X^* \times X}$ would have inner-product-like properties similar to (2.8) above.

[1] $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$.

Definition 15 *Let X be a Banach space. A continuous and strictly increasing function $\varphi\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that $\varphi(0) = 0$ and $\lim_{t\to\infty} \varphi(t) = \infty$ is called a* GAUGE FUNCTION. *The set-valued mapping $J_\varphi\colon X \to 2^{X^*}$ defined by*

$$J_\varphi(x) := \{x^* \in X^* : \langle x^*, x\rangle = \|x^*\| . \|x\| \ \ \textit{and} \ \ \|x^*\| = \varphi(\|x\|)\}$$

is called DUALITY MAPPING *associated with the gauge function φ. The duality mapping associated with the gauge function $\varphi(t) = t$ is called* NORMALIZED DUALITY MAPPING. *Finally, a* SELECTION *of the duality mapping J_φ is a single-valued function $j_\varphi\colon X \to X^*$ satisfying $j_\varphi(x) \in J_\varphi(x)$ for each $x \in X$.*

Suppose we are given $x \in X$. From Hahn-Banach Theorem it follows that there exists $x^* \in X^*$ such that $\|x^*\| = 1$ and $\langle x^*, x\rangle = \|x\|$, which means that $y^* := x^*\varphi(\|x\|) \in J_\varphi(x)$. Hence $J_\varphi(x) \neq \varnothing$ for any $x \in X$.

Remark 16 *The same reasoning of the previous paragraph implies, in view of* (2.4), *that the vector x^* belongs to the set*

$$\{x^* \in X^* : \langle x^*, x\rangle = \|x^*\| . \|x\| \ \ \textit{and} \ \ \|x^*\| = \varphi(\|x\|)\}$$

if and only if the vector $\operatorname{Re} x^*$ *belongs to*

$$\{x^* \in X^*_{\mathbb{R}} : \langle x^*, x\rangle = \|x^*\| . \|x\| \ \ \textit{and} \ \ \|x^*\| = \varphi(\|x\|)\},$$

which means that the two above sets can be identified to each other by the use of the isomorphismus $Tx^* = \operatorname{Re} x^*$. Therefore, one is allowed to write

$$\langle j_\varphi(x), y\rangle \leq \|j_\varphi(x)\| . \|y\|$$

even if X is a complex Banach space. The above inequality should actually be interpreted as

$$\operatorname{Re}\langle j_\varphi(x), y\rangle \leq \|\operatorname{Re} j_\varphi(x)\| . \|y\| = \|j_\varphi(x)\| . \|y\| .$$

The Asplund's Theorem (2.9) *below and the Xu-Roach inequalities in Theorem 18 should be interpreted in the same way.*

With the special notation J_p, where $p > 1$ is fixed, we denote the duality mapping associated with the gauge function $\varphi(t) = t^{p-1}$. In particular, J_2 is the normalized duality mapping. From definition, we conclude that, for all $x, y \in X$,

$$\begin{aligned} \|j_p(x)\| &= \varphi(\|x\|) = \|x\|^{p-1}, \\ \langle j_p(x), x\rangle &= \|j_p(x)\| . \|x\| = \|x\|^p \ \text{ and} \\ \langle j_p(x), y\rangle &\leq \|j_p(x)\| . \|y\| = \|x\|^{p-1} \|y\| . \end{aligned}$$

Further, each selection j_2 of the normalized duality mapping has the inner-product properties shown in (2.8).

The connection between the subdifferential and the duality mapping is given by the very important Asplund's Theorem [10, Lemma 4.3 and Theo. 4.4, Ch. I]: Let X be a Banach space, $x \in X$ an arbitrary vector and φ a gauge function. Then the function $\psi(t) := \int_0^t \varphi(s)\mathrm{d}s$, $t \geq 0$ is convex in $\mathbb{R}_0^+$ and

$$J_\varphi(x) = \partial(\psi(\|x\|)) . \tag{2.9}$$

For the gauge function $\varphi(t) = t^{p-1}$ we have $\psi(\|x\|) = \frac{1}{p}\|x\|^p$ and conclude that $J_p = \partial\left(\frac{1}{p}\|\cdot\|^p\right)$. In particular, for the normalized duality mapping it holds $J_2 = \partial\left(\frac{1}{2}\|\cdot\|^2\right)$. In a Hilbert space, $\partial\left(\frac{1}{2}\left\|\cdot\right\|^2\right)(x) = \left(\frac{1}{2}\left\|\cdot\right\|^2\right)'(x) = x$, which means that $J_2 = I$ is the identity operator. Unfortunately, this very nice property is true only in Hilbert spaces. In fact, one can prove that the normalized duality mapping J_2 is linear in X if and only if X is a Hilbert space [10, Prop. 4.8, Ch. I].

The Asplund's Theorem is the key to connect the properties of the duality mappings with convexity and smoothness properties of a Banach space. For instance, an interesting consequence of Asplund's Theorem is the fact that a Banach space X is smooth if and only if each duality mapping J_φ is single valued (cf [10, Cor. 4.5, Ch. I]). In this case,

$$\langle J_\varphi(x), y\rangle = \frac{\mathrm{d}}{\mathrm{dt}}\psi\left(\|x+ty\|\right)\Big|_{t=0}, \text{ for all } x, y \in X. \tag{2.10}$$

Further, X is uniformly smooth if and only if each duality mapping is single-valued and uniformly continuous in the unit sphere [10, Theo. 2.16, Ch. II].

The next properties of the duality mapping J_φ are collected from [10, Prop. 4.7, Ch. I]: Let $x, y \in X$. The duality mapping inherits the monotonicity property of the subdifferential:

$$\langle j_\varphi(x) - j_\varphi(y), x - y\rangle \geq 0.$$

Further, $J_\varphi(-x) = -J_\varphi(x)$ and

$$J_\varphi(\lambda x) = \frac{\varphi(\lambda\|x\|)}{\varphi(\|x\|)} J_\varphi(x) \text{ for all } \lambda > 0.$$

In particular, $J_2(\lambda x) = \lambda J_2(x)$ is homogeneous. The inverse of φ is a gauge function too and if $J^*_{\varphi^{-1}} \colon X^* \to X^{**}$ is the duality mapping on X^* associated with the gauge function φ^{-1}, then

$$x^* \in J_\varphi(x) \Longrightarrow x \in J^*_{\varphi^{-1}}(x^*). \tag{2.11}$$

Finally, if φ_1 and φ_2 are gauge functions, then

$$\varphi_2(\|x\|) J_{\varphi_1}(x) = \varphi_1(\|x\|) J_{\varphi_2}(x).$$

In particular, for $\varphi_1(t) = t^{r-1}$ and $\varphi_2(t) = t^{p-1}$ with $p, r > 1$ it holds

$$J_r(x) = \|x\|^{r-p} J_p(x). \tag{2.12}$$

From [10, Cor. 3.13, Ch. II], we see that the range[2] of J_φ is dense in X^*, i.e., $\overline{R(J_\varphi)} = X^*$. If X is reflexive, then this result becomes $R(J_\varphi) = X^*$ (actually, this is an equivalent condition to the reflexivity of X, see [10, Cor 3.4, Ch. II]). We conclude that in case of X being reflexive, the reciprocal of (2.11) is true and $R\left(J^*_{\varphi^{-1}}\right) = X^{**} \cong X$. Assuming that X is smooth, then each duality mapping is single valued and if X is additionally reflexive (this is the case for instance, if X is uniformly smooth), then J_φ is invertible and satisfies

$$J_\varphi^{-1} = J^*_{\varphi^{-1}} \colon X^* \to X^{**} \cong X. \tag{2.13}$$

In particular, $J_p^{-1} = J^*_{p^*}$. Lastly, if the norm-functions in X and X^* are F-differentiable, then X is reflexive, J_φ is single-valued, continuous and invertible with continuous inverse satisfying (2.13). This result is valid for example, if X is uniformly smooth (then the norm in X is uniformly F-differentiable on the unit sphere) and uniformly convex (then X^* is uniformly smooth and the same result holds true in this space).

The strict monotonicity of the duality mapping is equivalent to the strict convexity of X, i.e., X is strictly convex if and only if, each duality mapping satisfies

$$\langle j_\varphi(x) - j_\varphi(y), x - y\rangle > 0 \text{ for all } x, y \in X \text{ with } x \neq y.$$

[2] We define the range of $J_\varphi \colon X \to 2^{X^*}$ as $R(J_\varphi) := \cup_{x\in X} J_\varphi(x)$, which makes sense even if J_φ is not single-valued.

Example 17 *The dual space of the Lebesgue space $L^p(\Omega)$, $1 < p < \infty$ is given by $(L^p(\Omega))^* = L^{p^*}(\Omega)$ and the duality mapping J_p: $L^p(\Omega) \to L^{p^*}(\Omega)$ can be calculated using the formula* (2.10)*. In fact, the duality mapping J_p is associated with the gauge function $\varphi(t) = t^{p-1}$, which means that $\psi(t) = \int_0^t \varphi(s)ds = \frac{1}{p}t^p$. Let $f, g \in L^p(\Omega)$ be given. Then by* (2.10)*,*

$$\begin{aligned}
\langle J_p(f), g\rangle_{L^{p^*}\times L^p} &= \frac{\mathrm{d}}{\mathrm{dt}}\psi(\|f+tg\|_{L^p})\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{dt}}\left(\frac{1}{p}\|f+tg\|_{L^p}^p\right)\Big|_{t=0} \\
&= \frac{1}{p}\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega |f(x)+tg(x)|^p\,\mathrm{dx}\Big|_{t=0} \\
&= \int_\Omega |f(x)+tg(x)|^{p-1}\operatorname{sgn}(f(x)+tg(x))\,g(x)\,\mathrm{dx}\Big|_{t=0} \\
&= \int_\Omega |f(x)|^{p-1}\operatorname{sgn}(f(x))\,g(x)\,\mathrm{dx} = \left\langle |f|^{p-1}\operatorname{sgn}(f), g\right\rangle_{L^{p^*}\times L^p}.
\end{aligned}$$

This means that $J_p(f) = |f|^{p-1}sgn(f)$, where the equality is understood pointwise. Using now (2.12) *we conclude that the duality mapping J_r in $L^p(\Omega)$ is given by*

$$J_r(f) = \|f\|_{L^p}^{r-p}|f|^{p-1}\operatorname{sgn}(f). \tag{2.14}$$

2.6 Bregman distances

In 1991, Xu and Roach proved in their famous paper [53] two very important results, which are nowadays known as Xu-Roach inequalities. We start this section formulating these results in form of a theorem.

Theorem 18 (Xu-Roach) *Let X be a Banach space and $p > 1$.*

(A) If X is uniformly convex, then there exists a positive constant $\widetilde{K}_p$ such that for all $x, y \in X$ and $j_p(x) \in J_p(x)$,

$$\|x-y\|^p \geq \|x\|^p - p\langle j_p(x), y\rangle + \sigma_p(x,y)$$

with

$$\sigma_p(x,y) := \widetilde{K}_p \int_0^1 \frac{(\|x-ty\| \vee \|x\|)^p}{t}\delta_X\left(\frac{t\|y\|}{2(\|x-ty\|\vee\|x\|)}\right)\mathrm{dt},$$

where δ_X is the modulus of convexity, see Definition 12.

(B) If X is uniformly smooth, then there exists a positive constant $\widetilde{C}_p$ such that for all $x, y \in X$

$$\|x-y\|^p \leq \|x\|^p - p\langle J_p(x), y\rangle + \widetilde{\sigma}_p(x,y)$$

with

$$\widetilde{\sigma}_p(x,y) := \widetilde{C}_p \int_0^1 \frac{(\|x-ty\| \vee \|x\|)^p}{t}\rho_X\left(\frac{t\|y\|}{\|x-ty\|\vee\|x\|}\right)\mathrm{dt},$$

where ρ_X is the modulus of smoothness (Definition 12).

Assuming that the space X is $s-$convex for some $s > 1$ and using the definition of $s-$convexity,

$$\sigma_p(x,y) \geq \frac{\widetilde{K}_p K_s}{2^s}\|y\|^s \int_0^1 (\|x-ty\|\vee\|x\|)^{p-s}t^{s-1}\mathrm{dt}.$$

Since for all $t \in [0,1]$,

$$\|x-ty\|\vee\|x\| \leq \|x\| + \|y\| \leq 2(\|x\|\vee\|y\|),$$

it follows that for $p \leq s$,

$$\sigma_p(x,y) \geq pK_{p,s}(\|x\| \vee \|y\|)^{p-s}\|y\|^s, \tag{2.15}$$

where $K_{p,s} = \widetilde{K}_p K_s 2^{p-2s}/ps > 0$. Similarly, if the Banach space X is assumed to be $s-$smooth and $p \geq s$, then there exists a positive constant $C_{p,s}$ such that for all $x, y \in X$

$$\widetilde{\sigma}_p(x,y) \leq pC_{p,s}(\|x\| \vee \|y\|)^{p-s}\|y\|^s. \tag{2.16}$$

In particular, if $p = s$, then

$$\frac{1}{p}\|x-y\|^p \geq \frac{1}{p}\|x\|^p - \langle j_p(x), y\rangle + \frac{\overline{K}_p}{p}\|y\|^p, \tag{2.17}$$

and

$$\frac{1}{p}\|x-y\|^p \leq \frac{1}{p}\|x\|^p - \langle J_p(x), y\rangle + \frac{\overline{C}_p}{p}\|y\|^p, \tag{2.18}$$

respectively, with $\overline{K}_p := pK_{p,s}$ and $\overline{C}_p := pC_{p,s}$.

In [9, Cor. 4.17 and Cor. 5.8] it is shown that the existence of $\overline{K}_p$ and $\overline{C}_p$ in inequalities (2.17) and (2.18) are actually equivalent conditions to $p-$convexity and $p-$smoothness of X respectively.

Note that inequalities (2.17) and (2.18) reduce to the polarization identity (2.5) in Hilbert spaces (for $p = s = 2$, $\overline{K}_p = \overline{C}_p = 1$). Trying to mimic this identity in a general Banach space, we introduce the *Bregman distance.*

Definition 19 *Let X be a Banach space and $\Omega\colon X \to \mathbb{R}$ a convex and subdifferentiable functional. The* BREGMAN DISTANCE *associated to Ω is the function $\Delta_\Omega\colon X \times X \to \mathbb{R}$ defined by*

$$\Delta_\Omega(x,y) := \Omega(x) - \Omega(y) - \inf\{\langle \xi, x-y\rangle : \xi \in \partial\Omega(y)\}.$$

Despite its name, the Bregman distance is not a metric because it does not satisfy the reflexivity for example ($\Delta_\Omega(x,y) \neq \Delta_\Omega(y,x)$ in general). It does not satisfy the important triangle inequality either. But, from definition of subdifferential immediately follows that $\Delta_\Omega(x,y) \geq 0$ for all $x, y \in X$. Additionally, $x = y$ implies $\Delta_\Omega(x,y) = 0$.

Let φ be a gauge function. Then, from Asplund's Theorem, the function $\Omega(x) := \psi(\|x\|) = \int_0^{\|x\|}\varphi(s)\mathrm{ds}$ is convex and $\partial\Omega(x) = J_\varphi(x)$. It follows that in this case,

$$\Delta_\Omega(x,y) = \psi(\|x\|) - \psi(\|y\|) - \inf\{\langle \xi, x-y\rangle : \xi \in J_\varphi(y)\}.$$

We denote by Δ_p, $p > 1$, the Bregman distance associated to the particular gauge function $\varphi(t) = t^{p-1}$. This means

$$\Delta_p(x,y) = \frac{1}{p}\|x\|^p - \frac{1}{p}\|y\|^p - \inf\{\langle \xi, x-y\rangle : \xi \in J_p(y)\}. \tag{2.19}$$

Assume from now on that the duality mapping is single-valued (this is the case in smooth Banach spaces for instance). Hence, the above equality becomes

$$\begin{aligned}
\Delta_p(x,y) &= \frac{1}{p}\|x\|^p - \frac{1}{p}\|y\|^p - \langle J_p(y), x-y\rangle \\
&= \frac{1}{p}\|x\|^p - \frac{1}{p}\|y\|^p + \|y\|^p - \langle J_p(y), x\rangle \\
&= \frac{1}{p}\|x\|^p + \frac{1}{p^*}\|y\|^p - \langle J_p(y), x\rangle \\
&= \frac{1}{p}\|x\|^p - \langle J_p(y), x\rangle + \frac{1}{p^*}\|J_p(y)\|^{p^*}.
\end{aligned}$$

Observe the similarity between this formula and the polarization identity (2.5). Since in Hilbert spaces the normalized duality mapping is the identity operator, we conclude that in these spaces $\Delta_2(x,y) = \frac{1}{2}\|x-y\|^2$. Further,

$$\Delta_p(x,y) \geq \frac{1}{p}\|x\|^p + \frac{1}{p^*}\|y\|^p - \|y\|^{p-1}\|x\|.$$

Now, if $(x_n)_{n\in\mathbb{N}} \subset X$ is a sequence and $x \in X$ is a fixed vector, then the inequality $\Delta_p(x,x_n) \leq \rho$ implies

$$\|x_n\|^{p-1}\left(\frac{1}{p^*}\|x_n\| - \|x\|\right) \leq \rho.$$

Considering now the cases $\frac{1}{p^*}\|x_n\| - \|x\| \leq \frac{1}{2p^*}\|x_n\|$ and $\frac{1}{p^*}\|x_n\| - \|x\| > \frac{1}{2p^*}\|x_n\|$, we conclude the implication

$$\Delta_p(x,x_n) \leq \rho \implies \|x_n\| \leq 2p^*\left(\|x\| \vee \rho^{1/p}\right). \tag{2.20}$$

Therefore, $(x_n)_{n\in\mathbb{N}}$ is bounded whenever $\Delta_p(x,x_n)$ is bounded. A similar result can be proven if $\Delta_p(x_n,x) \leq \rho$. If the duality mapping is single-valued and continuous (this is the case, for instance, in a uniformly smooth Banach space) then the continuity is handed down to both arguments of the Bregman distance Δ_p.

If X is strictly convex, then $x = y$ whenever $\Delta_p(x,y) = 0$ because X strictly convex implies that $x \longmapsto \frac{1}{p}\|x\|^p$ is strictly convex, which in turn implies that Δ_p is strictly convex on its first argument. Now, if $x \neq y$, we find for $\lambda \in (0,1)$

$$0 \leq \Delta_p(\lambda x + (1-\lambda)y, y) < \lambda\Delta_p(x,y) + (1-\lambda)\Delta_p(y,y) = \lambda\Delta_p(x,y),$$

which implies that $\Delta_p(x,y) \neq 0$.

A straightforward calculation leads to the THREE POINTS IDENTITY:

$$\Delta_p(z,y) - \Delta_p(z,x) = \Delta_p(x,y) + \langle J_p(y) - J_p(x), x - z\rangle \tag{2.21}$$

for all $x,y,z \in X$. It is also easy to verify the identity $\Delta_p(x,y) = \Delta_{p^*}(J_p(y), J_p(x))$.

The Xu-Roach Theorem states that in an arbitrary uniformly convex Banach space it holds

$$\frac{1}{p}\sigma_p(y, y-x) \leq \Delta_p(x,y),$$

for all $x,y \in X$. Using (2.15) we conclude that in an $s-$convex Banach space there exists, for each $p \leq s$, a constant $K_{p,s}$ such that

$$K_{p,s}(\|y\| \vee \|x-y\|)^{p-s}\|x-y\|^s \leq \Delta_p(x,y) \tag{2.22}$$

for all $x,y \in X$. In particular, if the sequence $(x_n)_{n\in\mathbb{N}} \subset X$ is bounded (or $p = s$), then there exists a constant $C > 0$ such that

$$\|x_n - x_m\|^s \leq C\Delta_p(x_n,x_m). \tag{2.23}$$

In the same way, in an uniformly smooth Banach space it is true that

$$\Delta_p(x,y) \leq \frac{1}{p}\widetilde{\sigma}_p(y, y-x)$$

for all $x,y \in X$, which in view of (2.16) implies that in a $s-$smooth Banach space it holds, for each $p \geq s$,

$$\Delta_p(x,y) \leq C_{p,s}(\|y\| \vee \|x-y\|)^{p-s}\|x-y\|^s = C_{p,s}\left(\|y\|^{p-s}\|x-y\|^s \vee \|x-y\|^p\right) \tag{2.24}$$

for all $x, y \in X$. One can additionally prove for an arbitrary $s-$smooth Banach space, that for each $p > 1$, there exists a positive constant $\overline{C}_{p,s}$ satisfying

$$\left\| J_p(x) - J_p(y) \right\| \leq 2^{s-p} \overline{C}_{p,s} \left(\|x\| \vee \|y\| \right)^{p-s} \|x - y\|^{s-1},$$

for all $x, y \in X$. Since

$$\|x\| \vee \|y\| \leq \|x - y\| + \|y\| \leq 2 \left(\|x - y\| \vee \|y\| \right),$$

we additionally have for $p \geq s$,

$$\begin{aligned} \left\| J_p(x) - J_p(y) \right\| &\leq \overline{C}_{p,s} \left(\|x - y\| \vee \|y\| \right)^{p-s} \|x - y\|^{s-1} \\ &= \overline{C}_{p,s} \left(\|x - y\|^{p-1} \vee \|y\|^{p-s} \|x - y\|^{s-1} \right). \end{aligned} \tag{2.25}$$

Chapter 3

The Inexact Newton Method K-REGINN

We aim to find an approximate solution to the nonlinear ill-posed problem

$$F(x) = y \tag{3.1}$$

with F operating between *Banach spaces* X and Y, that is, $F\colon \mathrm{D}(F) \subset X \to Y$, where $\mathrm{D}(F)$ denotes the domain of definition of F. We suppose to have full knowledge of this operator. An approximation y^δ for y satisfying

$$\left\| y - y^\delta \right\| \leq \delta,$$

and the *noise level* $\delta > 0$ are assumed to be known as well. Suppose now that a solution x^+ of (3.1) exists and for the ease of presentation, assume for now that it is unique. We aim to find for each pair $\left(y^\delta, \delta\right)$ satisfying the above inequality, a vector x_δ such that the *regularization property* holds[1]:

$$x_\delta \to x^+ \text{ as } \delta \to 0. \tag{3.2}$$

The basis of our work is the Newton-type algorithm `REGINN` (`REG`ularization based on `IN`exact `N`ewton iteration). We first explain the original idea of this method as introduced in [43] and later present our *Kaczmarz* version `K-REGINN`, which is a generalization of the original algorithm. `REGINN`, as described in [43], improves the current iterate x_n via

$$x_{n+1} = x_n + s_n \tag{3.3}$$

by a correction step s_n obtained from approximately solving a local linearization of (3.1):

$$A_n s = b_n^\delta \tag{3.4}$$

where $A_n := F'(x_n)$ is the Fréchet derivative of F at x_n and $b_n^\delta := y^\delta - F\left(x_n\right)$ is the corresponding nonlinear residual. For a fixed number n, `REGINN` typically applies an iterative solver to (3.4), called *inner iteration*, to generate a sequence $\left(s_{n,k}\right)_{k\in\mathbb{N}}$ which approximates a solution of this system. The inner iteration terminates in the first index $k = k_n$ satisfying

$$\left\| A_n s_{n,k} - b_n^\delta \right\| < \mu \left\| b_n^\delta \right\| \tag{3.5}$$

with a pre-defined constant $0 < \mu < 1$. The Newton iteration (3.3), also called the *outer iteration,* is now realized defining $s_n := s_{n,k_n}$. The algorithm is finally terminated with the

[1]The approximate solution x_δ actually depends on δ and y^δ : $x_\delta = x_{(\delta,y^\delta)}$.

discrepancy principle: Stop in the first iteration $n = N(\delta)$ satisfying[2]

$$\left\| y^\delta - F\left(x_{N(\delta)}\right)\right\| \le \tau\delta,$$

with $\tau > 1$ being a constant[3].

In order to define `REGINN` in a more general framework, we assume that (3.1) splits into $d \in \mathbb{N}$ "smaller" sub-problems, that is, Y factorizes into Banach spaces $Y_0, \ldots, Y_{d-1}$: $Y = Y_0 \times Y_1 \times \cdots \times Y_{d-1}$. Accordingly, $F = (F_0, F_1, \ldots, F_{d-1})^\top$, $F_j \colon \mathrm{D}(F) \subset X \to Y_j$, and $y = (y_0, y_1, \ldots, y_{d-1})^\top$. The equation (3.1) is now equivalent to the system

$$F_j(x) = y_j, \quad j = 0, \ldots, d-1. \tag{3.6}$$

Our task can be recast as: for each d pairs $\left(y_j^{\delta_j}, \delta_j\right)$ satisfying

$$\|y_j - y_j^{\delta_j}\| \le \delta_j,\ j = 0, ..., d-1, \tag{3.7}$$

find a vector x_δ such that the regularization property (3.2) holds for

$$\delta := \max\{\delta_j : j = 0, \ldots, d-1\} > 0. \tag{3.8}$$

The approximations $y_j^{\delta_j}$ to y_j and the respective positive noise levels δ_j as well as the operators F_j, $j = 0, \ldots, d-1$ are assumed to be known.

We emphasize that systems like (3.6) arise quite naturally in applications where the data is measured by d individual experiments or observations. For instance, in electrical impedance tomography, one wants to find the conductivity of an object by applying, say, d current patterns at the boundary and measuring the resulting voltages at the boundary as well (see the numerical experiments in Chapter 5).

Our goal now is to introduce a Kaczmarz variant of `REGINN` (in short `K-REGINN`). In contrast to traditional iterative methods, the Kaczmarz strategy, also known as *loping* strategy, aims to find a solution of the original problem (3.1) processing the equations (3.6) cyclically, using one equation each time stead of using all of them at the same time. This kind of cycling strategy was initiated by Kaczmarz [27] in the context of linear systems in finite dimensional spaces, first analyzed in the context of inverse problems by Kowar and Scherzer [31] and further investigated by several authors [20, 7, 40, 39, 37].

The idea of `K-REGINN` is to determine s_n from (3.4) similarly as explained in (3.5) but with

$$A_n := F'_{[n]}(x_n) \text{ and } b_n^\delta := y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n),$$

where $[n] := n \bmod d$ denotes the remainder of integer division. Thus, the subsystems are processed cyclically, breaking the large-scale system (3.1) into handy pieces. Observe that the inner iteration works with a fixed equation of (3.6). When the inner iteration terminates, the current vector x_n is updated in the outer iteration (3.3) generating the vector x_{n+1} and the current equation is replaced by the next one. Finally, the outer iteration is stopped using a variant of the discrepancy principle.

We go now into details and explain `K-REGINN` more precisely: Start the outer iteration with $x_0 \in \mathrm{D}(F)$ and the inner iteration setting $s_{n,0} := 0$. With n fixed, generate the sequence $(s_{n,k})_{k\in\mathbb{N}}$. Update now the outer iteration using $x_{n+1} := x_n + s_{n,k_n}$, where the final (inner) index k_n is determined as follows: choose $\tau > 1$ and $\mu \in (0,1)$. Define $k_n = 0$ in case of

$$\left\| b_n^\delta \right\| \le \tau\delta_{[n]}. \tag{3.9}$$

[2] The number N is chosen by a posteriori strategy, it thus depends actually on δ and y^δ: $N = N\left(\delta, y^\delta\right)$. But we stick to the simpler notation $N = N(\delta)$.

[3] The idea of using (3.5) was originally introduced in 1982 by Dembo et al [13] in the context of nonlinear well-posed problems in finite dimensional spaces.

Algorithm 1 `K-REGINN`

Input: x_N; (y^{δ_j}, δ_j); F_j; F_j', $j = 0, \ldots, d-1$; μ; τ;

Output: x_N with $\|y_j^{\delta_j} - F_j(x_N)\| \leq \tau\delta_j$, $j = 0, \ldots, d-1$;

$\ell := 0$; $x_0 := x_N$; $c := 0$;
while $c < d$ **do**
  for $j = 0 : d-1$ **do**
    $n := \ell d + j$;
    $b_n^\delta := y_j^{\delta_j} - F_j(x_n)$; $A_n := F_j'(x_n)$;
    if $\|b_n^\delta\| \leq \tau\delta_j$ **then**
      $x_{n+1} := x_n$; $c := c+1$;
    else
      $k := 0$; $s_{n,0} := 0$; choose $k_{\max,n} \in \mathbb{N}$;
      repeat
        calculate $s_{n,k+1} := f(s_{n,k})$ from (3.4)   % The meaning of f is explained in
                                % Remark 20
        $k := k+1$;
      until $\left\|b_n^\delta - A_n s_{n,k}\right\| < \mu \left\|b_n^\delta\right\|$ or $k = k_{\max,n}$
      $x_{n+1} := x_n + s_{n,k}$; $c := 0$;
    end if
  end for
  $\ell := \ell + 1$;
end while
$x_N := x_{\ell d - c}$;

Otherwise choose arbitrarily $k_n \in \{1, ..., k_{REG}\}$ with

$$k_{REG} := \min\left\{k \in \mathbb{N} : \left\|A_n s_{n,k} - b_n^\delta\right\| < \mu \left\|b_n^\delta\right\|\right\}. \tag{3.10}$$

Note that the definition of k_n can be seen as

$$k_n = k_{REG} \wedge k_{\max,n}, \tag{3.11}$$

where $(k_{\max,n})_{n\in\mathbb{N}}$ is an arbitrary sequence of natural numbers and $k_{REG} := 0$ if (3.9) is verified. Observe further that if $k_{\max} < \infty$, where

$$k_{\max} := \max\{k_{\max,n} : n \in \mathbb{N}\}, \tag{3.12}$$

then the sequence $(k_n)_{n\in\mathbb{N}}$ is bounded: $k_n \leq k_{\max}$ for all $n \in \mathbb{N}$.

The equality $x_{n+1} = x_n$ holds whenever (3.9) holds, which means that the algorithm does not alter x_n anymore in case of (3.9) being verified d times in a row. Stop therefore the outer iteration as soon as the discrepancy principle (3.9) is satisfied d consecutively times. Our approximate solution of (3.1) is then x_N where $N = N(\delta)$ is the smallest number satisfying

$$\left\|y_j^{\delta_j} - F_j(x_N)\right\| \leq \tau\delta_j, \quad j = 0, \ldots, d-1. \tag{3.13}$$

See Algorithm 1 for an implementation in pseudocode. See also Remark 20.

Remark 20 *The function $f\colon X \to X$ in the repeat looping of Algorithm 1 represents a general procedure to generate the inner iteration sequence $(s_{n,k})_{0\leq k\leq k_n}$ from* (3.4)*. Although* *the function f depends on the Banach spaces X and Y and on the particular method used to approximate the solution of* (3.4)*, we do not consider any particular method to perform this task in our convergence analysis in Chapter 4. Instead of this, we only assume that some properties of this sequence hold true, without caring about how it is generated. In next section however, we adapt some classical and well-known methods from Hilbert to more general Banach spaces to provide some practical examples of how this sequence could be generated in order to have the required properties used in the convergence analysis of Chapter 4.*

3.1 Solving the inner iteration

In this section we present various iterative methods which can be employed to approximately solve the linear system $A_n s = b_n^\delta$ in order to find the vector s_{n,k_n} (see (3.10) and (3.11)) used to update the outer iteration of `K-REGINN`. We recall that

$$A_n := F'_{[n]}(x_n) \text{ and } b_n^\delta := y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n)$$

represent respectively the F-derivative of the forward operator $F_{[n]}$ at the current iterate x_n and the nonlinear residual. The numbers $p, r > 1$ are fixed and p^* and r^* represent their conjugate numbers respectively.

We suppose in the whole of this section that exact data ($\delta = 0$ in (3.8)) is given. The objectives are to avoid unnecessary complications at this point of the text as well as to ease the notation, and for this last reason, we temporarily omit the superscript δ. We would like to stress however, that all the results presented here can similarly be proven for the noisy data case. Later, in Chapter 4 we turn back to the old notation and consider noisy data again.

The concept in next definition is essential to understand the ideas of the primal gradient methods of Subsection 3.1.1 below.

Definition 21 *Let X be a vector space and $\varphi\colon D(\varphi) \subset X \to \mathbb{R}$ a functional with domain $D(\varphi)$ having non-empty interior. The vector $v \in X$ is a* DESCENT DIRECTION *for φ from $x \in int(\mathrm{D}(\varphi))$ if there exists a positive number $\bar{t}$ such that $\varphi(x + tv) < \varphi(x)$ for all $0 < t \leq \bar{t}$.*

`K-REGINN` updates the current vector x_n adding the vector s_{n,k_n} found in the inner iteration: $x_{n+1} = x_n + s_{n,k_n}$. The vector x_{n+1} is now a new approximation for a solution of (3.1). For this reason, we would like the vectors $s_{n,k}$ to be descent directions for the functional $\psi_n(x) := \frac{1}{r}\left\|F_{[n]}(x) - y_{[n]}\right\|^r$ from x_n.

The function $\frac{1}{r}\|\cdot\|^r$ is F-differentiable if the uniform smoothness of $Y_{[n]}$ is assumed, see Section 2.2. Assume now the F-differentiability of $F_{[n]}$. In this case the chain rule applies to the auxiliary function $\varphi_n(t) := \psi_n(x_n + ts_{n,k})$:

$$\begin{aligned}
\varphi_n'(0) &= \left\langle J_r\left(F_{[n]}(x_n + ts_{n,k}) - y_{[n]}\right), F'_{[n]}(x_n + ts_{n,k})\, s_{n,k}\right\rangle\Big|_{t=0} \\
&= \langle J_r(-b_n), A_n s_{n,k}\rangle = \langle J_r(-b_n), As_{n,k} - b_n\rangle - \langle J_r(-b_n), -b_n\rangle \\
&\leq \|b_n\|^{r-1}\left(\|A_n s_{n,k} - b_n\| - \|b_n\|\right).
\end{aligned}$$

Assuming additionally that $\|A_n s_{n,k} - b_n\| < \|b_n\|$, it follows that

$$\lim_{t \to 0} \frac{\varphi_n(t) - \varphi_n(0)}{t} = \varphi_n'(0) < 0,$$

which implies that there exists $\bar{t} > 0$ such that $\varphi_n(t) < \varphi_n(0)$ for all $0 < t < \bar{t}$. This is equivalent to $\psi_n(x_n + ts_{n,k}) < \psi(x_n)$ for all $0 < t < \bar{t}$, which means that $s_{n,k}$ is a descent direction for ψ_n from x_n.

Though the assumption under the space $Y_{[n]}$ facilitates the above proof through the use of the chain rule in the F-differentiable functions $\frac{1}{r}\left\|\cdot\right\|^{r}$ and $F_{[n]}$, it is not an essential condition. The result actually holds true under weaker restrictions on the space $Y_{[n]}$, which guarantees only G-differentiability of norm-functions, as the next proposition shows.

Proposition 22 *Let X and Y be Banach spaces with Y being smooth and let $F\colon D(F)\subset X\to Y$ be a Gâteaux-differentiable function in $x\in int(\mathrm{D}\left(F\right))$. Further, let $y\in Y$ be a fixed vector and define $A=F'\left(x\right)$ and $b=y-F\left(x\right)$. If $s\in X$ satisfies the inequality $\left\|As-b\right\|<\left\|b\right\|$, then s is a descent direction for the functional $\psi\left(\cdot\right):=\left\|F\left(\cdot\right)-y\right\|$ from x.*

Proof. Define the auxiliary functions $\psi_0\left(t\right):=\frac{1}{r}\left\|b-tAs\right\|^{r}, r>1, \psi_1\left(t\right):=\left\|F\left(x+ts\right)-y\right\|$ and $\psi_2\left(t\right):=\left\|b-tAs\right\|$. As Y is smooth, the duality mapping $J_r\colon Y\to Y^*$ is single-valued and satisfies (see (2.10))

$$\left\langle J_r\left(v\right),w\right\rangle=\frac{\mathrm{d}}{\mathrm{dt}}\left[\frac{1}{r}\left\|v+tw\right\|^{r}\right]\Big|_{t=0}. \tag{3.14}$$

Therefore,

$$\begin{aligned}\lim_{t\to0}\frac{\psi_0\left(t\right)-\psi_0\left(0\right)}{t}&=\psi_0'\left(0\right)=\left\langle J_r\left(b\right),-As\right\rangle=\left\langle J_r\left(b\right),b-As\right\rangle-\left\langle J_r\left(b\right),b\right\rangle\\&\le\left\|b\right\|^{r-1}\left(\left\|As-b\right\|-\left\|b\right\|\right)<0.\end{aligned}$$

The result implies that there exist small numbers $\bar{t},\gamma>0$ such that

$$\frac{\psi_0\left(t\right)-\psi_0\left(0\right)}{t}\le-\gamma<0$$

for all $0<t\le\bar{t}$, which in turn implies that

$$\psi_2\left(t\right)\le\left(-\gamma tr+\psi_2\left(0\right)^{r}\right)^{\frac{1}{r}}.$$

Now,

$$\begin{aligned}\psi_1\left(t\right)&\le\left\|F\left(x+ts\right)-F\left(x\right)-tF'\left(x\right)s\right\|+\psi_2\left(t\right)\\&\le\left\|F\left(x+ts\right)-F\left(x\right)-tF'\left(x\right)s\right\|+\left(\psi_2\left(0\right)^{r}-\gamma tr\right)^{\frac{1}{r}}\end{aligned}$$

and as $\psi_1\left(0\right)=\psi_2\left(0\right)$,

$$\frac{\psi_1\left(t\right)-\psi_1\left(0\right)}{t}\le\frac{\left\|F\left(x+ts\right)-F\left(x\right)-tF'\left(x\right)s\right\|}{t}+\frac{\left(\psi_2\left(0\right)^{r}-\gamma tr\right)^{\frac{1}{r}}-\psi_2\left(0\right)}{t}$$

for all $0<t\le\bar{t}$. Finally, $\left\|F\left(x+ts\right)-F\left(x\right)-tF'\left(x\right)s\right\|/t\to0$ as $t\to0$ because F is Gâteaux-differentiable at x and

$$\lim_{t\to0}\frac{\left(\psi_2\left(0\right)^{r}-\gamma tr\right)^{\frac{1}{r}}-\psi_2\left(0\right)}{t}\overset{\text{L'Hospital}}{=}\lim_{t\to0}-\gamma\left(\psi_2\left(0\right)^{r}-\gamma tr\right)^{-\frac{1}{r^*}}=-\frac{\gamma}{\psi_2\left(0\right)^{\frac{r}{r^*}}}<0.$$

Hence, there exists a number $\bar{t}_1>0$ such that $\psi_1\left(t\right)<\psi_1\left(0\right)$ for all $0<t\le\bar{t}_1$, i.e., $\left\|F\left(x+ts\right)-y\right\|<\left\|F\left(x\right)-y\right\|$. ■

3.1.1 Primal gradient methods

We forget K-REGINN for a while, skip the index n of the outer iteration and concentrate only in the inner iteration. $A\colon X \to Y$ represents a linear operator and $b \in Y$ is a fixed vector.

In Hilbert spaces, the gradient of the functional $\varphi(s) := \frac{1}{2}\|As - b\|^2$ is given by $\nabla\varphi(s) = A^*(As - b)$ and the vector $-\nabla\varphi(s_k)$ is a descent direction for φ from s_k whenever it is not zero. We prove in the next proposition a similar result for smooth Banach spaces.

Proposition 23 *Let X and Y be Banach spaces with Y being smooth. Further, let $A\colon X \to Y$ be linear, $b \in Y$ a fixed vector and $\varphi\colon X \to \mathbb{R}$ the convex functional $\varphi(s) := \frac{1}{r}\|As - b\|^r$ with $r > 1$ fixed. Then φ is G-differentiable in X and if ∇_k represents the G-derivative of φ at s_k, the vector $-j_{p^*}^*(\nabla_k)$ with $p^* > 1$ fixed is either equal zero or a descent direction for φ from s_k.*

Proof. Like in (2.2), we see that

$$\partial\varphi(s) = A^*\partial\left(\frac{1}{r}\|\cdot\|^r\right)(As - b) = A^*J_r(As - b).$$

As Y is smooth, the duality mapping $J_r\colon Y \to Y^*$ is single-valued, and so is the subdifferential $\partial\varphi\colon X \to X^*$, which implies that φ is G-differentiable in X. Further, its G-derivative at s_k is given by $\nabla_k = A^*J_r(As_k - b)$. Suppose that $\nabla_k \neq 0$. Using the auxiliary function

$$\Phi(t) := \varphi\left(s_k + t\left(-j_{p^*}^*(\nabla_k)\right)\right) = \frac{1}{r}\left\|(As_k - b) + t\left(-Aj_{p^*}^*(\nabla_k)\right)\right\|^r$$

we find applying (3.14),

$$\begin{aligned}\lim_{t\to 0}\frac{\Phi(t) - \Phi(0)}{t} &= \Phi'(0) \overset{(3.14)}{=} \left\langle J_r(As_k - b), -Aj_{p^*}^*(\nabla_k)\right\rangle \\ &= -\left\langle \nabla_k, j_{p^*}^*(\nabla_k)\right\rangle = -\|\nabla_k\|^{p^*} < 0.\end{aligned}$$

It follows that there exists a number $\bar{t} > 0$ such that for all $0 < t \leq \bar{t}$ it holds $\Phi(t) < \Phi(0)$, this is, $\varphi\left(s_k + t\left(-j_{p^*}^*(\nabla_k)\right)\right) < \varphi(s_k)$. ■

Proposition 23 shows that in a smooth Banach space Y, the sequence generated by the iterative method

$$s_{k+1} := s_k - \lambda_k j_{p^*}^*(\nabla_k), \tag{3.15}$$

with $p^* > 1$, $s_0 \in X$ and $\lambda_k > 0$ small enough, satisfies the inequality $\varphi(s_{k+1}) < \varphi(s_k)$, i.e.,

$$\|As_{k+1} - b\| < \|As_k - b\| \tag{3.16}$$

as long as $\nabla_k \neq 0$. If additionally $s_0 := 0$, then $\|As_k - b\| < \|b\|$ for all[4] $k \in \mathbb{N}$. The iterative methods defined in this way are called *primal gradient*[5] methods. Algorithm 2 codifies K-REGINN with a primal gradient method in the inner iteration in a smooth Banach space Y. The pieces highlighted in red represent the part of the algorithm exclusively related to (3.15).

If the step-size λ_k in (3.15) can be chosen independently on k, the associated gradient method is called *Landweber* method[6] (in short LW), that is, λ_{LW} =constant. The *Steepest Descent* method (SD) is defined choosing a step-size λ_{SD} satisfying

$$\lambda_{SD} \in \arg\min_{\lambda\in\mathbb{R}^+}\varphi\left(s_k - \lambda j_{p^*}^*(\nabla_k)\right),$$

[4] In view of Proposition 22, we see that for $A = A_n$ and $b = b_n$, the vectors $s_k = s_{n,k}$ are descent directions for the functional $\left\|F_{[n]}(\cdot) - y_{[n]}\right\|$ from x_n.

[5] Here the iteration occurs in the primal space X, in contrast with the so-called *dual methods* where the iteration happens in the dual space X^*, see Subsection 3.1.2.

[6] To homage the relevant work of L. Landweber [33].

Algorithm 2 K-REGINN with primal gradient inner iteration

Input: x_N; (y^{δ_j}, δ_j); F_j; F_j', $j = 0, \ldots, d-1$; μ; τ; $p, r > 1$;

Output: x_N with $\|y_j^{\delta_j} - F_j(x_N)\| \le \tau\delta_j$, $j = 0, \ldots, d-1$;

$\ell := 0$; $x_0 := x_N$; $c := 0$;
while $c < d$ **do**
  for $j = 0 : d-1$ **do**
    $n := \ell d + j$;
    $b_n^\delta := y_j^{\delta_j} - F_j(x_n)$; $A_n := F_j'(x_n)$;
    if $\|b_n^\delta\| \le \tau\delta_j$ **then**
      $x_{n+1} := x_n$; $c := c + 1$;
    else
      $k := 0$; $s_{n,0} := 0$; choose $k_{\max,n} \in \mathbb{N}$;
      repeat
        $\nabla_{n,k} := A_n^* J_r(A_n s_{n,k} - b_n^\delta)$;
        choose $\lambda_{n,k} > 0$ and $j_{p^*}^*(\nabla_{n,k}) \in J_{p^*}^*(\nabla_{n,k})$;
        $s_{n,k+1} := s_{n,k} - \lambda_{n,k} j_{p^*}^*(\nabla_{n,k})$;
        $k := k + 1$;
      until $\left\|b_n^\delta - A_n s_{n,k}\right\| < \mu \left\|b_n^\delta\right\|$ or $k = k_{\max,n}$
      $x_{n+1} := x_n + s_{n,k}$; $c := 0$;
    end if
  end for
  $\ell := \ell + 1$;
end while
$x_N := x_{\ell d - c}$;

if such a minimizer exists. Assuming this is the case and additionally assuming that the function φ is F-differentiable (the second assumption is true for instance, if Y is uniformly smooth), one can apply the chain rule to find

$$\begin{aligned}
\frac{d}{d\lambda}\varphi\left(s_k - \lambda j_{p^*}^*(\nabla_k)\right)\Big|_{\lambda=\lambda_{SD}} &= \left\langle \varphi'\left(s_k - \lambda j_{p^*}^*(\nabla_k)\right), -j_{p^*}^*(\nabla_k)\right\rangle\Big|_{\lambda=\lambda_{SD}} \\
&= -\left\langle A^* J_r(A s_{k+1} - b), j_{p^*}^*(\nabla_k)\right\rangle = -\left\langle \nabla_{k+1}, j_{p^*}^*(\nabla_k)\right\rangle .
\end{aligned}$$

It follows that, similarly to Hilbert spaces, the gradient of two consecutive iterates are "orthogonal" in the sense that $\left\langle j_{p^*}^*(\nabla_k), \nabla_{k+1}\right\rangle = 0$. Further, the inequality $\varphi(s_{k+1}) \le \varphi\left(s_k - \lambda j_{p^*}^*(\nabla_k)\right)$ is immediately verified for all $\lambda \ge 0$ and consequently (3.16) holds. Due to the nonlinearity of the duality mapping, an explicit formula for λ_{SD} is nevertheless difficult to be achieved. The uniqueness of a minimizer is guaranteed for instance, if A is injective and Y is strictly convex because in this case the function $\frac{1}{r}\|\cdot\|^r$ is strictly convex and $Ax_1 - b \neq Ax_2 - b$ whenever $x_1 \neq x_2$, which implies the strict convexity of φ and consequently the desired result.

Suppose now that Y is a $r-$smooth Banach space, then there exists a positive number

$\overline{C}_r$ (see (2.18)) such that for all $\lambda \geq 0$ and $s_{k+1} = s_k - \lambda j_{p^*}^* (\nabla_k)$,

$$\begin{aligned}
\frac{1}{r} \|As_{k+1} - b\|^r &= \frac{1}{r} \left\|(As_k - b) - \lambda A j_{p^*}^* (\nabla_k)\right\|^r \\
&\leq \frac{1}{r} \|As_k - b\|^r - \left\langle J_r (As_k - b), \lambda A j_{p^*}^* (\nabla_k)\right\rangle + \frac{\overline{C}_r}{r} \left\|\lambda A j_{p^*}^* (\nabla_k)\right\|^r \\
&= \frac{1}{r} \|As_k - b\|^r - \lambda \|\nabla_k\|^{p^*} + \frac{\overline{C}_r}{r} \lambda^r \left\|A j_{p^*}^* (\nabla_k)\right\|^r,
\end{aligned}$$

which implies that

$$\varphi (s_{k+1}) - \varphi (s_k) \leq -\lambda \|\nabla_k\|^{p^*} + \frac{\overline{C}_r}{r} \lambda^r \left\|A j_{p^*}^* (\nabla_k)\right\|^r =: f(\lambda). \tag{3.17}$$

The above inequality is verified for each method defined via $s_{k+1} = s_k - \lambda j_{p^*}^* (\nabla_k)$ and in particular, since the step-size λ_{SD} minimizes the difference $\varphi (s_{k+1}) - \varphi (s_k)$, inequality (3.17) holds for this method using an arbitrary $\lambda \geq 0$ in the rightmost side. The optimality condition $f'(\lambda) = 0$ drives to the step-size

$$\lambda_{MSD}^{r-1} := C_0 \frac{\|\nabla_k\|^{p^*}}{\left\|A j_{p^*}^* (\nabla_k)\right\|^r}, \tag{3.18}$$

with $C_0 := 1/\overline{C}_r$. The associated gradient method is called *Modified Steepest Descent* (MSD) method. If Y is a Hilbert space and $r = 2$, then the polarization identity (2.5) shows that $\overline{C}_r = 1$ can be chosen. In this case, the SD and MSD methods coincide and have the same step-size:

$$\lambda = \frac{\|\nabla_k\|^{p^*}}{\left\|A j_{p^*}^* (\nabla_k)\right\|^2}.$$

If X is a Hilbert space too, then this step-size is given by the expression (for $p = r = 2$):

$$\lambda = \frac{\|\nabla_k\|^2}{\|A\nabla_k\|^2} = \frac{\|A^* (As_k - b)\|^2}{\|AA^* (As_k - b)\|^2}. \tag{3.19}$$

Changing the definition of C_0 with an arbitrary number satisfying the inequality $0 < C_0 < r/\overline{C}_r$, we observe that the choice $\lambda_k \in (0, \lambda_{MSD}]$ implies that $f(\lambda_k) < 0$, which in view of (3.17) implies (3.16). The inequalities $0 < \lambda_k \leq \lambda_{MSD}$ in combination with (3.17) additionally imply that

$$\begin{aligned}
\varphi (s_{k+1}) - \varphi (s_k) &\leq f(\lambda_k) \leq -\lambda_k \|\nabla_k\|^{p^*} + \lambda_k \frac{\overline{C}_r}{r} \lambda_{MSD}^{r-1} \left\|A j_{p^*}^* (\nabla_k)\right\|^r \\
&= -\lambda_k \|\nabla_k\|^{p^*} + \lambda_k \frac{\overline{C}_r C_0}{r} \|\nabla_k\|^{p^*} = -C_1 \lambda_k \|\nabla_k\|^{p^*} < 0,
\end{aligned} \tag{3.20}$$

where $C_1 := 1 - \overline{C}_r C_0/r > 0$. The above result is true for each primal gradient method with step-size satisfying $\lambda_k \in (0, \lambda_{MSD}]$ with $0 < C_0 < r/\overline{C}_r$. It holds for SD too (in principle not for $\lambda_k = \lambda_{SD}$ in the rightmost term because we do not know whether $0 < \lambda_{SD} \leq \lambda_{MSD}$, but for an arbitrary $\lambda_k \in (0, \lambda_{MSD}]$). Now, if $\lambda_k \in [c\lambda_{MSD}, \lambda_{MSD}]$ with $0 < c < 1$ being independent of k,

$$\begin{aligned}
\left(\lambda_k \|\nabla_k\|^{p^*}\right)^{r-1} &\geq \left(c\lambda_{MSD} \|\nabla_k\|^{p^*}\right)^{r-1} \\
&\geq c^{r-1} C_0 \frac{\|\nabla_k\|^{p^* - r(p^*-1)}}{\|A\|^r} \|\nabla_k\|^{p^*(r-1)} = \frac{c^{r-1} C_0}{\|A\|^r} \|\nabla_k\|^r,
\end{aligned}$$

which implies that $\lambda_k \|\nabla_k\|^{p^*} \gtrsim \|\nabla_k\|^{r^*}$. From (3.20),

$$\sum_{k=0}^{\infty} \|\nabla_k\|^{r^*} \lesssim \sum_{k=0}^{\infty} \lambda_k \|\nabla_k\|^{p^*} \lesssim \sum_{k=0}^{\infty} \varphi(s_k) - \varphi(s_{k+1}) \leq \varphi(s_0) < \infty.$$

It follows that

$$\|\nabla_k\| \to 0 \text{ as } k \to \infty. \tag{3.21}$$

Hence, there exists a constant $C_2 > 0$ independent of k such that $\|\nabla_k\| \leq C_2$. The result is true for each primal gradient method with step-size λ_k in the interval $[c\lambda_{MSD}, \lambda_{MSD}]$ with $0 < c < 1$ fixed but arbitrary. Although we do not know if the inequality $c\lambda_{MSD} \leq \lambda_{SD} \leq \lambda_{MSD}$ is true, (3.21) is ensured for SD method too because (3.20) holds for this method with an arbitrary $\lambda_k \in (0, \lambda_{MSD}]$. In particular, (3.20) holds for SD with $\lambda_k = \lambda_{MSD}$ for example, which implies (3.21).

Finally, the inequality $p \leq r$ implies that $p^* - r(p^* - 1) \leq 0$, which in turn implies

$$\lambda_{MSD}^{r-1} \geq C_0 \frac{\|\nabla_k\|^{p^*-r(p^*-1)}}{\|A\|^r} \geq C_0 \frac{C_2^{p^*-r(p^*-1)}}{\|A\|^r}.$$

Choosing

$$\lambda_{LW} := \frac{C_0^{r^*-1} C_2^{r^*-p^*}}{\|A\|^{r^*}}, \tag{3.22}$$

we conclude that $\lambda_{LW} \in (0, \lambda_{MSD}]$ and due to (3.20), the inequality

$$\varphi(s_{k+1}) - \varphi(s_k) \leq -C_2 \lambda_{LW} \|\nabla_k\|^{p^*}$$

is valid for LW method. As λ_{LW} is constant, the last inequality immediately implies (3.21).

In summary, we have proven that:

- If it is well-defined, the SD method always satisfies (3.16). Additionally, if Y is $r-$smooth, then (3.21) holds true.
- The LW method is well-defined, satisfies (3.16), (3.20) and (3.21) whenever Y is $r-$smooth and $p \leq r$.
- Y $r-$smooth implies that each primal gradient method with step-size $\lambda_k \in [c\lambda_{MSD}, \lambda_{MSD}]$ and $0 < c < 1$ (in particular, MSD itself) satisfies (3.16), (3.20) and (3.21).

Lemma 24 *Let X and Y be Banach spaces and $(s_k)_{k\in\mathbb{N}}$ be a sequence generated by the iterative method* (3.15) *with $s_0 = 0$. Suppose that* (3.21) *and* (3.20) *hold. Then there exists a constant $C \geq 0$ such that*

$$\lim_{k\to\infty} \|As_k - b\|^r \leq \frac{1}{1+C}\left(\|As - b\|^r + C\|b\|^r\right), \text{ for all } s \in X. \tag{3.23}$$

Proof. From (3.21), it is possible to choose a subsequence $\left(\nabla_{k_j}\right)_{j\in\mathbb{N}}$ such that $\left\|\nabla_{k_j}\right\| \leq \|\nabla_m\|$ for $m \leq k_j$. Let $s \in X$ be an arbitrary vector. As $\nabla_{k_j} \in \partial\varphi\left(s_{k_j}\right)$, it follows from definition of subgradient that

$$\begin{aligned}\varphi\left(s_{k_j}\right) &\leq \varphi(s) + \left\langle \nabla_{k_j}, s_{k_j}\right\rangle - \left\langle \nabla_{k_j}, s\right\rangle \overset{(3.15)}{=} \varphi(s) - \sum_{m=0}^{k_j-1} \lambda_m \left\langle \nabla_{k_j}, j_{p^*}^*(\nabla_m)\right\rangle - \left\langle \nabla_{k_j}, s\right\rangle \\ &\leq \varphi(s) + \sum_{m=0}^{k_j-1} \lambda_m \|\nabla_m\|^{p^*} + \left\|\nabla_{k_j}\right\| . \|s\|,\end{aligned}$$

and employing (3.20) we arrive at

$$\varphi\left(s_{k_j}\right) \leq \varphi(s) + C\left(\varphi(s_0) - \varphi\left(s_{k_j}\right)\right) + \left\|\nabla_{k_j}\right\| . \|s\|,$$

with $C := 1/C_1 > 0$. Hence

$$(1+C)\,\varphi\left(s_{k_j}\right) \leq \varphi(s) + C\varphi(s_0) + \left\|\nabla_{k_j}\right\| . \|s\|.$$

Now, (3.20) implies that $(\varphi(s_k))_{k\in\mathbb{N}}$ is a positive decreasing sequence, hence convergent. It follows that

$$\begin{aligned}\lim_{k\to\infty}\varphi(s_k) = \lim_{j\to\infty}\varphi\left(s_{k_j}\right) &\leq \lim_{j\to\infty}\frac{1}{1+C}\left[\varphi(s) + C\varphi(0) + \left\|\nabla_{k_j}\right\| . \|s\|\right] \\ &= \frac{1}{1+C}\left[\varphi(s) + C\varphi(0)\right].\end{aligned}$$

■

3.1.2 Dual gradient methods

We start this subsection using a motivation in Hilbert spaces to introduce a new method, which we call *Decreasing Error* method (in short DE). Using the best of our knowledge, a similar method was introduced by Fridman in [17] in the context of linear problems in Hilbert spaces. In this particular situation, this method results in the gradient method with the fastest decreasing error and this is the reason why it was called later, the Minimal Error method, see e.g. [41]. As far as we know, the adaptation of DE method to nonlinear problems in Banach spaces is novel.

Suppose momentarily that there exists a solution s^+ of the linear equation $As = b$. Define now the gradient method $s_{k+1} = s_k - \lambda_k\nabla_k$, with $\lambda_k > 0$ and ∇_k being the gradient of $\varphi(s) = \frac{1}{2}\|As - b\|^2$ at s_k, i.e., $\nabla_k = A^*(As_k - b)$. The main idea consists of finding a calculable upper bound $\overline{\lambda}_k > 0$ for the step-size λ_k such that the inequality $\lambda_k \leq \overline{\lambda}_k$ implies the monotonically reduction of the error in each step.

Using the polarization identity (2.5) we find

$$\frac{1}{2}\left\|s_{k+1} - s^+\right\|^2 = \frac{1}{2}\left\|s_k - s^+ - \lambda_k\nabla_k\right\|^2 = \frac{1}{2}\left\|s_k - s^+\right\|^2 - \lambda_k\left\langle s_k - s^+, \nabla_k\right\rangle + \frac{1}{2}\lambda_k^2\left\|\nabla_k\right\|^2.$$

Thus,

$$\frac{1}{2}\left\|s_{k+1} - s^+\right\|^2 - \frac{1}{2}\left\|s_k - s^+\right\|^2 = -\lambda_k\left\langle s_k - s^+, \nabla_k\right\rangle + \frac{1}{2}\lambda_k^2\left\|\nabla_k\right\|^2 := g(\lambda_k). \tag{3.24}$$

Observe that $g(\lambda_k) < 0$ if and only if

$$\lambda_k < 2\frac{\left\langle s_k - s^+, \nabla_k\right\rangle}{\|\nabla_k\|^2} = 2\frac{\left\langle A\left(s_k - s^+\right), As_k - b\right\rangle}{\|\nabla_k\|^2} = 2\frac{\|As_k - b\|^2}{\|\nabla_k\|^2}.$$

We conclude that the choice $\lambda_k \in \left(0, \overline{\lambda}_k\right]$ with the calculable step-size

$$\overline{\lambda}_k := C_0\frac{\|As_k - b\|^2}{\|\nabla_k\|^2} \tag{3.25}$$

and $C_0 < 2$ implies that $g(\lambda_k) < 0$, as we wanted. Observe that $C_0 = 1$ transforms $\overline{\lambda}_k$ in the optimal step-size (in the sense that the resulting method has, among all gradient methods, the error which decreases with maximal speed), which is obtained from $g'\left(\overline{\lambda}_k\right) = 0$. The gradient method associated with the step-size $\overline{\lambda}_k$ with $C_0 = 1$ is just the Minimal Error method introduced in [17]. Note further that the step-size (3.25) is simpler to be computed than the one of Steepest Descent method (3.19).

To enlarge the above results in order to guarantee their validity for nonlinear problems in Banach spaces, more general results as those shown in last subsection are required. For the proper adjustment of DE method to `K-REGINN`, we suppose for the rest of this subsection that X is an uniformly smooth and uniformly convex Banach space. Both restrictions together ensure that the duality mapping $J_p \colon X \to X^*$, $1 < p < \infty$, is single valued, continuous, invertible and with continuous inverse given by $J_p^{-1} = J_{p^*}^* \colon X^* \to X^{**} \cong X$. This result provides free access to the dual space X^* in the sense that it is always possible to transfer a vector from X to its dual space X^*, perform an iteration and then come back to the original space in a stable way.

We further assume the next assumption on the inverse problem $F(x) = y$:

Assumption 1 (a) *There exists a solution* $x^+ \in X$ *of equation* (3.1) *satisfying*

$$B_\rho\left(x^+, \Delta_p\right) := \left\{v \in X : \Delta_p\left(x^+, v\right) < \rho\right\} \subset \mathrm{D}\left(F\right),$$

where $\rho > 0$ *and* $p > 1$ *are fixed numbers and the Bregman distance* Δ_p *is defined in* (2.19).

(b) *All the functions* F_j, $j = 0, \ldots, d-1$, *are continuously Fréchet differentiable in* $B_\rho\left(x^+, \Delta_p\right)$ *and their derivatives satisfy*

$$\left\|F_j'(v)\right\| \leq M \textit{ for all } v \in B_\rho\left(x^+, \Delta_p\right) \textit{ and } j = 0, \ldots, d-1,$$

where $M > 0$ *is a constant.*

(c) *(Tangential Cone Condition (TCC)): There exists a constant* $0 \leq \eta < 1$ *such that*

$$\left\|F_j(v) - F_j(w) - F_j'(w)\left(v - w\right)\right\| \leq \eta \left\|F_j(v) - F_j(w)\right\|,$$

for all $v, w \in B_\rho\left(x^+, \Delta_p\right)$ *and* $j = 0, \ldots, d-1$.

Before starting the analysis in more general Banach spaces, we stay a little longer in Hilbert spaces and observe how could be possible to employ the DE method as inner iteration of `K-REGINN`. Similar to before, we define the iteration $s_{n,k+1} = s_{n,k} - \lambda_{n,k}\nabla_{n,k}$, where $\lambda_{n,k} > 0$ and $\nabla_{n,k} = A_n^*\left(A_n s_{n,k} - b_n\right)$, with $A_n := F_{[n]}'\left(x_n\right)$ and $b_n := y_{[n]} - F_{[n]}\left(x_n\right)$. Observe that the most suitable vector to be approximated in the inner iteration is not s^+, but $e_n := x^+ - x_n$. The reason is quite simple: if $s_{n,k_n} = e_n$, then $x_{n+1} = x_n + e_n = x^+$. Applying an idea similar to (3.24), we derive the equality

$$\begin{aligned}\frac{1}{2}\left\|s_{n,k+1} - e_n\right\|^2 - \frac{1}{2}\left\|s_{n,k} - e_n\right\|^2 &= -\lambda_{n,k}\left\langle s_{n,k} - e_n, \nabla_{n,k}\right\rangle + \frac{1}{2}\lambda_{n,k}^2\left\|\nabla_{n,k}\right\|^2 \\ &:= g\left(\lambda_{n,k}\right)\end{aligned} \tag{3.26}$$

and conclude that $g\left(\lambda_{n,k}\right) < 0$ if and only if

$$\lambda_{n,k} < 2\frac{\left\langle s_{n,k} - e_n, \nabla_{n,k}\right\rangle}{\left\|\nabla_{n,k}\right\|^2}.$$

Note that the term $\left\langle s_{n,k} - e_n, \nabla_{n,k}\right\rangle$ depends on the unavailable information e_n. But,

$$\begin{aligned}\left\langle s_{n,k} - e_n, \nabla_{n,k}\right\rangle &= \left\langle A_n\left(s_{n,k} - e_n\right), A_n s_{n,k} - b_n\right\rangle \\ &= \left\langle A_n s_{n,k} - b_n, A_n s_{n,k} - b_n\right\rangle - \left\langle A_n e_n - b_n, A_n s_{n,k} - b_n\right\rangle \\ &\geq \left\|A_n s_{n,k} - b_n\right\|^2 - \left\|A_n s_{n,k} - b_n\right\| . \left\|A_n e_n - b_n\right\| .\end{aligned} \tag{3.27}$$

Applying now the TCC, Assumption $1(c)$, and observing that $\|b_n\| \leq \frac{1}{\mu}\|A_n s_{n,k} - b_n\|$ for $k = 0, ..., k_n,$, see (3.11),

$$\|A_n e_n - b_n\| = \left\|F_{[n]}\left(x^+\right) - F_{[n]}\left(x_n\right) - F'_{[n]}\left(x_n\right)\left(x^+ - x_n\right)\right\| \tag{3.28}$$
$$\leq \eta\|b_n\| \leq \frac{\eta}{\mu}\|A_n s_{n,k} - b_n\|.$$

Therefore

$$\langle s_{n,k} - e_n, \nabla_{n,k}\rangle \geq K_1\|A_n s_{n,k} - b_n\|^2,$$

with $K_1 := 1 - \frac{\eta}{\mu}$ (which is positive for $\mu \in (\eta, 1)$). Thus, $\lambda_{n,k} \in \left(0, \overline{\lambda}_{n,k}\right)$ with

$$\overline{\lambda}_{n,k} := 2K_1 \frac{\|A_n s_{n,k} - b_n\|^2}{\|\nabla_{n,k}\|^2}$$

implies $\|s_{n,k+1} - e_n\| < \|s_{n,k} - e_n\|$.

At this point we clearly see why is important to observe the outer and inner iteration of K-REGINN simultaneously for this method. An arbitrary linear operator $A : X \to Y$ and an arbitrary vector $b \in Y$ do not necessarily verify (3.28). We actually need to use $A = A_n$ and $b = b_n$.

See that the monotonicity of the error in the outer iteration is inherited from the inner iteration because

$$\left\|x_{n+1} - x^+\right\| = \|s_{n,k_n} - e_n\| < ... < \|s_{n,0} - e_n\| = \left\|x_n - x^+\right\|.$$

Remark 25 *To expand the above ideas to more general Banach spaces, we need to assume that X is $s-$convex for some $s > 1$, which leave us with two possible approaches. In the first one, we assume that $s = p$, which means that the index s coincides with the index used to define the duality mapping J_p. In this case, the numbers p and s in inequalities* (2.22), (2.23), (2.24) *and* (2.25) *coincide. This framework strongly facilitates the convergence analysis of K-REGINN presented in Chapter 4. However, when performing the numerical experiments in Chapter 5 we are interested in the use of the Lebesgue spaces $L^p(\Omega)$ with $1 < p < 2$, which are not $p-$convex but $2-$convex spaces. This structure forces the use of the normalized duality mapping J_2 in $L^p(\Omega)$ instead of the standard option J_p. At first, this is not a big problem, because J_2 can be calculated in $L^p(\Omega)$ via* (2.14). *But the numerical experiments we have done suggested that this is not the best approach to guarantee good reconstructions.*

In order to fix this problem, an alternative and more general approach can be applied assuming the space X to be $s-$convex with $p \leq s$, which makes possible the use of the duality mapping J_p in the $s-$convex space $L^p(\Omega)$ ($s = \max\{p, 2\}$). For this reason, we have chosen to employ this approach to develop our theory. This procedure actually results in a much more complicated theory, but as a reward we are free to use a more suitable duality mapping and expect to achieve better reconstructions in our numerical experiments in Chapter 5.

We now clarify how to engage the dual gradient methods in combination with K-REGINN in Banach spaces. Assume that X is uniformly smooth and $s-$convex with some $s \geq p$. Suppose that the current outer iterate x_n of K-REGINN is well-defined and define the convex functional $\varphi_n(s) := \frac{1}{r}\|A_n s - b_n\|^r$, $r > 1$. The *dual gradient* methods are now defined by the following iteration:

$$J_p(z_{n,k+1}) := J_p(z_{n,k}) - \lambda_{n,k}\nabla_{n,k}, \tag{3.29}$$

where $\lambda_{n,k} > 0$, $\nabla_{n,k} \in \partial\varphi_n(s_{n,k}) = A_n^* J_r(A_n s_{n,k} - b_n)$, $s_{n,k} := z_{n,k} - x_n$ and $s_{n,0} := 0$. Observe that $z_{n,0} = x_n + s_{n,0} = x_n$ as well as $z_{n,k_n} = x_n + s_{n,k_n} = x_{n+1}$. We codify

Algorithm 3 `K-REGINN` with dual gradient inner iteration

Input: x_N; (y^{δ_j}, δ_j); F_j; F_j', $j = 0, \ldots, d-1$; μ; τ; $p, r > 1$;

Output: x_N with $\|y_j^{\delta_j} - F_j(x_N)\| \le \tau\delta_j$, $j = 0, \ldots, d-1$;

$\ell := 0$; $x_0 := x_N$; $c := 0$;
while $c < d$ **do**
  for $j = 0 : d-1$ **do**
    $n := \ell d + j$;
    $b_n^\delta := y_j^{\delta_j} - F_j(x_n)$; $A_n := F_j'(x_n)$;
    if $\|b_n^\delta\| \le \tau\delta_j$ **then**
      $x_{n+1} := x_n$; $c := c + 1$;
    else
      $k := 0$; $s_{n,0} := 0$; choose $k_{\max,n} \in \mathbb{N}$;
      repeat
        choose $\lambda_{n,k} > 0$ and $\nabla_{n,k} \in A_n^* J_r(A_n s_{n,k} - b_n^\delta)$;
        $z_{n,k} := s_{n,k} + x_n$;
        $J_p(z_{n,k+1}) := J_p(z_{n,k}) - \lambda_{n,k}\nabla_{n,k}$;
        $s_{n,k+1} := J_{p^*}^*(J_p(z_{n,k+1})) - x_n$;
        $k := k + 1$;
      until $\left\|b_n^\delta - A_n s_{n,k}\right\| < \mu \left\|b_n^\delta\right\|$ or $k = k_{\max,n}$
      $x_{n+1} := x_n + s_{n,k}$; $c := 0$;
    end if
  end for
  $\ell := \ell + 1$;
end while
$x_N := x_{\ell d - c}$;

`K-REGINN` with a dual gradient method in the inner iteration in Algorithm 3. The fragments highlighted in red represent here the part of the algorithm corresponding to (3.29).

To imitate the polarization identity, which is necessary to derive (3.26), we change the functional $\frac{1}{2}\|x^+ - \cdot\|^2$ into the Bregman distance $\Delta_p(x^+, \cdot)$. Assume that[7] $x_n \in B_\rho(x^+, \Delta_p)$, that is, $\Delta_p(x^+, x_n) < \rho$. In view of (2.20), this inequality implies that $\|x_n\| \le C_{\rho,x^+}$, with

$$C_{\rho,x^+} := 2p^*\left(\left\|x^+\right\| \vee \rho^{1/p}\right). \tag{3.30}$$

Our goal now is to define a calculable step-size $\lambda_{DE} = \lambda_{DE}(n,k) > 0$ such that the inequalities $0 < \lambda_{n,k} \le \lambda_{DE}$ will imply the monotonicity of the error in the inner iteration:

$$\Delta_p\left(x^+, z_{n,k+1}\right) < \Delta_p\left(x^+, z_{n,k}\right). \tag{3.31}$$

Further, we prove that the step-size λ_{LW} of the Landweber method and λ_{MSD} of the Modified Steepest Descent method satisfy the required inequality and consequently the monotonicity property (3.31) is ensured for these methods.

The definition of λ_{DE} depends on an uniformly bound in n and k for the generated sequence $(z_{n,k})_{0\le k\le k_n}$, $n \in \mathbb{N}$, see (3.32) and (3.39). But in order to prove that this

[7]In Theorem 38, we will prove that if `K-REGINN` starts with $x_0 \in B_\rho(x^+, \Delta_p)$, then all the outer iterations x_n belong to the same ball, see 4.10.

sequence is uniformly bounded, it is necessary to use the property (3.31), which in turn depends on the definition of λ_{DE}. This reasoning suggests that an induction argument is necessary to prove all these properties at the same time. Note that $z_{n,0} = x_n \in B_\rho(x^+, \Delta_p)$ and therefore $\|z_{n,0}\| \leq C_{\rho,x^+}$. We prove now by induction that

$$\|z_{n,k}\| \leq C_{\rho,x^+} \tag{3.32}$$

for $k = 0, ..., k_n$. Assume that (3.32) holds true for some $k < k_n$ as well as $\Delta_p(x^+, z_{n,\ell}) < \Delta_p(x^+, z_{n,\ell-1})$ for $\ell = 1, ..., k$.

The three points identity (2.21) is applied to replace (3.26). Definition (3.29) together with $e_n = x^+ - x_n$ yields

$$\begin{aligned} \Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, z_{n,k}\right) &= \Delta_p\left(z_{n,k}, z_{n,k+1}\right) \\ &\quad + \left\langle J_p\left(z_{n,k+1}\right) - J_p\left(z_{n,k}\right), s_{n,k} - e_n\right\rangle \\ &= \Delta_p\left(z_{n,k}, z_{n,k+1}\right) - \lambda_{n,k}\left\langle \nabla_{n,k}, s_{n,k} - e_n\right\rangle =: g\left(\lambda_{n,k}\right). \end{aligned} \tag{3.33}$$

Making use of the properties of the duality mapping, we obtain, similarly to (3.27),

$$\begin{aligned} \left\langle \nabla_{n,k}, s_{n,k} - e_n\right\rangle &= \left\langle j_r\left(A_n s_{n,k} - b_n\right), A_n\left(s_{n,k} - e_n\right)\right\rangle \\ &= \left\langle j_r\left(A_n s_{n,k} - b_n\right), A_n s_{n,k} - b_n\right\rangle - \left\langle j_r\left(A_n s_{n,k} - b_n\right), A_n e_n - b_n\right\rangle \\ &\geq \left\|A_n s_{n,k} - b_n\right\|^r - \left\|A_n s_{n,k} - b_n\right\|^{r-1}\left\|A_n e_n - b_n\right\|. \end{aligned} \tag{3.34}$$

As X is $s-$convex, X^* is s^*-smooth and since $p^* \geq s^*$, there exists a positive constant C_{p^*,s^*} (see (2.24)) such that

$$\begin{aligned} \Delta_p\left(z_{n,k}, z_{n,k+1}\right) &= \Delta_{p^*}\left(J_p\left(z_{n,k+1}\right), J_p\left(z_{n,k}\right)\right) \\ &\leq C_{p^*,s^*}\Big(\left\|J_p\left(z_{n,k}\right)\right\|^{p^*-s^*}\left\|J_p\left(z_{n,k+1}\right) - J_p\left(z_{n,k}\right)\right\|^{s^*} \\ &\qquad \vee \left\|J_p\left(z_{n,k+1}\right) - J_p\left(z_{n,k}\right)\right\|^{p^*}\Big) \\ &\leq C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)}\lambda_{n,k}^{s^*}\left\|\nabla_{n,k}\right\|^{s^*} \vee \lambda_{n,k}^{p^*}\left\|\nabla_{n,k}\right\|^{p^*}\right). \end{aligned} \tag{3.35}$$

The last inequality is the exact point where the bound (3.32) is used. We impose now a restriction on $\lambda_{n,k}$: if it is possible to find $\lambda_{n,k}$ satisfying

$$C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)}\lambda_{n,k}^{s^*}\left\|\nabla_{n,k}\right\|^{s^*} \vee \lambda_{n,k}^{p^*}\left\|\nabla_{n,k}\right\|^{p^*}\right) \overset{!}{\leq} C_0\lambda_{n,k}\left\|A_n s_{n,k} - b_n\right\|^r \tag{3.36}$$

for some $0 < C_0 < 1$, then putting it together with (3.34) in (3.33) we arrive at

$$g\left(\lambda_{n,k}\right) \leq \lambda_{n,k}\left\|A_n s_{n,k} - b_n\right\|^{r-1}\left[\left\|A_n e_n - b_n\right\| - \left(1 - C_0\right)\left\|A_n s_{n,k} - b_n\right\|\right]. \tag{3.37}$$

The above inequality is the key to prove that $g(\lambda_{n,k}) < 0$ and it is fundamental for our convergence analysis in Chapter 4. Observe that it does not depend on the TCC (Assumption 1(c), page 31), but if we use it, together with $k < k_n$ and the definition (3.11), we obtain like in (3.28) the bound

$$g\left(\lambda_{n,k}\right) \leq -\lambda_{n,k}C_3\left\|A_n s_{n,k} - b_n\right\|^r < 0, \tag{3.38}$$

with $C_3 := 1 - C_0 - \eta/\mu$ (which is positive if $\eta < 1 - C_0$ and $\eta/(1 - C_0) < \mu < 1$). This inequality ensure that (3.31) holds and accordingly

$$\Delta_p\left(x^+, z_{n,k+1}\right) < \Delta_p\left(x^+, z_{n,k}\right) < ... < \Delta_p\left(x^+, z_{n,0}\right) < \rho.$$

It follows that $\|z_{n,k+1}\| \leq C_{\rho,x^+}$, which completes the induction proof.

It remains only to find $\lambda_{DE} > 0$ such that (3.36) holds for all $\lambda_{n,k} \le \lambda_{DE}$. Looking at (3.36) , we see that this is the case if

$$\lambda_{DE} := C_1 \lambda_{DE,s} \wedge C_2 \lambda_{DE,p} \tag{3.39}$$

where

$$\lambda_{DE,\ell} := \frac{\|A_n s_{n,k} - b_n\|^{r(\ell-1)}}{\|\nabla_{n,k}\|^{\ell}},$$

with $C_1 := C_0^{s-1}/C_{p^*,s^*}^{s-1} C_{\rho,x^+}^{s-p}$ and $C_2 := C_0^{p-1}/C_{p^*,s^*}^{p-1}$.

Remark 26 *As $z_{n,k} - x^+ = s_{n,k} - e_n$, it comes from* (3.34) *and* (3.28) ,

$$\|\nabla_{n,k}\| \left\|z_{n,k} - x^+\right\| \ge \langle \nabla_{n,k}, s_{n,k} - e_n \rangle \ge \left(1 - \frac{\eta}{\mu}\right) \|A_n s_{n,k} - b_n\|^r .$$

Since the sequences $(z_{n,k} - x^+)_{k \le k_n}$, $n \in \mathbb{N}$, are uniformly bounded, we obtain

$$\|A_n s_{n,k} - b_n\|^r \lesssim \|\nabla_{n,k}\| .$$

Finally, $p \le s$ implies $\|A_n s_{n,k} - b_n\|^{r(s-p)} \lesssim \|\nabla_{n,k}\|^{s-p}$ and consequently

$$\frac{\|A_n s_{n,k} - b_n\|^{r(s-1)}}{\|\nabla_{n,k}\|^{s}} \lesssim \frac{\|A_n s_{n,k} - b_n\|^{r(p-1)}}{\|\nabla_{n,k}\|^{p}}.$$

Therefore, $\lambda_{DE,s} \lesssim \lambda_{DE,p}$ for all n and k. This means that $\lambda_{DE,s} \lesssim \lambda_{DE} \lesssim \lambda_{DE,p}$. Further, if a small enough constant replaces the constant C_1 in definition of λ_{DE}, then $\lambda_{DE} = \lambda_{DE,s}$ can be chosen and the property

$$\lambda_{n,k} \le \lambda_{DE} \Longrightarrow \Delta_p\left(x^+, z_{n,k+1}\right) < \Delta_p\left(x^+, z_{n,k}\right)$$

still holds.

Remark 27 *As long as a minimizer s_n^+ of $\varphi_n(s) = \frac{1}{r}\|A_n s - b_n\|^r$ exists and $\lambda_{n,k} \in (0, \lambda_{DE}]$, the sequence $(\Delta_p(z_n^+, z_{n,k}))_{k \le k_n}$ with $z_n^+ := x_n + s_n^+$ is also monotonically decreasing. In fact, proceeding like in* (3.33) *and* (3.35) ,

$$\Delta_p\left(z_n^+, z_{n,k+1}\right) - \Delta_p\left(z_n^+, z_{n,k}\right) \le h\left(\lambda_{n,k}\right)$$

with

$$h\left(\lambda_{n,k}\right) := C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)} \lambda_{n,k}^{s^*} \|\nabla_{n,k}\|^{s^*} \vee \lambda_{n,k}^{p^*} \|\nabla_{n,k}\|^{p^*}\right) - \lambda_{n,k}\left\langle \nabla_{n,k}, s_{n,k} - s_n^+\right\rangle .$$

Now, as $\|A_n s_n^+ - b_n\| \le \|A_n e_n - b_n\|$, we obtain like in (3.34) *and* (3.28) ,

$$\left\langle \nabla_{n,k}, s_{n,k} - s_n^+\right\rangle \ge \left(1 - \frac{\eta}{\mu}\right) \|A_n s_{n,k} - b_n\|^r$$

From (3.36) , *it follows that for all $\lambda_{n,k} \in (0, \lambda_{DE}]$ it is true that*

$$h\left(\lambda_{n,k}\right) \le -\left(1 - C_0 - \frac{\eta}{\mu}\right) \lambda_{n,k} \|A_n s_{n,k} - b_n\|^r < 0.$$

Notice that

$$\lambda_{DE,s} \gtrsim \frac{\|A_n s_{n,k} - b_n\|^{r(s-1)}}{\|\nabla_{n,k}\|^s} \geq \frac{1}{M^s} \|A_n s_{n,k} - b_n\|^{s-r},$$

which implies that

$$\lambda_{DE} \gtrsim \|A_n s_{n,k} - b_n\|^t, \text{ with } t := s - r > -r. \tag{3.40}$$

Employing the TCC (Assumption 1(c), page 31), we see that

$$\|b_n\| - \|A_n e_n\| \leq \|b_n - A_n e_n\| \leq \eta \|b_n\|.$$

Hence

$$\|b_n\| \leq \frac{M}{1-\eta} \|e_n\| \leq \frac{M}{1-\eta} \left(\|x^+\| + C_{\rho,x^+}\right). \tag{3.41}$$

Thus, the sequence of residuals $(b_n)_{n\in\mathbb{N}}$ is bounded. Since

$$\|s_{n,k}\| \leq \|z_{n,k}\| + \|x_n\| \leq 2C_{\rho,x^+}, \tag{3.42}$$

the sequences $(s_{n,k})_{k\leq k_n}$ and consequently $(A_n s_{n,k} - b_n)_{k\leq k_n}$ are uniformly bounded for all $n \in \mathbb{N}$. This implies that there exists a constant $C_4 > 0$ independent on n and k such that $\|A_n s_{n,k} - b_n\| \leq C_4$ for all $k \leq k_n$ and $n \in \mathbb{N}$. Thus, as $\|A_n s_{n,k} - b_n\| \geq \mu \|b_n\|$ for all $k = 0, ..., k_n - 1$,

$$\lambda_{DE} \gtrsim \|A_n s_{n,k} - b_n\|^{s-r} \geq \frac{C_4^{-r}\mu^s}{M^s} \|b_n\|^s.$$

In particular, the Landweber method with step-size (independent of k) defined by $\lambda_{LW} := C_5 \|b_n\|^s$, where $C_5 > 0$ is a small constant, is well-defined as inner iteration of K-REGINN and satisfies $\lambda_{LW} \in (0, \lambda_{DE}]$. Consequently, inequality (3.37) is satisfied for this method.

Even a small constant in n and k can be used as Landweber step-size if $s \leq r$, because in this case

$$\lambda_{DE} \gtrsim \|A_n s_{n,k} - b_n\|^{s-r} \geq C_4^{s-r}. \tag{3.43}$$

From definition (3.29), we can see that

$$s_{n,k+1} = J_{p^*}^* \left(J_p \left(x_n + s_{n,k}\right) - \lambda_{n,k} \nabla_{n,k}\right) - x_n$$

and since the Steepest Descent method is the gradient method whose the step-size minimizes the residual, the most natural manner to define the Steepest Descent method for the dual gradient methods seems to be choosing a step-size satisfying

$$\lambda_{SD} \in \arg\min_{\lambda\in\mathbb{R}^+} \varphi_n \left(J_{p^*}^* \left(J_p \left(x_n + s_{n,k}\right) - \lambda \nabla_{n,k}\right) - x_n\right). \tag{3.44}$$

If such a number exists, it follows immediately that

$$\varphi_n \left(s_{n,k+1}\right) \leq \varphi \left(J_{p^*}^* \left(J_p \left(x_n + s_{n,k}\right) - \lambda \nabla_{n,k}\right) - x_n\right)$$

for all $\lambda \geq 0$, and in particular, $\|A_n s_{n,k+1} - b_n\| \leq \|A_n s_{n,k} - b_n\|$ is achieved by picking $\lambda = 0$. However, an explicit expression for λ_{SD} is hard to find. Even simple inequalities involving λ_{SD} are not easy to be obtained. Unfortunately we were not able to prove that $\lambda_{SD} \in (0, \lambda_{DE}]$ to include it in our convergence analysis of Chapter 4. But, the Modified Steepest Descent method defined in (3.18) can be useful here, because it already has an explicit step-size. For the dual methods it is nevertheless convenient to alter its exponents. Notice that for $\ell \in \{p, s\}$,

$$\begin{aligned}\|\nabla_{n,k}\|^{p^* r^*(\ell-1)} &= \left\langle J_{p^*}^* \left(\nabla_{n,k}\right), \nabla_{n,k}\right\rangle^{r^*(\ell-1)} = \left\langle A_n J_{p^*}^* \left(\nabla_{n,k}\right), j_r \left(A_n s_{n,k} - b_n\right)\right\rangle^{r^*(\ell-1)} \\ &\leq \left\|A_n J_{p^*}^* \left(\nabla_{n,k}\right)\right\|^{r^*(\ell-1)} \|A_n s_{n,k} - b_n\|^{r(\ell-1)},\end{aligned}$$

and therefore,

$$\frac{\|\nabla_{n,k}\|^{p^* r^*(\ell-1)-\ell}}{\left\|A_n J_{p^*}^* \left(\nabla_{n,k}\right)\right\|^{r^*(\ell-1)}} \leq \frac{\|A_n s_{n,k} - b_n\|^{r(\ell-1)}}{\|\nabla_{n,k}\|^{\ell}}.$$

Thus, defining[8]

$$\lambda_{MSD} := K_1 \lambda_{MSD,s} \wedge K_2 \lambda_{MSD,p} \tag{3.45}$$

with

$$\lambda_{MSD,\ell} := \frac{\|\nabla_{n,k}\|^{p^* r^*(\ell-1)-\ell}}{\left\|A_n J_{p^*}^* \left(\nabla_{n,k}\right)\right\|^{r^*(\ell-1)}},$$

$0 < K_1 \leq C_1$ and $0 < K_2 \leq C_2$, we conclude that $\lambda_{MSD} \leq \lambda_{DE}$ and inequality (3.37) automatically holds for this method. We remark that this step-size still coincide with the one of Steepest Descent method (3.19) in Hilbert spaces if $p = s = r = 2$ and $K_1 = K_2 = 1$.[9]

Further

$$\lambda_{MSD,\ell} \geq \frac{\|\nabla_{n,k}\|^{p^* r^*(\ell-1)-\ell}}{M^{r^*(\ell-1)} \|\nabla_{n,k}\|^{(p^*-1) r^*(\ell-1)}} \gtrsim \|\nabla_{n,k}\|^{t},$$

with $t := r^*(\ell - 1) - \ell \geq -1 > -r$. Now, if $\ell \leq r$ (i.e., if $p \leq s \leq r$), we conclude that $t \leq 0$, which implies that

$$\lambda_{MSD} \gtrsim \|A_n s_{n,k} - b_n\|^{t}, \text{ with } -r < t \leq 0. \tag{3.46}$$

Using the same argument as that used for the DE method (see (3.43)), we conclude that

$$\lambda_{MSD} \gtrsim C_4^t, \tag{3.47}$$

where $C_4 > 0$ is an uniform upper bound to $\|A_n s_{n,k} - b_n\|$.

We investigate now what should be the optimal step-size for the dual gradient methods. Defining the vector $z_{n,\lambda} \in X$ as

$$J_p\left(z_{n,\lambda}\right) := J_p\left(z_{n,k}\right) - \lambda \nabla_{n,k},$$

we intend to find a step-size $\lambda = \lambda_{opt}$ such that the corresponding dual gradient method minimizes $f(\lambda) := \Delta_p\left(x^+, z_{n,\lambda}\right) - \Delta_p\left(x^+, z_{n,k}\right)$, yielding a dual gradient method with the fastest decreasing error in the inner iteration of `K-REGINN`. Looking to the three points identity (2.21), we see that this difference can be written as

$$\begin{aligned}
f(\lambda) &= \Delta_p\left(z_{n,k}, z_{n,\lambda}\right) + \left\langle J_p\left(z_{n,\lambda}\right) - J_p\left(z_{n,k}\right), z_{n,k} - x^+ \right\rangle \\
&= \left(\frac{1}{p}\|z_{n,k}\|^p - \left\langle J_p\left(z_{n,\lambda}\right), z_{n,k}\right\rangle + \frac{1}{p^*}\left\|J_p\left(z_{n,\lambda}\right)\right\|^{p^*}\right) - \lambda\left\langle \nabla_{n,k}, z_{n,k} - x^+\right\rangle \\
&= \left(\frac{1}{p}\|z_{n,k}\|^p - \left\langle J_p\left(z_{n,k}\right) - \lambda\nabla_{n,k}, z_{n,k}\right\rangle + \frac{1}{p^*}\left\|J_p\left(z_{n,k}\right) - \lambda\nabla_{n,k}\right\|^{p^*}\right) \\
&\quad - \lambda\left\langle \nabla_{n,k}, z_{n,k} - x^+\right\rangle.
\end{aligned}$$

The uniformly smoothness of X^* implies the Fréchet-differentiability of $\frac{1}{p^*}\|\cdot\|^{p^*}$. Applying the chain rule, we obtain

$$\begin{aligned}
f'(\lambda) &= \left\langle \nabla_{n,k}, z_{n,k}\right\rangle - \left\langle J_{p^*}\left(J_p\left(z_{n,k}\right) - \lambda\nabla_{n,k}\right), \nabla_{n,k}\right\rangle - \left\langle \nabla_{n,k}, z_{n,k} - x^+\right\rangle \\
&= -\left\langle \nabla_{n,k}, J_{p^*}^*\left(J_p\left(z_{n,k}\right) - \lambda\nabla_{n,k}\right) - x^+\right\rangle,
\end{aligned}$$

[8] The step-size of the MSD method is defined slightly differently for the primal and dual gradient methods. The same occurs with the LW method. We however keep the same notation λ_{MSD} and λ_{LW} for these step-sizes in both situations.

[9] In Hilbert spaces, $C_{p^*,s^*} = 1/2$ can be chosen and therefore, the unique restriction we have is $1 = K_\ell \leq C_\ell = C_0/C_{p^*,s^*} = 2C_0$, $\ell = 1, 2$. This means that the inequality $1/2 \leq C_0 < 1$ needs to be observed.

which makes evident that f' is continuous. From (3.28) and (3.34),

$$f'(0) = -\left\langle \nabla_{n,k}, z_{n,k} - x^+ \right\rangle \leq -\left(1 - \frac{\eta}{\mu}\right) \|A_n s_{n,k} - b_n\|^r < 0$$

in the inner iteration of K-REGINN. Further,

$$\|J_p(z_{n,k}) - \lambda \nabla_{n,k}\| \geq \lambda \|\nabla_{n,k}\| - \|J_p(z_{n,k})\| \to \infty \text{ as } \lambda \to \infty,$$

which in view of (2.20) implies that $\Delta_p(x^+, z_{n,\lambda}) = \Delta_{p^*}(J_p(z_{n,\lambda}), J_p(x^+))$ cannot remain unbounded as λ grows to infinite. It follows that $f(\lambda) \to \infty$ as $\lambda \to \infty$ and since f is continuous, it needs to be increasing at least in a small open interval. The derivative of f is therefore positive in this interval and since it is a continuous function and $f'(0) < 0$, there exists a $\lambda = \lambda_{opt} > 0$ such that $f'(\lambda_{opt}) = 0$. The uniqueness of λ_{opt} follows from the injectivity of f' : let $\lambda_1, \lambda_2 \in \mathbb{R}$ satisfy $f'(\lambda_1) = f'(\lambda_2)$. Assuming that $\lambda_1 \neq \lambda_2$ we derive from the strictly monotonicity of the duality mapping in strictly convex Banach spaces the contradiction

$$\begin{aligned}
0 &= (\lambda_1 - \lambda_2)\left(f'(\lambda_1) - f'(\lambda_2)\right) \\
&= (\lambda_1 - \lambda_2)\left\langle \nabla_{n,k}, J_{p^*}^*(J_p(z_{n,\lambda_2})) - J_{p^*}^*(J_p(z_{n,\lambda_1}))\right\rangle \\
&= \left\langle (J_p(z_{n,k}) - \lambda_2 \nabla_{n,k}) - (J_p(z_{n,k}) - \lambda_1 \nabla_{n,k}), J_{p^*}^*(J_p(z_{n,\lambda_2})) - J_{p^*}^*(J_p(z_{n,\lambda_1}))\right\rangle \\
&= \left\langle J_p(z_{n,\lambda_2}) - J_p(z_{n,\lambda_1}), J_{p^*}^*(J_p(z_{n,\lambda_2})) - J_{p^*}^*(J_p(z_{n,\lambda_1}))\right\rangle > 0.
\end{aligned}$$

Of course λ_{opt} is a minimizer of f because f' is negative in zero and positive for large values. Additionally, the unique minimizer λ_{opt} solves the nonlinear equation

$$\left\langle \nabla_{n,k}, J_{p^*}^*(J_p(z_{n,k}) - \lambda_{opt} \nabla_{n,k}) - x^+ \right\rangle = 0. \tag{3.48}$$

Using again the three points identity, we conclude that for $z_{n,k+1} := z_{n,\lambda_{opt}}$,

$$\begin{aligned}
\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, z_{n,k}\right) &= -\Delta_p(z_{n,k+1}, z_{n,k}) + \left\langle J_p(z_{n,k+1}) - J_p(z_{n,k}), z_{n,k+1} - x^+\right\rangle \\
&= -\Delta_p(z_{n,k+1}, z_{n,k}) - \lambda_{opt}\left\langle \nabla_{n,k}, z_{n,k+1} - x^+\right\rangle
\end{aligned}$$

and consequently

$$\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, z_{n,k}\right) = -\Delta_p(z_{n,k+1}, z_{n,k}).$$

An explicit formula for λ_{opt} is hard to achieve and to calculate λ_{opt}, even numerically, is very challenging since the nonlinear equation (3.48) depends on the unavailable information x^+. It is easy to confirm that $\lambda_{opt} = \langle \nabla_{n,k}, s_{n,k} - e_n \rangle / \|\nabla_{n,k}\|^2$ in Hilbert spaces, which still depends on the unavailable information x^+. This leaves little hope to exactly determine λ_{opt} for nonlinear problems using only available information[10]. However, using the property (3.48) a lower bound for λ_{opt} can be achieved:

$$\begin{aligned}
\lambda_{opt}\left\langle \nabla_{n,k}, z_{n,k} - x^+\right\rangle &= -\langle \lambda_{opt} \nabla_{n,k}, z_{n,k+1} - z_{n,k}\rangle \\
&= \langle J_p(z_{n,k+1}) - J_p(z_{n,k}), z_{n,k+1} - z_{n,k}\rangle \\
&= \Delta_p(z_{n,k+1}, z_{n,k}) + \Delta_p(z_{n,k}, z_{n,k+1}) \\
&= \Delta_{p^*}(J_p(z_{n,k}), J_p(z_{n,k+1})) + \Delta_{p^*}(J_p(z_{n,k+1}), J_p(z_{n,k})).
\end{aligned}$$

The inequality (3.31) obviously hold for $k = 0, ..., k_n - 1$, and accordingly, the uniform bound (3.32) is true. Proceeding like in (3.35), we obtain

$$\begin{aligned}
\lambda_{opt}\langle \nabla_{n,k}, s_{n,k} - e_n\rangle &= \lambda_{opt}\left\langle \nabla_{n,k}, z_{n,k} - x^+\right\rangle \\
&\leq 2C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)} \lambda_{opt}^{s^*} \|\nabla_{n,k}\|^{s^*} \vee \lambda_{opt}^{p^*} \|\nabla_{n,k}\|^{p^*}\right).
\end{aligned}$$

[10] In Hilbert spaces, the last expression reduces to the calculable step-size $\lambda_{opt} = \|As_k - b\|^2 / \|\nabla_k\|^2$ if the problem (3.1) is linear, see (3.25).

Thus

$$\lambda_{opt} \geq \left(\frac{1}{2^{s-1} C_{p^*,s^*}^{s-1} C_{\rho,x^+}^{s-p}} \right) \frac{\langle \nabla_{n,k}, s_{n,k} - e_n \rangle^{s-1}}{\|\nabla_{n,k}\|^s} \wedge \left(\frac{1}{2^{p-1} C_{p^*,s^*}^{p-1}} \right) \frac{\langle \nabla_{n,k}, s_{n,k} - e_n \rangle^{p-1}}{\|\nabla_{n,k}\|^p}.$$

Finally, from (3.34) and (3.28) we obtain, like in Remark 26,

$$\lambda_{opt} \geq \left(\frac{1-\eta/\mu}{2} \right)^{s-1} C_1 \lambda_{DE.s} \wedge \left(\frac{1-\eta/\mu}{2} \right)^{p-1} C_2 \lambda_{DE,p} \geq \left(\frac{1-\eta/\mu}{2} \right)^{s-1} \lambda_{DE}.$$

Remark 28 *Unfortunately, we do not know if $\lambda_{opt} \leq \lambda_{DE}$ to include this method in the convergence analysis of Chapter 4. However, as λ_{opt} minimizes f, we can conclude that*

$$\Delta_p \left(x^+, z_{n,\lambda_{opt}} \right) - \Delta_p \left(x^+, z_{n,k} \right) = f\left(\lambda_{opt} \right) \leq f\left(\lambda_{DE} \right) \overset{(3.38)}{\leq} -C_3 \lambda_{DE} \left\| A_n s_{n,k} - b_n \right\|^r,$$

which is enough to prove that the inner iteration of `K-REGINN` *terminates ($k_n < \infty$) and consequently the monotonicity property* (3.31) *as well as the bound* (3.32) *is transferred to the outer iteration, see Theorem 38. This implies that at least weak convergence holds for this method, see Corollary 41.*

We finish this subsection discussing how the dual gradient methods look like when $k_{\max} = 1$ is used (see (3.12)). In this case, `K-REGINN` performs only one inner iteration each outer iteration whenever the inequality (3.9) is not verified. It follows that

$$\begin{aligned} J_p\left(x_{n+1}\right) &= J_p\left(z_{n,1}\right) = J_p\left(z_{n,0}\right) - \lambda_{n,0} A_n^* j_r\left(A_n s_{n,0} - b_n\right) \\ &= J_p\left(x_n\right) - \lambda_{n,0} F_{[n]}'\left(x_n\right)^* j_r\left(F_{[n]}\left(x_n\right) - y_{[n]}\right). \end{aligned}$$

We conclude that this procedure is equivalent to apply a dual gradient method directly to the nonlinear system (3.6). Thus the convergence of `K-REGINN` using these methods as inner iteration implies in particular the convergence of a Kaczmarz version of the dual gradient methods themselves. Analogously, the iteration

$$x_{n+1} = x_n - \lambda_{n,0} j_{p^*}^* \left(F_{[n]}'\left(x_n\right)^* J_r\left(F_{[n]}\left(x_n\right) - y_{[n]}\right)\right)$$

represents the primal gradient methods in case $k_{\max} = 1$.

3.1.3 Tikhonov methods

The linear system $A_n s = b_n$ often inherits the ill-posedness of the nonlinear equation $F_{[n]}\left(x_n\right) = y_{[n]}$. Therefore, it is crucial to apply a regularization technique to reconstruct stable approximations. The plan in this subsection is to employ Tikhonov regularization [51] to the referred linear system in order to generate the inner iteration of `K-REGINN`. To alleviate the notation, we forget `K-REGINN` once again and skip the index n of the outer iteration for a while. Let X and Y be Banach spaces, $A \colon X \to Y$ linear and $b \in Y$ fixed. Define for each $k \in \mathbb{N}_0$, the *Tikhonov functional* $T_k \colon X \to \mathbb{R}_0^+$ as

$$T_k\left(s\right) := \frac{1}{r} \left\| As - b \right\|^r + \alpha_k \Omega\left(s\right), \tag{3.49}$$

with $r > 1$, $\alpha_k > 0$ and $\Omega \colon X \to \mathbb{R}_0^+$ being subdifferentiable and satisfying $\left(\overline{s}_m\right)_{m\in\mathbb{N}}$ bounded whenever $\left(\Omega\left(\overline{s}_m\right)\right)_{m\in\mathbb{N}}$ bounded. Define now the *Tikhonov iteration* as

$$s_{k+1} \in \arg\min_{s\in X} T_k\left(s\right), \; s_0 = 0. \tag{3.50}$$

Algorithm 4 `K-REGINN` with Tikhonov inner iteration

Input: x_N; (y^{δ_j}, δ_j); F_j; F_j', $j = 0, \ldots, d-1$; μ; τ; $r > 1$;

Output: x_N with $\|y_j^{\delta_j} - F_j(x_N)\| \le \tau\delta_j$, $j = 0, \ldots, d-1$;

$\ell := 0$; $x_0 := x_N$; $c := 0$;
while $c < d$ **do**
  for $j = 0 : d-1$ **do**
    $n := \ell d + j$;
    $b_n^\delta := y_j^{\delta_j} - F_j(x_n)$; $A_n := F_j'(x_n)$;
    if $\|b_n^\delta\| \le \tau\delta_j$ **then**
      $x_{n+1} := x_n$; $c := c+1$;
    else
      $k := 0$; $s_{n,0} := 0$; choose $k_{\max,n} \in \mathbb{N}$;
      repeat
        choose $\alpha_{n,k} > 0$ and $\Omega_k \colon X \to \mathbb{R}_0^+$ subdifferentiable;
        choose $s_{n,k+1} \in \arg\min_{s\in X} \left(\frac{1}{r} \|A_n s - b_n^\delta\|^r + \alpha_{n,k}\Omega_k(s) \right)$;
        $k := k+1$;
      until $\left\|b_n^\delta - A_n s_{n,k}\right\| < \mu \left\|b_n^\delta\right\|$ or $k = k_{\max,n}$
      $x_{n+1} := x_n + s_{n,k}$; $c := 0$;
    end if
  end for
  $\ell := \ell + 1$;
end while
$x_N := x_{\ell d - c}$;

Algorithm 4 displays the combination of `K-REGINN` with a Tikhonov method in the inner iteration with the parts corresponding to the Tikhonov methods highlighted in red. To guarantee the well-definedness of the inner iteration, the space X is assumed to be reflexive, which is a sufficient condition for the existence of a minimizer of T_k in (3.49) as we will see in a moment.

Remark 29 *The* PENALIZATION TERM Ω *is defined above in a relatively general form. Therefore, the Tikhonov functional* (3.49) *can represent many different functionals and the associated iteration* (3.50) *accordingly plays the role of various Tikhonov-based iterations. In this subsection we analyze four different methods, namely, the* TIKHONOV-PHILLIPS *and* ITERATED-TIKHONOV *methods, where* $\Omega(s) := \frac{1}{p}\|s - \overline{x}\|^p$ *for some fixed vector* $\overline{x} \in X$ *and* $\Omega(s) := \frac{1}{p}\|s - s_k\|^p$ *respectively, as well as their* BREGMAN VARIATIONS, $\Omega(s) := \Delta_p(x + s, x + \overline{x})$ *for some* $x \in X$ *fixed and* $\Omega(s) = \Delta_p(x + s, x + s_k)$ *respectively. Although our main intention is to prove some properties of these methods, we keep the analysis as general as possible and only introduce the referred methods at the specific points we need them.*

The existence of a minimizer of T_k is assured in case of X being reflexive as we prove now. Indeed, following ideas from [48, Prop. 4.1], let $(\overline{s}_m)_{m\in\mathbb{N}} \subset X$ be a sequence satisfying $T_k(\overline{s}_m) \to \inf\{T_k(s) : s \in X\}$ as $m \to \infty$. Thus the sequence $(T_k(\overline{s}_m))_{m\in\mathbb{N}}$, and consequently $\Omega(\overline{s}_m) \le T_k(\overline{s}_m)/\alpha_k$ is bounded. It follows that $(\overline{s}_m)_{m\in\mathbb{N}}$ is bounded and since X is reflexive, there exists a subsequence $\left(\overline{s}_{m_j}\right)_{j\in\mathbb{N}}$ and a vector $s^+ \in X$ such that

$\overline{s}_{m_j} \rightharpoonup s^+$ as $j \to \infty$. Hence $A\overline{s}_{m_j} - b \rightharpoonup As^+ - b$ and since the norm-function is lower semi-continuous,

$$\left\|As^+ - b\right\| \leq \liminf_{j\to\infty} \left\|A\overline{s}_{m_j} - b\right\|.$$

As Ω is subdifferentiable, it is also lower semi-continuous. Therefore

$$\Omega\left(s^+\right) \leq \liminf_{j\to\infty} \Omega\left(\overline{s}_{m_j}\right).$$

Both results together prove that

$$T_k\left(s^+\right) \leq \liminf_{j\to\infty} T_k\left(\overline{s}_{m_j}\right) = \lim_{m\to\infty} T_k\left(\overline{s}_m\right) = \inf_{s\in X} T_k\left(s\right).$$

Therefore s^+ is a minimizer of T_k as we wanted.

Utilizing the optimality condition $0 \in \partial T_k\left(s_{k+1}\right)$ we find

$$0 \in A^* J_r\left(As_{k+1} - b\right) + \alpha_k \partial\Omega\left(s_{k+1}\right),$$

and conclude that there exist a selection $j_r\colon Y \to Y^*$ and $s^*_{k+1} \in \partial\Omega\left(s_{k+1}\right)$ such that

$$s^*_{k+1} = -\frac{1}{\alpha_k} A^* j_r\left(As_{k+1} - b\right).$$

If the functional Ω is strictly convex, the minimizer of T_k is unique and consequently, s_{k+1} is unique in (3.50). If Ω is Gâteaux-differentiable in s_{k+1}, the above equality becomes

$$\Omega'\left(s_{k+1}\right) = -\frac{1}{\alpha_k} A^* j_r\left(As_{k+1} - b\right),$$

where Ω' represents the G-derivative of Ω.

The *Tikhonov-Phillips* method (TP) is defined by choosing a strictly decreasing zero-sequence $\left(\alpha_k\right)_{k\in\mathbb{N}}$ and $\Omega\left(s\right) := \frac{1}{p}\left\|s - \overline{x}\right\|^p$, with $p > 1$ and $\overline{x} \in X$ being independent on k.

Assume from now on that X is smooth and strictly convex. The first restriction is equivalent to the G-differentiability of $\frac{1}{p}\left\|\cdot\right\|^p$ and consequently $\Omega'\left(s\right) = J_p\left(s - \overline{x}\right)$. The second restriction in turn, is equivalent to the strict convexity of $\frac{1}{p}\left\|\cdot\right\|^p$, which implies that T_k is strictly convex and therefore the minimizer of this functional is unique. Choosing $\overline{x} := 0$, the TP method results in the implicit iteration

$$J_p\left(s_{k+1}\right) = -\frac{1}{\alpha_k} A^* j_r\left(As_{k+1} - b\right). \tag{3.51}$$

It is easy to confirm that $s_{k+1} = \left(A^*A + \alpha_k I\right)^{-1} A^*b$ in Hilbert spaces (using $p = r = 2$).

Definition (3.50) immediately implies that $T_{k-1}\left(s_k\right) \leq T_{k-1}\left(0\right)$ for all $k \in \mathbb{N}$ and as $\Omega\left(0\right) = 0$,

$$\left\|As_k - b\right\| \leq \left\|b\right\|. \tag{3.52}$$

Used as inner iteration of `K-REGINN`, the iteration (3.51) results in a Kaczmarz variation of the well-known Levenberg-Marquardt method [21]. `K-REGINN` with the use of TP method in the inner iteration and the choice $\overline{x} := x_0 - x_n$ is transformed into a Kaczmarz version of the `IRGN` (`I`teratively `R`egularized `G`auss-`N`ewton method, due to Bakushinskii [3], see also [29]). Observe that, if $\overline{x} \neq 0$, then $\Omega\left(0\right) = \frac{1}{p}\left\|\overline{x}\right\|^p \neq 0$, which means that the inequality $T_{k-1}\left(s_k\right) \leq T_{k-1}\left(0\right)$ alone is not enough to ensure (3.52). However, inequality (3.16) can be proven for this method, as Lemma 30 below shows.

A variation of TP method is defined employing the Bregman distance instead of the standard norm in X. The functional $\frac{1}{p}\left\|\cdot - \overline{x}\right\|^p$ is then replaced by $\Omega\left(s\right) := \Delta_p\left(x + s, x + \overline{x}\right)$

where $x \in X$ is a fixed vector. The restrictions on X still guarantee the strict convexity and G-differentiability of Ω in this case. Moreover, $\Omega'(s) = J_p(x+s) - J_p(x+\overline{x})$ and we have for this variation,

$$J_p(x+s_{k+1}) = J_p(x+\overline{x}) - \frac{1}{\alpha_k} A^* j_r(As_{k+1} - b). \tag{3.53}$$

The two variations of TP method coincide for $p = 2$ whenever X is a Hilbert space, because in this situation $\frac{1}{2}\|s - \overline{x}\|^2 = \Delta_2(x+s, x+\overline{x})$. Similar to before, $\overline{x} = 0$ implies $\Omega(0) = 0$ and the property (3.52) follows again from $T_{k-1}(s_k) \leq T_{k-1}(0)$. Inequality (3.16) also holds for this method, even in case of $\overline{x} \neq 0$, as the next lemma demonstrate.

The first part of this lemma is a generalization to Banach spaces of [30, Theo. 2.16].

Lemma 30 *Let X and Y be Banach spaces with X being reflexive, $A\colon X \to Y$ linear, $b \in Y$ and $(\alpha_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ a strictly decreasing positive sequence with $\alpha_k \to 0$ as $k \to \infty$. Let $s_0 \in X$ and*

$$s_{k+1} \in \arg\min_{s\in X} T_k(s)$$

with

$$T_k(s) := \frac{1}{r}\|As - b\|^r + \alpha_k \Omega(s),$$

where $r > 1$, $\Omega\colon X \to \mathbb{R}_0^+$ is subdifferentiable and satisfies $(\overline{s}_m)_{m\in\mathbb{N}}$ bounded whenever $(\Omega(\overline{s}_m))_{m\in\mathbb{N}}$ bounded. Then $(s_k)_{k\in\mathbb{N}}$ is well-defined and satisfies

$$\|As_{k+1} - b\| \leq \|As_k - b\| \ \text{ and } \Omega(s_{k+1}) \geq \Omega(s_k) \tag{3.54}$$

for all $k \in \mathbb{N}$, where the above inequalities are strict in case of Y being strictly convex. Furthermore,

$$\lim_{k\to\infty} \|As_k - b\| = \inf_{s\in X} \|As - b\| \tag{3.55}$$

and consequently, inequality (3.23) *in Proposition 24 holds for $C = 0$.*

Proof. As shown at the beginning of this subsection, a minimizer of T_k exists for each $k \in \mathbb{N}_0$. Thus s_k and s_{k+1} are well-defined and fill the optimality conditions $0 \in \partial T_{k-1}(s_k)$ and $0 \in \partial T_k(s_{k+1})$. Therefore, there exist a selection j_r of the duality mapping J_r, $s_k^* \in \partial\Omega(s_k)$ and $s_{k+1}^* \in \partial\Omega(s_{k+1})$ such that

$$\begin{aligned} \alpha_{k-1} s_k^* &= -A^* j_r(As_k - b) \ \text{ and} \\ \alpha_k s_{k+1}^* &= -A^* j_r(As_{k+1} - b). \end{aligned} \tag{3.56}$$

Subtracting the second equation from the first and applying both sides of the resulting equality in the vector $s_k - s_{k+1}$,

$$\begin{aligned} &\alpha_{k-1}\langle s_k^* - s_{k+1}^*, s_k - s_{k+1}\rangle + (\alpha_{k-1} - \alpha_k)\langle s_{k+1}^*, s_k - s_{k+1}\rangle \\ &\quad = \langle j_r(As_{k+1} - b) - j_r(As_k - b), A(s_k - s_{k+1})\rangle. \end{aligned}$$

Observe now that $A(s_k - s_{k+1}) = -[(As_{k+1} - b) - (As_k - b)]$. From the monotonicity of the duality mapping (respectively the strict monotonicity in case of Y being a strictly convex space), we conclude that the first term in the left-hand side and the one in the right-hand side of above equality are non-negative (resp. positive) and non-positive (resp. negative) respectively. As $\alpha_{k-1} > \alpha_k$, we conclude that $\langle s_{k+1}^*, s_k - s_{k+1}\rangle \leq 0$ (resp. $\langle s_{k+1}^*, s_k - s_{k+1}\rangle < 0$). We proceed applying both sides of second equation of (3.56) in the vector $s_k - s_{k+1}$,

$$\begin{aligned} 0 \geq \alpha_k\langle s_{k+1}^*, s_k - s_{k+1}\rangle &= \langle j_r(As_{k+1} - b), A(s_{k+1} - s_k)\rangle \\ &= \langle j_r(As_{k+1} - b), As_{k+1} - b\rangle - \langle j_r(As_{k+1} - b), As_k - b\rangle \\ &\geq \|As_{k+1} - b\|^r - \|As_{k+1} - b\|^{r-1}\|As_k - b\|, \end{aligned}$$

which proves the first inequality in (3.54). To prove the second one, we only use $T_{k-1}(s_k) \leq T_{k-1}(s_{k+1})$ and apply the just proved inequality $\|As_{k+1} - b\| \leq \|As_k - b\|$.

Now we turn to (3.55). Because $(\|As_k - b\|)_{k\in\mathbb{N}}$ is a non-increasing sequence, it converges. Let $\epsilon > 0$ be given and $(\overline{s}_j)_{j\in\mathbb{N}} \subset X$ be a sequence satisfying

$$\frac{1}{r}\|A\overline{s}_j - b\|^r \to a := \inf_{s\in X}\frac{1}{r}\|As - b\|^r \text{ as } j \to \infty.$$

Then there exists a number $J \in \mathbb{N}$ such that for all $j \geq J$ and all $k \in \mathbb{N}$,

$$\begin{aligned} a &\leq \frac{1}{r}\|As_k - b\|^r \leq \frac{1}{r}\|As_k - b\|^r + \alpha_{k-1}\Omega(s_k) \\ &\leq \frac{1}{r}\|A\overline{s}_j - b\|^r + \alpha_{k-1}\Omega(\overline{s}_j) \leq a + \epsilon + \alpha_{k-1}\Omega(\overline{s}_j). \end{aligned}$$

In particular

$$a \leq \frac{1}{r}\|As_k - b\|^r \leq a + \epsilon + \alpha_{k-1}\Omega(\overline{s}_J)$$

for all $k \in \mathbb{N}$. Since $\alpha_k \to 0$ as $k \to \infty$,

$$a \leq \lim_{k\to\infty}\frac{1}{r}\|As_k - b\|^r \leq a + \epsilon.$$

As $\epsilon > 0$ is arbitrary, (3.55) follows. Inequality (3.23) with $C = 0$ is just (3.55). ■

We now impose some restrictions on the sequence $(\alpha_k)_{k\in\mathbb{N}}$ in order to prove an inequality similar to (3.67) for the Bregman variation of TP method (3.53). Choose $1 < \theta < (r+1)/r$ and assume that

$$1 < \frac{\alpha_{k-1}}{\alpha_k} \leq \theta, \text{ for all } k \in \mathbb{N}. \tag{3.57}$$

Define $z_k := x + s_k$ and observe that the inequality $T_{k-1}(s_k) \leq T_{k-1}(s_{k+1})$ implies

$$\begin{aligned} -\|As_{k+1} - b\|^r &\leq -\|As_k - b\|^r + r\alpha_{k-1}\Delta_p(z_{k+1}, x + \overline{x}) \\ &\leq -\|As_k - b\|^r + r\theta\alpha_k\Delta_p(z_{k+1}, x + \overline{x}), \end{aligned}$$

Now, as $T_k(s_{k+1}) \leq T_k(\overline{x})$,

$$\alpha_k\Delta_p(z_{k+1}, x + \overline{x}) \leq \frac{1}{r}\|A\overline{x} - b\|^r$$

and with help of (2.21) and (3.54), we obtain for $f(k) := \Delta_p(x^+, z_k) - \Delta_p(x^+, x + \overline{x})$ with $e := x^+ - x$,

$$\begin{aligned} f(k+1) &= \left\langle J_p(z_{k+1}) - J_p(x + \overline{x}), z_{k+1} - x^+\right\rangle - \Delta_p(z_{k+1}, x + \overline{x}) \\ &= -\frac{1}{\alpha_k}\langle j_r(As_{k+1} - b), A(s_{k+1} - e)\rangle - \Delta_p(z_{k+1}, x + \overline{x}) \\ &\leq \frac{1}{\alpha_k}\left(\|As_{k+1} - b\|^{r-1}\|Ae - b\| - \|As_{k+1} - b\|^r - \alpha_k\Delta_p(z_{k+1}, x + \overline{x})\right) \\ &\leq \frac{1}{\alpha_k}\left(\|As_{k+1} - b\|^{r-1}\|Ae - b\| - \|As_k - b\|^r + (r\theta - 1)\alpha_k\Delta_p(z_{k+1}, x + \overline{x})\right) \\ &\leq \frac{1}{\alpha_k}\left(\|As_k - b\|^{r-1}\|Ae - b\| - \|As_k - b\|^r + C_2\|A\overline{x} - b\|^r\right), \end{aligned}$$

with $C_2 := (r\theta - 1)/r < 1$. For $\overline{x} := 0$, $x := x_n$, $s_k := s_{n,k}$, $A := A_n$ and $b := b_n$, the TP method becomes

$$J_p(z_{n,k+1}) = J_p(x_n) - \frac{1}{\alpha_k}A_n^* j_r(A_n s_{n,k+1} - b_n) \tag{3.58}$$

and the above inequality takes the form

$$\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, x_n\right) \tag{3.59}$$
$$\leq \frac{1}{\alpha_k}\left[\|A_n s_{n,k} - b_n\|^{r-1}\left(\|A_n e_n - b_n\| - \|A_n s_{n,k} - b_n\|\right) + C_2\|b_n\|^r\right].$$

Moreover, the vector $z_{n,k+1}$ in (3.58) is the unique minimizer of the Tikhonov functional

$$T_{n,k}(z) := \frac{1}{r}\|A_n(z - x_n) - b_n\|^r + \alpha_k \Delta_p(z, x_n). \tag{3.60}$$

Remark 31 *Inequality* (3.59) *was achieved using only the conditions for existence and uniqueness of a minimizer of* (3.49) *together with the hypotheses of Lemma 30. These are actually very few requirements. The $s-$convexity of X, for instance, is not necessary either to the well-definedness of the TP method nor to prove* (3.59) *above. The TCC (Assumption 1(c), page 31) is not necessary as well.*

The main difficulties in studying the convergence of K-REGINN using the Tikhonov-Phillips-type methods as inner iteration arise from the fact that the iteration s_{k+1} is actually independent of the previous iteration s_k. This fact also indicates that some information is wasted during the inner iteration. It would be helpful if we could count on an iteration similar to (3.29) of the dual gradient methods, where the available information s_k is used to generate the update s_{k+1}. Motivated by these facts, we introduce the *Iterated-Tikhonov* method (IT), where the sequence $(\alpha_k)_{k\in\mathbb{N}}$ is chosen being independent of k , i.e., $\alpha_k = \alpha$ for all $k \in \mathbb{N}$. In contrast, the functional Ω is dependent on this variable: $\Omega(s) = \Omega_k(s) := \frac{1}{p}\|s - s_k\|^p$. Then from (3.50),

$$J_p(s_{k+1} - s_k) = -\frac{1}{\alpha}A^* j_r(A s_{k+1} - b),$$

for some selection $j_r\colon Y \to Y^*$. Adopting the Bregman distance variation

$$\Omega_k(s) := \Delta_p(x + s, x + s_k),$$

with $x \in X$ fixed, we arrive at

$$J_p(x + s_{k+1}) = J_p(x + s_k) - \frac{1}{\alpha}A^* j_r(A s_{k+1} - b). \tag{3.61}$$

The inequality $\|A s_{k+1} - b\| \leq \|A s_k - b\|$ is an immediate consequence of $T_k(s_{k+1}) \leq T_k(s_k)$. Moreover, $T_k(s_{k+1}) < T_k(s_k)$ for $s_{k+1} \neq s_k$ because in this case $\Omega_k(s_{k+1}) > 0$ and accordingly (3.16) applies. Since $s_0 = 0$, inequality (3.52) is also true.

Our target now is to prove a similar inequality to (3.37) for the Bregman variation of IT method. Similarly to what occurs with the dual gradient methods, this property needs to be proven by induction at the same time with (3.31) and (3.32). For this reason, it is necessary to observe the inner and outer iteration of K-REGINN simultaneously. Regarded as inner iteration of K-REGINN, IT method has a constant regularization parameter α in the inner iteration, but it can be chosen dependent on the index n of the outer iteration: $\alpha = \alpha_n$. Let Assumption 1, page 31, hold true and assume that X is uniformly smooth and $s-$convex with $p \leq s \leq r$. Define now $x = x_n$, $s_k = s_{n,k}$, $A = A_n$, $b = b_n$ and $\alpha = \alpha_n$ in (3.61), which results in

$$J_p(z_{n,k+1}) = J_p(z_{n,k}) - \frac{1}{\alpha_n}A_n^* j_r(A_n s_{n,k+1} - b_n), \tag{3.62}$$

with $z_{n,k} := x_n + s_{n,k}$. With these definitions, the vector $z_{n,k+1} \in X$ is the unique minimizer of the Tikhonov functional

$$T_{n,k}(z) := \frac{1}{r}\|A_n(z - x_n) - b_n\|^r + \alpha_n \Delta_p(z, z_{n,k}). \tag{3.63}$$

In order to prove (3.37), and based in the proof of this inequality for the dual gradient method DE presented in the last subsection, we assume that[11] $x_n \in B_\rho(x^+, \Delta_p)$ and observe that this implies that $\Delta_p(x^+, z_{n,0}) = \Delta_p(x^+, x_n) < \rho$ and consequently $\|z_{n,0}\| \le C_{\rho,x^+}$, with the constant $C_{\rho,x^+} > 0$ being defined in (3.30). Suppose now by induction that for some $k \in \mathbb{N}$, the inner iterates $z_{n,0}, ..., z_{n,k}$ are well-defined and satisfy the inequalities (3.31) and (3.32). Everything we need to complete the induction proof is that the inequality $\Delta_p(x^+, z_{n,k+1}) < \Delta_p(x^+, z_{n,k})$ holds. To this end, we adapt ideas of our previous work [39]. Applying the three points identity (2.21):

$$\begin{aligned}
\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, z_{n,k}\right) &= -\Delta_p(z_{n,k+1}, z_{n,k}) \\
&\quad + \left\langle J_p(z_{n,k+1}) - J_p(z_{n,k}), z_{n,k+1} - x^+\right\rangle \\
&\le \left\langle J_p(z_{n,k+1}) - J_p(z_{n,k}), z_{n,k+1} - x^+\right\rangle \\
&= -\frac{1}{\alpha_n}\langle j_r(A_n s_{n,k+1} - b_n), A_n(s_{n,k+1} - e_n)\rangle \\
&= \frac{1}{\alpha_n}\langle j_r(A_n s_{n,k+1} - b_n), (A_n e_n - b_n) - (A_n s_{n,k+1} - b_n)\rangle \\
&\le \frac{1}{\alpha_n}\left(\|A_n s_{n,k+1} - b_n\|^{r-1}\|A_n e_n - b_n\| - \|A_n s_{n,k+1} - b_n\|^r\right).
\end{aligned} \tag{3.64}$$

We would like to have the linearized residual in the $k-$th inner iterate, $\|A_n s_{n,k} - b_n\|$, instead of $\|A_n s_{n,k+1} - b_n\|$ in the rightmost term of above inequality. Observe however that, since the functional $\|\cdot\|^r$ is convex,

$$\|A_n s_{n,k} - b_n\|^r \le 2^{r-1}\left(\|A_n s_{n,k+1} - b_n\|^r + \|A_n(s_{n,k+1} - s_{n,k})\|^r\right),$$

which implies that

$$-\|A_n s_{n,k+1} - b_n\|^r \le -\frac{1}{2^{r-1}}\|A_n s_{n,k} - b_n\|^r + M^r\|z_{n,k+1} - z_{n,k}\|^r. \tag{3.65}$$

To estimate the last term in the right-hand side of (3.65) we recall the definition $\nabla_{n,k} = A_n^* j_r(A_n s_{n,k} - b_n)$ and make use of the $s-$convexity of X. Since X^* is s^*-smooth and as $p^* \ge s^*$, it follows from (2.25) and (3.32) that

$$\begin{aligned}
\|z_{n,k+1} - z_{n,k}\| &= \left\|J_{p^*}^*(J_p(z_{n,k+1})) - J_{p^*}^*(J_p(z_{n,k}))\right\| \\
&\le \overline{C}_{p^*,s^*}\Big(\|J_p(z_{n,k+1}) - J_p(z_{n,k})\|^{p^*-1} \\
&\qquad \vee \|J_p(z_{n,k})\|^{p^*-s^*}\|J_p(z_{n,k+1}) - J_p(z_{n,k})\|^{s^*-1}\Big) \\
&\le \overline{C}_{p^*,s^*}\left(\frac{1}{\alpha_n^{p^*-1}}\|\nabla_{n,k+1}\|^{p^*-1} \vee \frac{C_{\rho,x^+}^{p-s^*(p-1)}}{\alpha_n^{s^*-1}}\|\nabla_{n,k+1}\|^{s^*-1}\right).
\end{aligned}$$

As $p \le s \le r$, the inequality $(r-1)(p^*-1) - 1 \ge 0$ holds true and as

$$\|A_n s_{n,k+1} - b_n\| \le \|A_n s_{n,k} - b_n\| \le \|b_n\|,$$

[11] Similar to the DE method, this property needs to be assumed in order to prove inequality (3.32). This fact will be proven later in Theorem 38.

it follows that

$$\begin{aligned} \|\nabla_{n,k+1}\|^{p^*-1} &\leq M^{p^*-1} \|A_n s_{n,k+1} - b_n\|^{(r-1)(p^*-1)} \\ &\leq M^{p^*-1} \|b_n\|^{(r-1)(p^*-1)-1} \|A_n s_{n,k} - b_n\| . \end{aligned} \tag{3.66}$$

Proceeding similarly, replacing p^* with s^* and using at this time the inequality $(r-1)(s^*-1)-1 \geq 0$ we arrive at

$$M^r \|z_{n,k+1} - z_{n,k}\|^r \leq C(n)^r \|A_n s_{n,k} - b_n\|^r ,$$

with

$$C(n) := \frac{\overline{C}_{p^*,s^*} M^{p^*} \|b_n\|^{(r-1)(p^*-1)-1}}{\alpha_n^{p^*-1}} \vee \frac{\overline{C}_{p^*,s^*} C_{\rho,x^+}^{p-s^*(p-1)} M^{s^*} \|b_n\|^{(r-1)(s^*-1)-1}}{\alpha_n^{s^*-1}}.$$

Choose $0 < C_0 < 2^{-1/r^*}$ and observe that $C(n) \leq C_0$ if and only if $\alpha_n \geq \alpha_{\min,n}$ with

$$\alpha_{\min,n} := \frac{\overline{C}_{p^*,s^*}^{p-1} M^p \|b_n\|^{r-p}}{C_0^{p-1}} \vee \frac{\overline{C}_{p^*,s^*}^{s-1} C_{\rho,x^+}^{s-p} M^s \|b_n\|^{r-s}}{C_0^{s-1}}.$$

Hence, for $\alpha_n \geq \alpha_{\min,n}$,

$$M^r \|z_{n,k+1} - z_{n,k}\|^r \leq C_0^r \|A_n s_{n,k} - b_n\|^r .$$

Inserting it in (3.65), (3.65) in (3.64) and using the inequality

$$\|A_n s_{n,k+1} - b_n\| \leq \|A_n s_{n,k} - b_n\|$$

once again, we obtain

$$\begin{aligned} &\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, z_{n,k}\right) \\ &\qquad \leq \frac{1}{\alpha_n} \|A_n s_{n,k} - b_n\|^{r-1} \left(\|A_n e_n - b_n\| - C_1 \|A_n s_{n,k} - b_n\|\right), \end{aligned} \tag{3.67}$$

with $C_1 := 1/2^{r-1} - C_0^r > 0$, which has the same form as (3.37) with $\lambda_{n,k} = 1/\alpha_n$. Employing the TCC and choosing carefully the constant μ, we conclude, like in (3.28) and (3.38), that the right-hand side of (3.67) is negative, this is, $\Delta_p(x^+, z_{n,k+1}) < \Delta_p(x^+, z_{n,k})$, completing the induction proof.

Notice that, from TCC, the sequence of the residuals can be proven to be bounded, this is, $\|b_n\| \leq C_2$ (see (3.41)) and since $p \leq s \leq r$,

$$\alpha_{\min,n} \leq \frac{\overline{C}_{p^*,s^*}^{p-1} M^p C_2^{r-p}}{C_0^{p-1}} \vee \frac{\overline{C}_{p^*,s^*}^{s-1} C_{\rho,x^+}^{s-p} M^s C_2^{r-s}}{C_0^{s-1}} =: \alpha_{\min},$$

which means that the inequality $\alpha_n \geq \alpha_{\min}$ implies $\alpha_n \geq \alpha_{\min,n}$. Choosing, for instance, $\alpha_{\max} := K\alpha_{\min}$ with $K > 1$, we conclude that all the above results hold true for

$$\alpha_n \in [\alpha_{\min}, \alpha_{\max}] . \tag{3.68}$$

In particular, the choice α_n =constant is possible.

Remark 32 *See that, if a single step is given in each inner iteration, i.e., if $k_n = 1$ whenever (3.9) does not hold (this means that $k_{\max} = 1$ in (3.12)), then x_{n+1} minimizes*

$$T_n(x) = \frac{1}{r} \left\| y_{[n]} - F_{[n]}(x_n) - F'_{[n]}(x_n)(x - x_n) \right\|^r + \alpha_n \Delta_p(x, x_n) .$$

In Hilbert spaces this functional reads (with $p = r = 2$)

$$T_n(x) = \frac{1}{2}\left\|y_{[n]} - F_{[n]}(x_n) - F'_{[n]}(x_n)(x - x_n)\right\|^2 + \frac{\alpha_n}{2}\|x - x_n\|^2$$

revealing `K-REGINN-IT` *as a Kaczmarz version of the Levenberg-Marquardt method in Banach spaces. In case of a linear problem ($F_j = A_j$ are linear for all j's) we have*

$$T_n(x) = \frac{1}{r}\left\|y_{[n]} - A_{[n]}x\right\|^r + \alpha_n \Delta_p(x, x_n)$$

and the method is now a Kaczmarz version of the Iterated-Tikhonov method defined in [26].

3.1.4 Mixed gradient-Tikhonov methods

To apply a Tikhonov method in order to generate the inner iteration of `K-REGINN` as done in the last subsection, one needs to minimize a Tikhonov functional on each single step of the inner iteration. But, even finding an approximate minimizer for a Tikhonov functional is normally not a simple task and finding it exactly is almost impossible in general Banach spaces. Motivated by this reasoning, we propose to employ a dual gradient method to iteratively minimize a fixed Tikhonov functional and then use the resulting iteration as inner iteration of `K-REGINN`. The advantage of this procedure over a classical Tikhonov iteration as presented in Subsection 3.1.3 is clear: we do not intend to find a minimizer of a Tikhonov functional. We only apply a dual gradient method to iteratively improve the current iterate until finding an approximate minimizer good enough.

Assume Assumption 1 in page 31, and let X be uniformly smooth and $s-$convex with $p \leq s$. Define the Tikhonov functional

$$T_n(s) := \frac{1}{r}\|A_n s - b_n\|^r + \alpha_n \Omega_n(s), \tag{3.69}$$

where $r > 1$, $\alpha_n > 0$ and $\Omega_n \colon X \to \mathbb{R}_0^+$ is a subdifferentiable functional satisfying the condition: for each $n \in \mathbb{N}$ fixed, the sequence $(\overline{s}_m)_{m\in\mathbb{N}}$ is bounded whenever $(\Omega_n(\overline{s}_m))_{m\in\mathbb{N}}$ is bounded. This condition guarantees the existence of a minimizer of T_n (see the reasoning in the beginning of last subsection). Define now

$$J_p(z_{n,k+1}) := J_p(z_{n,k}) - \lambda_{n,k}\nabla T_{n,k}, \tag{3.70}$$

with $z_{n,0} := x_n$, $\lambda_{n,k} > 0$ and

$$\nabla T_{n,k} \in \partial T_n(s_{n,k}) = A_n^* J_r(A_n s_{n,k} - b_n) + \alpha_n \partial\Omega_n(s_{n,k})$$

with $s_{n,k} := z_{n,k} - x_n$. See Algorithm 5 for an implementation in pseudocode.

Since in each inner iteration the update $z_{n,k+1}$ is obtained following the direction of the gradient of a Tikhonov functional, we hope to obtain extra stability in comparison to the regular gradient methods, where the direction of the gradient of the linearized residual is chosen to update the current iterate. Compare (3.70) and (3.29).

We now apply a similar idea to that used to derive the Decreasing Error method in Subsection 3.1.2, in order to prove inequalities (3.31) and (3.32) at the same time. Assume that $x_n \in B_\rho(x^+, \Delta_p)$, which implies that $\|z_{n,0}\| \leq C_{\rho,x^+}$, see (3.30). Suppose by induction that for some $k \in \mathbb{N}$, the iterates $z_{n,0}, ..., z_{n,k}$ are well-defined and satisfy the inequalities (3.31) and (3.32). Again, everything we need in order to complete the induction proof is to prove that the inequality $\Delta_p(x^+, z_{n,k+1}) < \Delta_p(x^+, z_{n,k})$ holds. Making use of the three points identity (2.21) and inequality (2.24) we find, like in (3.33) and (3.35),

$$\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, z_{n,k}\right) \leq g(\lambda_{n,k}),$$

Algorithm 5 `K-REGINN` with mixed gradient-Tikhonov inner iteration

Input: x_N; (y^{δ_j}, δ_j); F_j; F_j', $j = 0, \ldots, d-1$; μ; τ; $p, r > 1$;

Output: x_N with $\|y_j^{\delta_j} - F_j(x_N)\| \le \tau\delta_j$, $j = 0, \ldots, d-1$;

$\ell := 0$; $x_0 := x_N$; $c := 0$;
while $c < d$ **do**
  for $j = 0 : d-1$ **do**
    $n := \ell d + j$;
    $b_n^\delta := y_j^{\delta_j} - F_j(x_n)$; $A_n := F_j'(x_n)$;
    if $\|b_n^\delta\| \le \tau\delta_j$ **then**
      $x_{n+1} := x_n$; $c := c + 1$;
    else
      $k := 0$; $s_{n,0} := 0$; choose $k_{\max,n} \in \mathbb{N}$; $\alpha_n > 0$ and $\Omega_n \colon X \to \mathbb{R}_0^+$ subdifferentiable;
      repeat
        choose $\lambda_{n,k} > 0$ and $\nabla T_{n,k} \in A_n^* J_r(A_n s_{n,k} - b_n^\delta) + \alpha_n \partial\Omega_n(s_{n,k})$;
        $z_{n,k} := s_{n,k} + x_n$;
        $J_p(z_{n,k+1}) := J_p(z_{n,k}) - \lambda_{n,k}\nabla T_{n,k}$;
        $s_{n,k+1} := J_{p^*}^*(J_p(z_{n,k+1})) - x_n$;
        $k := k + 1$;
      until $\left\|b_n^\delta - A_n s_{n,k}\right\| < \mu\left\|b_n^\delta\right\|$ or $k = k_{\max,n}$
      $x_{n+1} := x_n + s_{n,k}$; $c := 0$;
    end if
  end for
  $\ell := \ell + 1$;
end while
$x_N := x_{\ell d - c}$;

with

$$g(\lambda_{n,k}) := -\lambda_{n,k}\langle \nabla T_{n,k}, s_{n,k} - e_n\rangle + C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)}\lambda_{n,k}^{s^*}\|\nabla T_{n,k}\|^{s^*} \vee \lambda_{n,k}^{p^*}\|\nabla T_{n,k}\|^{p^*}\right)$$

Observe now that,

$$\langle \nabla T_{n,k}, s_{n,k} - e_n\rangle = \langle A_n^* j_r(A_n s_{n,k} - b_n), s_{n,k} - e_n\rangle + \alpha_n\left\langle s_{n,k}^*, s_{n,k} - e_n\right\rangle,$$

with some vector $s_{n,k}^* \in \partial\Omega_n(s_{n,k})$. Like in (3.34),

$$\langle A_n^* j_r(A_n s_{n,k} - b_n), s_{n,k} - e_n\rangle \ge \|A_n s_{n,k} - b_n\|^r - \|A_n s_{n,k} - b_n\|^{r-1}\|A_n e_n - b_n\|,$$

and as $s_{n,k}^* \in \partial\Omega_n(s_{n,k})$, it follows from definition of subgradient that,

$$\left\langle s_{n,k}^*, s_{n,k} - e_n\right\rangle \ge \Omega_n(s_{n,k}) - \Omega_n(e_n) \ge -\Omega_n(e_n).$$

If an upper bound $C_3 > 0$ for $\Omega_n(e_n)$ is known, i.e., if

$$\Omega_n(e_n) \le C_3 \tag{3.71}$$

for all $n \in \mathbb{N}$, then choosing $0 < \alpha_n \le (C_4/C_3)\|b_n\|^r$ with $0 < C_4 < 1$,

$$\langle \nabla T_{n,k}, s_{n,k} - e_n\rangle \ge \|A_n s_{n,k} - b_n\|^r - \|A_n s_{n,k} - b_n\|^{r-1}\|A_n e_n - b_n\| - C_4\|b_n\|^r.$$

Proceeding similarly to (3.39), we define $\overline{\lambda}_{DE} = \overline{\lambda}_{DE}(n,k)$ as

$$\overline{\lambda}_{DE} := C_1 \overline{\lambda}_{DE,s} \wedge C_2 \overline{\lambda}_{DE,p} \tag{3.72}$$

with C_1 and C_2 like in (3.39) and $\overline{\lambda}_{DE,\ell}$ being the same as $\lambda_{DE,\ell}$ but with $\nabla T_{n,k}$ replacing $\nabla_{n,k}$, i.e.,

$$\overline{\lambda}_{DE,\ell} := \frac{\|A_n s_{n,k} - b_n\|^{r(\ell-1)}}{\|\nabla T_{n,k}\|^{\ell}}.$$

Thus, $\lambda_{n,k} \in \left(0, \overline{\lambda}_{DE}\right]$ implies that

$$g(\lambda_{n,k}) \leq \lambda_{n,k} \left[\|A_n s_{n,k} - b_n\|^{r-1} \left(\|A_n e_n - b_n\| - C_5 \|A_n s_{n,k} - b_n\| \right) + C_4 \|b_n\|^r \right]. \tag{3.73}$$

where $C_5 = 1 - C_0$. See that $C_5 > C_4 > 0$ if $0 < C_0 < 1 - C_4$.

Inequality (3.73) above was achieved using neither the TCC nor the definition of k_n. However, as we will see later in Theorem 38, if the constant η in the TCC is small enough and the constant μ which stops the inner iteration of `K-REGINN` is well chosen, then the right-hand side of (3.73) is negative. Consequently, $g(\lambda_{n,k}) < 0$ and the induction proof is complete.

We assume now that the functionals Ω_n satisfy the property: $\Omega_n(s_{n,k})$ is uniformly bounded in n and k whenever $(s_{n,k})$ is uniformly bounded in n and k. In this case, since the sequences $(s_{n,k})_{0 \leq k \leq k_n}$, $n \in \mathbb{N}$ are uniformly bounded (see (3.42)), it follows that,

$$\alpha_n \|\Omega_n(s_{n,k})\| \lesssim \alpha_n \lesssim \|b_n\|^r \leq \frac{1}{\mu^r} \|A_n s_{n,k} - b_n\|^r$$

for $k = 0, ..., k_n - 1$. Thus,

$$\begin{aligned} \|\nabla T_{n,k}\| &\leq M \|A_n s_{n,k} - b_n\|^{r-1} + \alpha_n \|\Omega_n(s_{n,k})\| \\ &\lesssim \|A_n s_{n,k} - b_n\|^{r-1} \vee \|A_n s_{n,k} - b_n\|^r \\ &\lesssim \|A_n s_{n,k} - b_n\|^{r-1}, \end{aligned}$$

because $\|A_n s_{n,k} - b_n\|$ is uniformly bounded (see (3.41) and (3.42)). Consequently,

$$\overline{\lambda}_{DE,\ell} \gtrsim \|A_n s_{n,k} - b_n\|^{\ell - r}$$

and therefore,

$$\overline{\lambda}_{DE} \gtrsim \|A_n s_{n,k} - b_n\|^{s-r} \wedge \|A_n s_{n,k} - b_n\|^{p-r} \gtrsim \|A_n s_{n,k} - b_n\|^{s-r} \tag{3.74}$$

for $k = 0, ..., k_n - 1$.

Similarly to (3.43), one can prove for $s \leq r$ that $\overline{\lambda}_{DE,\ell} \gtrsim C_6^{-r}$, where $C_6 > 1$ is an upper bound to $\|A_n s_{n,k} - b_n\|$. This reasoning implies that the Landweber method, with constant step-size given by

$$\lambda_{LW} := C_6^{-r}, \tag{3.75}$$

satisfies $\lambda_{LW} \leq \overline{\lambda}_{DE}$, which implies that inequality (3.73) holds for this method.

Suppose from now on that the functional Ω_n is defined by $\Omega_n(s) := \frac{1}{p}\|s - \overline{x}_n\|^p$, with $(\overline{x}_n)_{n\in\mathbb{N}} \subset X$ being independent of k. Then the iteration (3.70) assumes the form

$$J_p(z_{n,k+1}) = J_p(z_{n,k}) - \lambda_{n,k}\left(A_n^* j_r(A_n s_{n,k} - b_n) + \alpha_n J_p(s_{n,k} - \overline{x}_n)\right). \tag{3.76}$$

In this case, as X is $s-$convex and hence strictly convex, Ω_n is strictly convex too and the minimizer of T_n, defined in (3.69), is unique.

In order to find an upper bound C_3 for $\Omega_n(e_n) = \frac{1}{p}\|e_n - \overline{x}_n\|^p$, an upper bound for $\|x^+\|$ normally needs to be known. For example, if $\overline{x}_n := -x_n$, i.e., $\Omega_n(s) = \frac{1}{p}\|s + x_n\|^p$, then we need a bound for $\Omega_n(e_n) = \frac{1}{p}\|x^+\|^p$. If the functional $\Omega_n(s) = \frac{1}{p}\|s - (x_0 - x_n)\|^p$ is considered, then

$$\Omega_n(e_n) = \frac{1}{p}\left\|x^+ - x_0\right\|^p \le \frac{1}{p}\left(\left\|x^+\right\| + \|x_0\|\right)^p.$$

The bound $\Omega_n(e_n) \le \left(\|x^+\| + C_{x^+,\rho}\right)^p/p$ holds in case of $\Omega_n(s) = \frac{1}{p}\|s\|^p$ is chosen and $x_n \in B_\rho(x^+, \Delta_p)$ is assumed.

We finish this subsection giving a practical example: consider the iteration (3.76) with $\overline{x}_n := 0$, which is equivalent to the dual gradient iteration (3.70) applied to the Tikhonov functional

$$T_n(s) = \frac{1}{r}\|A_n s - b_n\|^r + \alpha_n \frac{1}{p}\|s\|^p.$$

If one wants to use SD method to iteratively minimize this functional, it is necessary to define the step-size $\lambda_{n,k}$ in (3.70) such that the number $T_n(s_{n,k+1})$ is as small as possible. As

$$s_{n,k+1} = J_{p^*}^*\left(J_p(s_{n,k} + x_n) - \lambda_{n,k}\nabla T_{n,k}\right) - x_n,$$

we conclude that $\lambda_{n,k} \in \arg\min_{\lambda \in \mathbb{R}_0^+} h(\lambda)$, with $h\colon\ \mathbb{R}_0^+ \to \mathbb{R}_0^+$ defined by

$$h(\lambda) := T_n\left(J_{p^*}^*\left(J_p(s_{n,k} + x_n) - \lambda\nabla T_{n,k}\right) - x_n\right).$$

It is possible to employ now an iterative method to find a minimizer of h if it exists. Although $\mathrm{D}(h) \subset \mathbb{R}$, this optimization problem is often difficult to be solved and can be computationally expensive, particularly if the Banach space Y has poor convexity and smoothness properties.

In Hilbert spaces however, the use of polarization identity (with $p = r = 2$) yields

$$h(\lambda) = T_n(s_{n,k} - \lambda\nabla T_{n,k}) = T_n(s_{n,k}) - \lambda\|\nabla T_{n,k}\|^2 + \frac{1}{2}\lambda^2\left(\|A\nabla T_{n,k}\|^2 + \alpha_n\|\nabla T_{n,k}\|^2\right).$$

Requiring $h'(\lambda_{SD}) = 0$, we obtain the explicit step-size

$$\lambda_{SD} = \frac{\|\nabla T_{n,k}\|^2}{\|A\nabla T_{n,k}\|^2 + \alpha_n\|\nabla T_{n,k}\|^2}.$$

The result implies in particular that $1/\left(M^2 + \alpha_n\right) \le \lambda_{SD} \le 1/\alpha_n$. Finally, from the modified version of Young's inequality[12], it follows that for all $a, b \ge 0$ and $\epsilon > 0$,

$$(a+b)^2 \le a^2 + 2\left(\frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2}\right) + b^2 = \frac{1+\epsilon}{\epsilon}a^2 + (1+\epsilon)\,b^2.$$

Thus,

$$\begin{aligned}\|\nabla T_{n,k}\|^4 &= \langle\nabla T_{n,k}, \nabla T_{n,k}\rangle^2 = \left(\langle A\nabla T_{n,k}, A_n s_{n,k} - b_n\rangle + \alpha_n\langle\nabla T_{n,k}, s_{n,k}\rangle\right)^2 \\ &\le \frac{1+\epsilon}{\epsilon}\|A\nabla T_{n,k}\|^2\|A_n s_{n,k} - b_n\|^2 + (1+\epsilon)\,\alpha_n^2\|\nabla T_{n,k}\|^2\|s_{n,k}\|^2,\end{aligned}$$

for all $\epsilon > 0$. Now, if

$$\alpha_n \le \frac{\|A_n s_{n,k} - b_n\|^2}{\epsilon\|s_{n,k}\|^2},$$

[12] For all $a, b \ge 0$ and $\epsilon > 0$ it holds: $ab \le a^2/2\epsilon + \epsilon b^2/2$.

it follows that

$$\|\nabla T_{n,k}\|^4 \leq \frac{1+\epsilon}{\epsilon} \|A_n s_{n,k} - b_n\|^2 \left(\|A\nabla T_{n,k}\|^2 + \alpha_n \|\nabla T_{n,k}\|^2\right),$$

and consequently,

$$\lambda_{SD} \leq \frac{1+\epsilon}{\epsilon} \frac{\|A_n s_{n,k} - b_n\|^2}{\|\nabla T_{n,k}\|^2},$$

for all $\epsilon > 0$. Choosing $0 < \epsilon \leq 1/(1 - C_0)$, we find $(1+\epsilon)/\epsilon \leq C_0$, which implies that $\lambda_{SD} \leq \overline{\lambda}_{DE}$, see (3.72).

Chapter 4

Convergence Analysis of K-REGINN

Although we do not consider in this chapter any specific method to generate the inner iteration of K-REGINN, some properties of the inner iteration sequence $(s_{n,k})_{k\in\mathbb{N}}$ are required in form of assumptions. Only these assumptions are then used to prove general results for the outer iteration sequence $(x_n)_{n\in\mathbb{N}}$. Many of these required properties were previously proven for the sequences generated by the methods presented in Section 3.1, which means that these sequences can be used as inner iteration of K-REGINN and the respective results proved in this chapter are assured for these methods. We point out however, that any sequence satisfying the necessary requirements assumed in this chapter can be used as inner iteration of K-REGINN and all the respective results will accordingly hold for them.

We assume again that only noisy data is available, that is, $\delta > 0$ in (3.8) is given. To stress this fact, we accordingly add a superscript δ in the residual b_n : $b_n^\delta = y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n)$.

Lemma 33 *Let X and Y be Banach spaces with Y being smooth. Let Assumption 1 in page 31 hold true and $k_n = k_{REG}$. If x_n is well-defined and $k_{REG} < \infty$, then s_{n,k_n} is a descent direction for the functional $\psi_n(\cdot) := \left\| F_{[n]}(\cdot) - y_{[n]}^{\delta_{[n]}} \right\|$ from x_n.*

Proof. The result follows from Proposition 22 and $\left\| A_n s_{n,k_n} - b_n^\delta \right\| < \mu \left\| b_n^\delta \right\| < \left\| b_n^\delta \right\|$. ■

The above lemma shows that $\left\| F_{[n]}(x_n + t s_{n,k_n}) - y_{[n]}^{\delta_{[n]}} \right\| < \left\| F_{[n]}(x_n) - y_{[n]}^{\delta_{[n]}} \right\|$ for $t > 0$ small enough. But, as we will see later in Theorem 35, if the right assumptions are required, then $t = 1$ can be used and consequently $\left\| F_{[n]}(x_{n+1}) - y_{[n]}^{\delta_{[n]}} \right\| < \left\| F_{[n]}(x_n) - y_{[n]}^{\delta_{[n]}} \right\|$.

4.1 Termination and qualified decreasing residual

To show termination of K-REGINN, we need only some few restrictions on the sequence $(s_{n,k})_{k\in\mathbb{N}}$. Keeping this goal in mind, we use for the first results a level set based analysis. The LEVEL SET associated to the point $\overline{x} \in \mathrm{D}(F)$ is defined as

$$\pounds(\overline{x}) := \left\{ x \in \mathrm{D}(F) : \left\| F_j(x) - y_j^\delta \right\| \leq \left\| F_j(\overline{x}) - y_j^\delta \right\|,\ j = 0, ..., d-1 \right\}.$$

For any $\overline{x} \in \mathrm{D}(F)$, it holds $\overline{x} \in \pounds(\overline{x})$ and $x \in \pounds(\overline{x}) \Longrightarrow \pounds(x) \subset \pounds(\overline{x})$.

In the first step towards the termination of K-REGINN, we prove that k_n is finite. For that, we require the first assumption on the sequence $(s_{n,k})_{k\in\mathbb{N}}$.

Assumption 2 *If the iterate x_n of K-REGINN is well-defined, then there exists a constant $C \geq 0$ independent on n and k such that*

$$\lim_{k\to\infty} \left\| A_n s_{n,k} - b_n^\delta \right\|^r \leq \frac{1}{1+C} \left(\left\| A_n s - b_n^\delta \right\|^r + C \left\| b_n^\delta \right\|^r \right),\ \textit{for all } s \in X.$$

The Assumption 2 holds true if the sequence $(s_{n,k})_{k\in\mathbb{N}}$ is generated from the following methods:

- Each primal gradient method, defined by iteration (3.15), presented in Subsection 3.1.1, with step-size $\lambda_k \in [c\lambda_{MSD}, \lambda_{MSD}]$ and $0 < c < 1$, where λ_{MSD} is defined in (3.18) (this includes the MSD method itself) and assuming that $Y_{[n]}$ is $r-$smooth (see Lemma 24).
- The primal gradient LW method, defined by iteration (3.15) with the constant step-size (3.22) and with the additional requirement $p \leq r$. For this method, we also assume the hypotheses of Lemma 24. In particular, the space $Y_{[n]}$ needs to be $r-$smooth.
- The Tikhonov-Phillips method presented in Subsection 3.1.3 under the hypotheses of Lemma 30.

We did not prove this property for the dual gradient methods, Iterated-Tikhonov and mixed gradient-Tikhonov methods. Therefore, the results shown in this section, which are based on Assumption 2 above are in principle not guaranteed for these methods. However, the inequalities (3.37), (3.67) and (3.73) are strong enough to guarantee not only the results of this section, but even the stronger results of the next one (see Theorem 38 below).

Assumption 2 is a weaker version of

$$\lim_{k\to\infty}\left\|A_n s_{n,k} - b_n^{\delta}\right\| = \inf_{s\in X}\left\|A_n s - b_n^{\delta}\right\|,$$

which is equivalent to[1]

$$\lim_{k\to\infty} A_n s_{n,k} = P_{\overline{R(A_n)}} b_n^{\delta}$$

in Hilbert spaces. This last property is exactly the Assumption (2.2) in [36], used to prove [36, Lemma 2.3], which is a similar version of next proposition. Assumption 2 combined with a careful choice of μ implies that the inner iteration stops after finitely many iterations.

Proposition 34 *Assume Assumption 2 and that the iterate x_n is well-defined. Assume further that the TCC (Assumption 1(c), page 31) holds for x_n, x^+ and $F_{[n]}$. Then for $\tau > (1+\eta)/(1-\eta)$ and $\mu \in (\mu_{\min}, 1)$ with*

$$\mu_{\min}^r := \frac{\left(\frac{1+\eta}{\tau} + \eta\right)^r + C}{1+C},$$

the stop index k_n is finite.

Proof. $\eta < 1$ together with the restriction on τ imply $(1+\eta)/\tau + \eta < 1$, which in turn implies that the interval of picking the parameter μ is non-empty. If inequality (3.9) is satisfied, then $k_n = 0$. Otherwise, $\left\|b_n^{\delta}\right\| > \tau\delta_{[n]}$ and we define $e_n := x^+ - x_n$, use (3.7) and the TCC to obtain

$$\begin{aligned}\left\|A_n e_n - b_n^{\delta}\right\| &= \left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n) - F_{[n]}'(x_n)\left(x^+ - x_n\right)\right\| \\ &\leq \left\|y_{[n]}^{\delta_{[n]}} - y_{[n]}\right\| + \left\|F_{[n]}\left(x^+\right) - F_{[n]}(x_n) - F_{[n]}'(x_n)\left(x^+ - x_n\right)\right\| \\ &\leq \delta_{[n]} + \eta\left\|F_{[n]}\left(x^+\right) - F_{[n]}(x_n)\right\| \\ &\leq \delta_{[n]} + \eta\left(\delta_{[n]} + \left\|b_n^{\delta}\right\|\right) < \left(\frac{1+\eta}{\tau} + \eta\right)\left\|b_n^{\delta}\right\|.\end{aligned} \tag{4.1}$$

[1] $P_{\overline{R(A_n)}}\colon Y \to Y$ is the orthogonal projection in $\overline{R(A_n)}$.

Now, we apply Assumption 2 with $s = e_n$ to prove

$$\begin{aligned}\lim_{k\to\infty}\left\|A_n s_{n,k} - b_n^\delta\right\|^r &\leq \frac{1}{1+C}\left(\left\|A_n e_n - b_n^\delta\right\|^r + C\left\|b_n^\delta\right\|^r\right) \\ &\leq \frac{\left(\frac{1+\eta}{\tau}+\eta\right)^r + C}{1+C}\left\|b_n^\delta\right\|^r < \mu^r\left\|b_n^\delta\right\|^r,\end{aligned}$$

which means that $k_n \leq k_{REG} < \infty$. ■

Under the hypotheses of above proposition, we conclude that x_n well-defined implies x_{n+1} well-defined, this is, the inner iteration terminates. The next theorem shows that the outer iteration also terminates after finitely many iterations if $d = 1$ (only a single equation in (3.6) is considered) and $k_n = k_{REG}$. This theorem is an adaptation of [36, Theo. 2.8].

Theorem 35 (Termination) *Let X and Y be Banach spaces with Y being smooth. Let $D(F)$ be open, $d = 1$, $k_n = k_{REG}$ and suppose that Assumption 2 holds true. Suppose additionally that Assumption 1 in page 31 holds true with $B_\rho(x^+, \Delta_p)$ replaced by $\pounds(\overline{x})$ for some $\overline{x} \in X$ fixed and start with $x_0 \in \pounds(\overline{x})$. Further, assume that the constant η in the TCC (Assumption 1(c)) is small enough to satisfy the inequality*

$$\frac{\eta^r + C}{1+C} < (1-2\eta)^r \tag{4.2}$$

and the constant τ in the discrepancy principle (3.13) *is large enough to verify the inequality*

$$\left(\frac{1+\eta}{\tau}+\eta\right)^r + C < (1+C)(1-2\eta)^r. \tag{4.3}$$

Define now the constants $0 < \mu_{\min} < \mu_{\max} < 1$ as

$$\mu_{\min}^r := \frac{\left(\frac{1+\eta}{\tau}+\eta\right)^r + C}{1+C}, \quad \mu_{\max} := 1-2\eta \tag{4.4}$$

and pick up $\mu \in (\mu_{\min}, \mu_{\max})$. Then, there exists a constant $0 < \Lambda < 1$ such that:

1. *all the iterates of `REGINN` are well-defined and belong to the level set $\pounds(x_0)$ as long as the discrepancy principle* (3.13) *is not satisfied;*
2. *there exists $N = N(\delta) \in \mathbb{N}$ satisfying* (3.13) *and*
$$N < \frac{\ln(\tau\delta/\left\|b_0^\delta\right\|)}{\ln\Lambda} + 1; \tag{4.5}$$
3. *the residual has the qualified decreasing behavior*
$$\left\|b_{n+1}^\delta\right\| \leq \Lambda\left\|b_n^\delta\right\| \tag{4.6}$$
for $n = 0, ..., N-1$.

Proof. We first discuss the choices of the constants. See that

$$\lim_{\eta\to 0^+}\left(\frac{\eta^r + C}{1+C}\right)^{\frac{1}{r}} + 2\eta = \left(\frac{C}{1+C}\right)^{\frac{1}{r}} < 1,$$

which implies that (4.2) is verified provided η small enough. Now, due to the restriction on η,

$$\lim_{\tau\to\infty}\frac{\left(\frac{1+\eta}{\tau}+\eta\right)^r + C}{1+C} = \frac{\eta^r + C}{1+C} < (1-2\eta)^r$$

and therefore, inequality (4.3) is verified whenever τ is large enough. Further, the restriction on τ implies that the constants $\mu_{\min}$ and $\mu_{\max}$ in (4.4) are well-defined and thus the interval for choosing the tolerance μ is non-empty. Finally, choosing $\Lambda \in \mathbb{R}$ satisfying

$$\mu + \frac{\eta\,(1+\mu)}{1-\eta} < \Lambda < 1,$$

we observe that the restriction $\mu < \mu_{\max}$ implies that $\mu + \eta\,(1+\mu)\,/\,(1-\eta) < 1$. Hence Λ is well-defined. Further, as $1 - 2\eta < 1$, we also have $\tau > (1+\eta)\,/\,(1-\eta)$ and Proposition 34 applies.

We use now an inductive argument: x_0 is well-defined and belongs to $\pounds\,(x_0)$. Suppose that the iterates $x_0, ..., x_n$ with $n \leq N-1$, are well-defined, belong to $\pounds\,(x_0)$ and satisfy inequality (4.6). From Proposition 34, $k_n < \infty$ and consequently $x_{n+1} = x_n + s_{n,k_n}$ is well-defined. We prove next that $x_{n+1} \in \mathrm{D}(F)$ and inequality (4.6) holds for this vector, which will prove that $x_{n+1} \in \pounds\,(x_0)$. In fact, for each $t \in \mathbb{R}$ define the vector

$$x_{n,t} := x_n + t s_{n,k_n} \in X$$

and see that for each $t \in \mathbb{R}$ such that $x_{n,t} \in \overline{\pounds\,(\overline{x})} \subset \mathrm{D}(F)$, it is true that (see Assumption 1(c))

$$\begin{aligned}\left\|F\,(x_{n,t}) - F\,(x_n) - F'\,(x_n)\,(x_{n,t} - x_n)\right\| &\leq \frac{\eta}{1-\eta}\left\|F'\,(x_n)\,(x_{n,t} - x_n)\right\| \\ &= t\frac{\eta}{1-\eta}\left\|F'\,(x_n)\,s_{n,k_n}\right\|,\end{aligned}$$

and from definition (3.10) of k_{REG},

$$\left\|F'\,(x_n)\,s_{n,k_n}\right\| - \left\|b_n^\delta\right\| \leq \left\|b_n^\delta - F'\,(x_n)\,s_{n,k_n}\right\| < \mu\left\|b_n^\delta\right\|,$$

which implies that $\|F'\,(x_n)\,s_{n,k_n}\| < (1+\mu)\left\|b_n^\delta\right\|$. It follows that

$$\begin{aligned}\left\|F\,(x_{n,t}) - y^\delta\right\| &\leq (1-t)\left\|b_n^\delta\right\| + t\left\|b_n^\delta - F'\,(x_n)\,s_{n,k_n}\right\| \\ &\quad + \left\|F\,(x_{n,t}) - F\,(x_n) - F'\,(x_n)\,t s_{n,k_n}\right\| \\ &< (1-t)\left\|b_n^\delta\right\| + t\mu\left\|b_n^\delta\right\| + t\frac{\eta\,(1+\mu)}{1-\eta}\left\|b_n^\delta\right\| \\ &= \left[1 - t\left(1 - \left(\mu + \frac{\eta\,(1+\mu)}{1-\eta}\right)\right)\right]\left\|b_n^\delta\right\| \leq [1 - t\,(1-\Lambda)]\left\|b_n^\delta\right\|.\end{aligned} \tag{4.7}$$

for each $t \in [0,1]$ such that $x_{n,t} \in \overline{\pounds\,(\overline{x})}$. Let

$$t_{\max} := \sup\left\{t \in [0,1] : x_{n,t} \in \pounds\,(x_0)\right\}.$$

We want to prove that $t_{\max} = 1$. Observe first that the above set where the supremum is taken is non-empty because from Lemma 33, s_{n,k_n} is a descent direction for the functional $\psi\,(\cdot) := \left\|F\,(\cdot) - y^\delta\right\|$ from x_n and since $\mathrm{D}(F)$ is open, there exists a constant $\overline{t} > 0$ such that for all $0 < t \leq \overline{t}$, $x_{n,t} \in \mathrm{D}(F)$ and

$$\left\|F\,(x_{n,t}) - y^\delta\right\| < \left\|F\,(x_n) - y^\delta\right\| \overset{\text{Induction}}{\leq} \left\|F\,(x_0) - y^\delta\right\|.$$

This implies that $x_{n,t} \in \pounds\,(x_0)$ for all $0 < t \leq \overline{t}$. Assume next that $t_{\max} < 1$, i.e., $x_{n,t_{\max}} \in \partial\pounds\,(x_0)$. As $x_0 \in \pounds\,(\overline{x})$, it follows that

$$x_{n,t_{\max}} \in \partial\pounds\,(x_0) \subset \overline{\pounds\,(x_0)} \subset \overline{\pounds\,(\overline{x})} \subset \mathrm{D}\,(F).$$

Thus,

$$\left\|F\left(x_{n,t_{\max}}\right)-y^{\delta}\right\|<\left[1-t_{\max}\left(1-\Lambda\right)\right]\left\|b_{n}^{\delta}\right\|<\left\|b_{n}^{\delta}\right\|\leq\left\|b_{0}^{\delta}\right\|,$$

which contradicts $x_{n,t_{\max}}\in\partial\pounds\left(x_{0}\right)$. Choosing $t=1$ in (4.7) we find (4.6). Now, from (4.6), $\left\|b_{n}^{\delta}\right\|\leq\Lambda^{n}\left\|b_{0}^{\delta}\right\|\rightarrow0$ as $n\rightarrow\infty$, which implies that (3.13) is satisfied for n large enough. Finally, from

$$\tau\delta<\left\|b_{N-1}^{\delta}\right\|\leq\Lambda^{N-1}\left\|b_{0}^{\delta}\right\|$$

we obtain

$$N-1<\frac{\ln\left(\tau\delta/\left\|b_{0}^{\delta}\right\|\right)}{\ln\left(\Lambda\right)}.$$

■

Note that the level set $\pounds\left(\overline{x}\right)$ can be replaced by $\pounds\left(x_{0}\right)$ or X in the above theorem.

Corollary 36 *Assume all the hypotheses of Theorem 35. Then,* $F\left(x_{N(\delta)}\right)\rightarrow F\left(x^{+}\right)$ *as* $\delta\rightarrow0$.

Proof. The result follows from

$$\left\|y-F\left(x_{N(\delta)}\right)\right\|\leq\left\|y^{\delta}-y\right\|+\left\|y^{\delta}-F\left(x_{N(\delta)}\right)\right\|\leq\left(1+\tau\right)\delta. \tag{4.8}$$

■

From Tangential Cone Condition, Assumption $1(c)$, page 31, it follows easily that

$$\left\|F_{j}(v)-F_{j}(w)\right\|\leq\frac{1}{1-\eta}\left\|F_{j}^{\prime}(w)\left(v-w\right)\right\|$$

for all $v,w\in B_{\rho}\left(x^{+},\Delta_{p}\right)$ and $j=0,\ldots,d-1$. Now, if $d=1$ and x^{+} is the unique solution of (3.1) in a neighborhood of x^{+}, then the null space of $A:=F^{\prime}\left(x^{+}\right)$ satisfies $N\left(A\right)=\{0\}$ because if $0\neq v\in N\left(A\right)$, then for $t>0$ small enough,

$$\left\|F(x^{+}-tv)-y\right\|=\left\|F(x^{+}-tv)-F(x^{+})\right\|\leq\frac{1}{1-\eta}t\left\|Av\right\|=0.$$

In general, the function $\left\|\cdot\right\|_{A}:X\rightarrow\mathbb{R},\ x\longmapsto\left\|Ax\right\|$ is a semi-norm in X and it is a norm in case of A being injective (it is a weaker norm than the standard norm, because $\left\|x\right\|_{A}=\left\|Ax\right\|\leq\left\|A\right\|\left\|x\right\|$, for all $x\in X$).

Corollary 37 *Assume all the hypotheses of Theorem 35. Then,* $\left\|x_{N(\delta)}-x^{+}\right\|_{A}\rightarrow0$ *as* $\delta\rightarrow0$

Proof. From inequality (4.8) of Corollary 36 and Assumption $1(c)$,

$$\begin{aligned}\left\|x_{N(\delta)}-x^{+}\right\|_{A}&=\left\|F^{\prime}\left(x^{+}\right)\left(x_{N(\delta)}-x^{+}\right)\right\|\\&\leq\left(\eta+1\right)\left\|F\left(x_{N(\delta)}\right)-F\left(x^{+}\right)\right\|\leq\left(\eta+1\right)\left(1+\tau\right)\delta.\end{aligned}$$

■

The boundedness of $\pounds\left(x_{0}\right)$ would certainly facilitate a convergence analysis of the family $\left(x_{N(\delta)}\right)_{\delta>0}\subset\pounds\left(x_{0}\right)$ using for instance, a weak convergence argument in reflexive Banach spaces. There is no reason however, to believe that this is the case. On the contrary, this set is expected to be unbounded for ill-posed problems, which stimulate us to concentrate in a local convergence analysis restricted to the bounded set $B_{\rho}\left(\Delta_{p},x^{+}\right)$.

4.2 Decreasing error and weak convergence

Assumption 3 *If the iterate x_n of* K-REGINN *is well-defined and the definitions $e_n := x^+ - x_n$ and $z_{n,k} := s_{n,k} + x_n$ with $s_{n,0} := 0$ are adopted, then there exist constants $0 \leq K_0 < K_1 \leq 1$ and positive numbers $\lambda_{n,k}$ satisfying*

$$\lambda_{n,k} \gtrsim \left\| A_n s_{n,k} - b_n^\delta \right\|^t, \text{ with } t > -r$$

such that

$$\begin{aligned}&\Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, \widehat{z}_{n,k}\right)\\ &\qquad \leq \lambda_{n,k}\left[\left\| A_n s_{n,k} - b_n^\delta \right\|^{r-1}\left(\left\| A_n e_n - b_n^\delta \right\| - K_1 \left\| A_n s_{n,k} - b_n^\delta \right\|\right) + K_0 \left\| b_n^\delta \right\|^r\right],\end{aligned}$$

where either $\widehat{z}_{n,k} = z_{n,k}$ or $\widehat{z}_{n,k} = x_n$.

The following methods satisfy the required properties of Assumption 3, with $\widehat{z}_{n,k} = z_{n,k}$, provided X is uniformly smooth and $s-$convex with $p \leq s$:

- The dual gradient method DE, presented in (3.29) and (3.39), Subsection 3.1.2 with $K_1 = 1 - C_0$, $0 < C_0 < 1$, $K_0 = 0$ (see (3.37)) and $t = s - r$ (see (3.40));
- Each dual gradient method, defined by iteration (3.29) in Subsection 3.1.2, with step-size satisfying $\lambda_{n,k} \in \left[c \left\| A_n s_{n,k} - b_n^\delta \right\|^t, \lambda_{DE}\right]$ where $c > 0$ and $t > -r$ (K_1 and K_0 are the same as the DE method). In particular, the MSD method as defined in (3.45), with $-r < t \leq 0$, see (3.46), and LW method with step-size being given by a small constant (see (3.43)) and with $s \leq r$. For this last method, it obviously holds $t = 0$;
- The Bregman variation of the IT method defined by the implicit iteration (3.62) in Subsection 3.1.3 with $K_1 = 1/2^{r-1} - C_0^r$, $0 < C_0^r < 1/2^{r-1}$, $\lambda_{n,k} = 1/\alpha_n$ and $K_0 = t = 0$ (see (3.67) and (3.68));
- Each of the mixed gradient-Tikhonov methods presented in Subsection 3.1.4 and defined by iteration (3.76), with step-size satisfying $\lambda_{n,k} \in \left[c \left\| A_n s_{n,k} - b_n^\delta \right\|^t, \overline{\lambda}_{DE}\right]$ with $c > 0$ and $t > -r$, where $\overline{\lambda}_{DE}$ is defined in (3.72). In particular, the DE method itself (see (3.74)) and the LW method with a small constant step-size (see (3.75)). For these methods, $0 < K_0 < 1 - C_0 = K_1$ with $0 < C_0 < 1$, as we can see in (3.73).

The Assumption 3 is verified for $\widehat{z}_{n,k} = x_n$ if X is uniformly smooth and uniformly convex for:

- The Bregman variation of TP method as presented in (3.58), Subsection 3.1.3, with $K_1 = 1$, $K_0 = (r\theta - 1)/r < 1$, $\lambda_{n,k} = 1/\alpha_k$ and $t = 0$ (see (3.59) and (3.57)).

Assumption 3 is a generalization of [36, Assumption (3.2)] and the next Theorem corresponds to Theorem 3.1 in the same reference.

Theorem 38 (Decreasing error) *Let X and Y be Banach spaces with X being uniformly smooth and uniformly convex. Assume that Assumption 3 and Assumption 1 in page 31 hold true with the constant η in the TCC satisfying*

$$0 \leq \eta < K_1 - K_0,$$

and the constant τ *in* (3.13) *satisfying*

$$\tau > \frac{\eta + 1}{K_1 - K_0 - \eta}. \tag{4.9}$$

Assume additionally that Assumption 2 in page 53 holds true with $C = 0$ *in case of* $\widehat{z}_{n,k} = x_n$ *and start with* $x_0 \in B_\rho\left(x^+, \Delta_p\right)$. *Then for* $\mu \in (\mu_{\min}, 1)$ *with*

$$\mu_{\min}^r := \frac{\frac{\eta+1}{\tau} + \eta + K_0}{K_1},$$

1. *all the iterates of* `K-REGINN` *are well-defined and belong to* $B_\rho\left(x^+, \Delta_p\right) \subset D(F)$ *as long as the discrepancy principle* (3.13) *is not satisfied;*
2. *the outer iteration terminates, this is, there exists a number* $N = N(\delta) \in \mathbb{N}$ *satisfying* (3.13);
3. *the iterates have the error decreasing behavior*
$$\Delta_p\left(x^+, x_{n+1}\right) \le \Delta_p\left(x^+, x_n\right) \tag{4.10}$$
for $n = 0, ..., N-1$, *where the equality holds if and only if* (3.9) *holds;*
4. *the generated sequence is bounded:*
$$\|x_n\| \le C_{\rho,x^+} \text{ for all } \delta > 0 \text{ and } n = 0, ..., N, \tag{4.11}$$
where $C_{\rho,x^+} > 0$ *is defined in* (3.30);
5. *there exists a constant* $C_0 > 0$ *such that the inequality*
$$C_0 \sum_{k=0}^{k_n - 1} \lambda_{n,k} \left\| A_n s_{n,k} - b_n^\delta \right\|^r \le \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right) \tag{4.12}$$
holds if $\widehat{z}_{n,k} := z_{n,k}$ *and*
$$C_0 \lambda_{n,k_n-1} \left\| A_n s_{n,k_n-1} - b_n^\delta \right\|^r \le \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right) \tag{4.13}$$
if $\widehat{z}_{n,k} := x_n$.

Proof. The restrictions on the constant η imply the well-definedness of τ and $\mu_{\min} < 1$. We use induction: as $x_0 \in B_\rho\left(x^+, \Delta_p\right)$, the bound $\|x_0\| \le C_{\rho,x^+}$ holds (see (2.20)). Suppose that the iterates $x_0, ..., x_n$, with $n < N$ are well-defined, belong to $B_\rho\left(x^+, \Delta_p\right)$ and satisfy (4.10) as well as (4.11). If (3.9) is verified, then $x_{n+1} = x_n$ and all the required properties hold true for x_{n+1}. Otherwise, as $\left\|b_n^\delta\right\| \le \frac{1}{\mu}\left\|A_n s_{n,k} - b_n^\delta\right\|$ for $k = 0, ..., k_n - 1$, we proceed like in (4.1) to validate

$$\left\| A_n e_n - b_n^\delta \right\| < \left(\frac{1+\eta}{\tau} + \eta \right) \left\| b_n^\delta \right\| \le C_{\eta,\tau,\mu} \left\| A_n s_{n,k} - b_n^\delta \right\|,$$

with $C_{\eta,\tau,\mu} := \left(\frac{1+\eta}{\tau} + \eta\right)/\mu$. Using now $\mu_{\min} < \mu < 1$, we get from Assumption 3,

$$\begin{aligned} \Delta_p\left(x^+, z_{n,k+1}\right) - \Delta_p\left(x^+, \widehat{z}_{n,k}\right) &< \lambda_{n,k}\left[\left(C_{\eta,\tau,\mu} - K_1\right) \left\| A_n s_{n,k} - b_n^\delta \right\|^r + K_0 \left\| b_n^\delta \right\|^r \right] \\ &\le \lambda_{n,k} \left(C_{\eta,\tau,\mu} - K_1 + \frac{K_0}{\mu^r} \right) \left\| A_n s_{n,k} - b_n^\delta \right\|^r \\ &\le -C_0 \lambda_{n,k} \left\| A_n s_{n,k} - b_n^\delta \right\|^r, \end{aligned} \tag{4.14}$$

with $C_0 := K_1 - \left(\frac{1+\eta}{\tau} + \eta + K_0\right)/\mu^r > 0$.

We study first the case $\widehat{z}_{n,k} = z_{n,k}$. Using the inequality $\lambda_{n,k} \gtrsim \left\|A_n s_{n,k} - b_n^\delta\right\|^t$, with $t > -r$ (see Assumption 3), we obtain for all $\ell \leq k_n$,

$$\begin{aligned}\left(\mu\left\|b_n^\delta\right\|\right)^{t+r}\ell &\leq \sum_{k=0}^{\ell-1}\left\|A_n s_{n,k} - b_n^\delta\right\|^{t+r} \\ &\lesssim \sum_{k=0}^{\ell-1} C_0\lambda_{n,k}\left\|A_n s_{n,k} - b_n^\delta\right\|^r \overset{(4.14)}{\leq} \sum_{k=0}^{\ell-1}\left(\Delta_p\left(x^+, z_{n,k}\right) - \Delta_p\left(x^+, z_{n,k+1}\right)\right) \\ &= \Delta_p\left(x^+, z_{n,0}\right) - \Delta_p\left(x^+, z_{n,\ell}\right) \leq \Delta_p\left(x^+, z_{n,0}\right) = \Delta_p\left(x^+, x_n\right) < \infty,\end{aligned}$$

which implies that $\ell < \infty$ and consequently $k_n < \infty$. This means that the inner iteration terminates. Using $\ell = k_n$ in the above inequality we obtain

$$0 < \sum_{k=0}^{k_n-1} C_0\lambda_{n,k}\left\|A_n s_{n,k} - b_n^\delta\right\|^r \leq \Delta_p\left(x^+, z_{n,0}\right) - \Delta_p\left(x^+, z_{n,k_n}\right) = \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right),$$

which implies (4.12) and (4.10). Using now (4.10), we conclude that

$$\Delta_p\left(x^+, x_{n+1}\right) \leq \Delta_p\left(x^+, x_n\right) \leq ... \leq \Delta_p\left(x^+, x_0\right) < \rho,$$

which means that $x_{n+1} \in B_\rho\left(x^+, \Delta_p\right) \subset \mathrm{D}(F)$ and that (4.11) is true (see (2.20)). It remains to prove that the outer iteration terminates. Define the set $I := \left\{n \in \mathbb{N}_0 : \left\|b_n^\delta\right\| > \tau\delta_{[n]}\right\}$ and let $\overline{n}$ represent the largest number in I (which in principle could possibly be ∞). For any $n \in I$ the inequality $\left\|b_n^\delta\right\|^{t+r} \geq (\tau\delta_{\min})^{t+r}$ holds, where $\delta_{\min} := \min\{\delta_j : j = 0, ..., d-1\} > 0$. Now, as $k_n \geq 1$ for all $n \in I$ and $k_n = 0$ otherwise, we obtain from (4.12),

$$\begin{aligned}\sum_{n\in I}(\mu\tau\delta_{\min})^{t+r} &\leq \sum_{n\in I}\left(\mu\left\|b_n^\delta\right\|\right)^{t+r} k_n \leq \sum_{n=0}^{\overline{n}}\left(\mu\left\|b_n^\delta\right\|\right)^{t+r} k_n \\ &\lesssim \sum_{n=0}^{\overline{n}}\sum_{k=0}^{k_n-1}\lambda_{n,k}\left\|A_n s_{n,k} - b_n^\delta\right\|^r \\ &\lesssim \sum_{n=0}^{\overline{n}}\left(\Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right)\right) \leq \Delta_p\left(x^+, x_0\right) < \infty,\end{aligned}$$

which can only be true if $\overline{n} < \infty$. Then $N(\delta) = \overline{n} + 1 < \infty$.

We turn now to the proof in the case $\widehat{z}_{n,k} = x_n$. Observe first that $\tau > (1+\eta)/(1-\eta)$. Now, as $C = 0$ in Assumption 2, $0 < K_1 \leq 1$ and $K_0 \geq 0$, we conclude that $\mu_{\min}$ is larger than that one in Proposition 34, which implies that the results of this proposition are valid here and therefore k_n is finite. Using $k = k_n - 1$ in (4.14),

$$\begin{aligned}\Delta_p\left(x^+, x_{n+1}\right) - \Delta_p\left(x^+, x_n\right) &= \Delta_p\left(x^+, z_{n,k_n}\right) - \Delta_p\left(x^+, \widehat{z}_{n,k}\right) \\ &\leq -C_0\lambda_{n,k_n-1}\left\|A_n s_{n,k_n-1} - b_n^\delta\right\|^r,\end{aligned}$$

which is (4.13). Consequently, $x_{n+1} \in B_\rho\left(x^+, \Delta_p\right)$ and (4.10) as well as (4.11) is true. Defining the set I and the number $\delta_{\min}$ as before,

$$\begin{aligned}\sum_{n\in I}\mu^{t+r}(\tau\delta_{\min})^{t+r} &\leq \sum_{n\in I}\left(\mu\left\|b_n^\delta\right\|\right)^{t+r} \leq \sum_{n\in I}\left\|A_n s_{n,k_n-1} - b_n^\delta\right\|^{t+r} \\ &\lesssim \sum_{n=0}^{\overline{n}}\lambda_{n,k_n-1}\left\|A_n s_{n,k_n-1} - b_n^\delta\right\|^r \\ &\lesssim \sum_{n=0}^{\overline{n}}\left(\Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right)\right) \leq \Delta_p\left(x^+, x_0\right) < \infty.\end{aligned}$$

As before, we conclude that $N(\delta) < \infty$. ■

Remark 39 *Assuming the hypotheses of Theorem 38, the results of Corollaries 36 and 37 are clearly true. Furthermore, the monotonicity estimate* (4.10) *can be proven in a more general setting:*

$$\Delta_p(\vartheta_n, x_{n+1}) \leq \Delta_p(\vartheta_n, x_n), \tag{4.15}$$

where ϑ_n represents a solution of the $[n]$-th equation in $B_\rho(x^+, \Delta_p)$: $y_{[n]} = F_{[n]}(\vartheta_n)$. We remark further that, for those iterations satisfying $k_n = k_{REG}$ and where inequality (3.9) *is violated by x_n, we find following [21],*

$$\begin{aligned}\left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_{n+1})\right\| &\leq \left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n) - F'_{[n]}(x_n)(x_{n+1} - x_n)\right\| \\ &\qquad + \left\|F_{[n]}(x_{n+1}) - F_{[n]}(x_n) - F'_{[n]}(x_n)(x_{n+1} - x_n)\right\| \\ &\leq \mu\left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n)\right\| + \eta\left\|F_{[n]}(x_{n+1}) - F_{[n]}(x_n)\right\| \\ &\leq (\mu + \eta)\left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n)\right\| + \eta\left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_{n+1})\right\|,\end{aligned}$$

so that

$$\left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_{n+1})\right\| \leq \Lambda\left\|y_{[n]}^{\delta_{[n]}} - F_{[n]}(x_n)\right\| \quad \text{with } \Lambda := \frac{\mu + \eta}{1 - \eta}.$$

Finally, if η satisfies $\eta + K_0 < (1 - 2\eta)^r K_1$ (this is possible for a small η because $K_0 < K_1$) and

$$\tau > \frac{\eta + 1}{(1 - 2\eta)^r K_1 - (\eta + K_0)}$$

then $\mu_{\min}^r < (1 - 2\eta)^r$ and restricting μ to $(\mu_{\min}, 1 - 2\eta)$ yields $\Lambda < 1$. In particular, if $d = 1$ (only one equation in (3.6) *is considered),* (4.6) *and* (4.5) *are true.*

Remark 40 *Inequality* (4.14) *together with $\widehat{z}_{n,k} = z_{n,k}$ implies*

$$\Delta_p\left(x^+, z_{n,k+1}\right) \leq \Delta_p\left(x^+, z_{n,k}\right) \leq \ldots \leq \Delta_p\left(x^+, z_{n,0}\right) = \Delta_p\left(x^+, x_n\right) < \rho,$$

which is the monotonicity of the inner iteration. In view of (2.20),

$$\|z_{n,k}\| \leq C_{\rho,x^+}. \tag{4.16}$$

Thus the sequences $(z_{n,k})_{0 \leq k \leq k_n}$, $n \leq N(\delta)$, are uniformly bounded in n, k and δ.

The monotonicity in the inner iteration does not follow directly from (4.14) *if $\widehat{z}_{n,k} = x_n$, but the uniform bound* (4.16) *still holds in this case because*

$$\Delta_p\left(x^+, z_{n,k+1}\right) \leq \Delta_p\left(x^+, x_n\right) < \rho.$$

Weak convergence of `K-REGINN` follows immediately from (4.11) and the reflexivity of X.

Corollary 41 (Weak convergence) *Let all the assumptions of Theorem 38 hold true. If the operators F_j, $j = 0, \ldots, d-1$, are weakly sequentially closed then for any sequence $\left(y_j^{(\delta_j)_i}\right)_{i \in \mathbb{N}}$ with $\delta^{(i)} = \max\left\{(\delta_j)_i : j = 0, \ldots, d-1\right\} \to 0$ as $i \to \infty$, the sequence $\left(x_{N(\delta^{(i)})}\right)_{i \in \mathbb{N}}$ contains a subsequence that converges weakly to a solution of* (3.1) *in $B_\rho(x^+, \Delta_p)$. If x^+ is the unique solution of* (3.1) *in $B_\rho(x^+, \Delta_p)$, then $\left(x_{N(\delta)}\right)_{\delta > 0}$ converges weakly to x^+ as $\delta = \max\{\delta_j : j = 0, ..., d-1\} \to 0$.*

Proof. This is a standard proof. See, e.g. [36, Cor. 3.5]. ■

4.3 Convergence without noise

From now on, we need to clearly differ between the noisy ($\delta > 0$) and the noise-free ($\delta = 0$) situations. For this reason we exclusively mark quantities by a superscript δ when the data is corrupted by noise: x_n^δ, b_n^δ, A_n^δ, etc. Thus, x_n, b_n, A_n, etc. originate from exact data. Note that the starting guess is chosen independently of δ: $x_0^\delta = x_0$.

Algorithm 1 is well defined in the noiseless situation when we set $\delta_j = 0$, $\tau = \infty$, and $\tau\delta_j = 0$. Then, the discrepancy principle (3.9) is replaced by $\|b_n\| = 0$, in which case $x_{n+1} = x_n$. Termination only occurs in the unlikely event that an iterate x_N satisfies $\|y_j - F_j(x_N)\| = 0$ for $j = 0, \ldots, d-1$, i.e., x_N solves (3.6). In general, Algorithm 1 does not stop but produces a sequence which converges strongly to a solution of (3.1) as we will prove in this section, see Theorem 43 below.

Except for the termination, all results of Theorem 38 hold true with an even larger interval for the selection of the tolerances: $\mu \in \left(\frac{\eta+K_0}{K_1}, 1\right)$. Accordingly, the constant in (4.14) is replaced by $C_0 := K_1 - (\eta + K_0)/\mu^r > 0$. Further, $N(\delta) = \infty$ in case we have no premature termination.

In the next assumption, we define how the regularizing sequence is generated in the inner iteration.

Assumption 4 *If the iterate x_n of K-REGINN is well-defined, then the inner iteration sequence is generated by*

$$J_p(z_{n,k+1}) = J_p(\widehat{z}_{n,k}) - \lambda_{n,k}\left(A_n^* j_r(v_{n,k}) + \gamma_n J_p(s_{n,k} - \overline{x}_n)\right),$$

where $(v_{n,k})_{k \le k_n} \subset Y_{[n]}$ satisfies $\|v_{n,k}\| \le \|A_n s_{n,k} - b_n\|$. The sequence $(\gamma_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ obeys the inequalities $0 \le \gamma_n \le K_2 \|b_n\|^r$ with $K_2 \ge 0$. Further, $(\overline{x}_n)_{n\in\mathbb{N}} \subset X$ is an arbitrary sequence and $K_2 = 0$ whenever this sequence is unbounded or in case $\widehat{z}_{n,k} = x_n$.

The following methods satisfy the above assumption:

- All the dual gradient methods, defined by iteration (3.29) in Subsection 3.1.2 and satisfying $\lambda_{n,k} \le \lambda_{DE}$. In particular, the DE method itself, the MSD and LW methods. Here, $\widehat{z}_{n,k} = z_{n,k}$, $v_{n,k} = A_n s_{n,k} - b_n$ and $K_2 = 0$;
- The Bregman variation of the Iterated-Tikhonov method (3.62), with $\widehat{z}_{n,k} = z_{n,k}$, $v_{n,k} = A_n s_{n,k+1} - b_n$ and $K_2 = 0$;
- The Bregman variation of Tikhonov-Phillips method (3.58) with $\widehat{z}_{n,k} = x_n$, $v_{n,k} = A_n s_{n,k+1} - b_n$ (see (3.54)) and $K_2 = 0$;
- The mixed gradient-Tikhonov methods, defined by iteration (3.76), with $\widehat{z}_{n,k} = z_{n,k}$, $v_{n,k} = A_n s_{n,k} - b_n$, $\gamma_n = \alpha_n$ and $K_2 = K_0/K_3$, where K_0 is defined in Assumption 3, page 58, and K_3 is an upper bound to the sequence $\left(\frac{1}{p}\|e_n - \overline{x}_n\|^p\right)_{n\in\mathbb{N}}$ (see Subsection 3.1.4).

With the next lemma we prepare our convergence proof for the exact data case.

Lemma 42 *Assume all the hypotheses from Theorem 38 but with $\mu \in \left(\frac{\eta+K_0}{K_1}, 1\right)$ and let Assumption 4 hold true. Then,*

$$\Delta_p(x_n, x_{n+1}) \lesssim \Delta_p(x^+, x_n) - \Delta_p(x^+, x_{n+1}) \tag{4.17}$$

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

Proof. From (2.21),

$$\Delta_p(x_n, x_{n+1}) = \Delta_p\left(x^+, x_{n+1}\right) - \Delta_p\left(x^+, x_n\right) + \left\langle J_p(x_{n+1}) - J_p(x_n), x^+ - x_n\right\rangle. \quad (4.18)$$

We first analyze the case $\widehat{z}_{n,k} = z_{n,k}$. From Assumption 4,

$$\begin{aligned}\left\langle J_p(x_{n+1}) - J_p(x_n), x^+ - x_n\right\rangle &= \langle J_p(z_{n,k_n}) - J_p(z_{n,0}), e_n\rangle \\ &= \sum_{k=0}^{k_n-1} \langle J_p(z_{n,k+1}) - J_p(z_{n,k}), e_n\rangle \\ &= \sum_{k=0}^{k_n-1} \lambda_{n,k}\left(\langle j_r(v_{n,k}), -A_n e_n\rangle + \gamma_n \langle J_p(s_{n,k} - \overline{x}_n), -e_n\rangle\right).\end{aligned}$$

If $(\overline{x}_n)_{n\in\mathbb{N}}$ is not bounded, then $\gamma_n \equiv 0$ ($K_2 = 0$), otherwise, since the sequences (e_n) and $(s_{n,k})$ are uniformly bounded, see (4.11) and (4.16),

$$\gamma_n \langle J_p(s_{n,k} - \overline{x}_n), -e_n\rangle \le \gamma_n \|s_{n,k} - \overline{x}_n\|^{p-1} \|e_n\| \lesssim \gamma_n \lesssim \|b_n\|^r. \quad (4.19)$$

From properties of j_r, the TCC (Assumption 1(c)) and $\|b_n\| \le \frac{1}{\mu}\|b_n - A_n s_{n,k}\|$ for $k = 0, ..., k_n - 1$, we obtain by making use of (4.12),

$$\begin{aligned}\left\langle J_p(x_{n+1}) - J_p(x_n), x^+ - x_n\right\rangle &\lesssim \sum_{k=0}^{k_n-1} \lambda_{n,k}\left(\|A_n s_{n,k} - b_n\|^{r-1}\|A_n e_n\| + \|b_n\|^r\right) \\ &\le \sum_{k=0}^{k_n-1} \lambda_{n,k}\left(\|A_n s_{n,k} - b_n\|^{r-1}(\eta+1)\|b_n\| + \|b_n\|^r\right) \\ &\le \left(\frac{\eta+1}{\mu} + \frac{1}{\mu^r}\right)\sum_{k=0}^{k_n-1} \lambda_{n,k}\|A_n s_{n,k} - b_n\|^r \\ &\lesssim \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right).\end{aligned}$$

Inserting this result in (4.18), we arrive at (4.17). We proceed now proving the result for the case $\widehat{z}_{n,k} = x_n$ (consequently, $K_2 = 0$). Similarly to above, but using (4.13) instead of (4.12), the inequality

$$\begin{aligned}\left\langle J_p(x_{n+1}) - J_p(x_n), x^+ - x_n\right\rangle &= \langle J_p(z_{n,k_n}) - J_p(x_n), e_n\rangle \\ &= -\lambda_{n,k_n-1}\langle j_r(v_{n,k_n-1}), A_n e_n\rangle \\ &\le \frac{\eta+1}{\mu}\lambda_{n,k_n-1}\|A_n s_{n,k_n-1} - b_n\|^r \\ &\lesssim \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right),\end{aligned}$$

is achieved, which concludes the proof. ■

Assumptions 1 to 4 are strong enough to prove the convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ to a solution of (3.1) as $n \to \infty$ if a single equation in (3.6) is considered ($d = 1$). However, it is very difficult to prove the same result for the Kaczmarz version of `REGINN` ($d > 1$) using only these assumptions. The boundedness of k_n and of $\lambda_{n,k}$ seems to be necessary conditions to this end. For this reason we enunciate the next assumption.

Assumption 5 *If $s \le r$ and $k_{\max} < \infty$ in* (3.12), *then there exist constants $0 < \lambda_{\min} \le \lambda_{\max}$ such that $\lambda_{\min} \le \lambda_{n,k} \le \lambda_{\max}$ for all $k = 0, ..., k_n$ and $n \in \mathbb{N}$.*

The above assumption is verified for:

- The dual gradient method LW with a small constant step-size. This method clearly satisfies $\lambda_{\min} = \lambda_{LW} = \lambda_{\max}$;
- The Bregman variation of Iterated-Tikhonov method, where the required inequalities are satisfied for $\lambda_{n,k} = 1/\alpha_n$. Consequently, $\lambda_{\min} = 1/\alpha_{\max}$ and $\lambda_{\max} = 1/\alpha_{\min}$ (see (3.68));
- The Bregman variation of Tikhonov-Phillips method with $\lambda_{n,k} = 1/\alpha_k$. In this case, $\lambda_{\min} = 1/\alpha_0$ and $\lambda_{\max} = 1/\alpha_{k_{\max}}$ (see (3.57)).

It is possible to prove the existence of $\lambda_{\min}$ for the dual gradient methods DE and MSD (see (3.43) and (3.47)). But unfortunately, this seems not to be the case for $\lambda_{\max}$. However, if we modify the definitions of these methods to $\lambda_{new} = \lambda_{old} \wedge \overline{\lambda}$ with a (possibly very large) constant $\overline{\lambda} \geq \lambda_{\min}$, Assumption 5 is immediately verified for these methods with $\lambda_{\max} := \overline{\lambda}$. Anyway, some results of next Theorem hold true for these methods in their original definition. In the following convergence proof we adapt ideas from our previous work [40].

Theorem 43 (Convergence without noise) *Let X and Y be Banach spaces with X being uniformly smooth and $s-$convex with $p \leq s$. Let Assumption 1 in page 31 hold true and start with $x_0 \in B_\rho(x^+, \Delta_p)$. Assume that Assumptions 3, page 58 and 4, page 62 hold true and that Assumption 2, page 53 is verified with $C = 0$ in case of $\widehat{z}_{n,k} = x_n$. Suppose that the constant η in the TCC satisfy $\eta < K_1 - K_0$ and that $\mu \in (\mu_{\min}, 1)$ with $\mu_{\min} := (\eta + K_0)/K_1 < 1$. If $d = 1$, then `REGINN` either stops after finitely many iterations with a solution of* (3.1) *or the generated sequence $(x_n)_{n\in\mathbb{N}} \subset B_\rho(x^+, \Delta_p)$ converges strongly in X to a solution of* (3.1). *If x^+ is the unique solution in $B_\rho(x^+, \Delta_p)$, then $x_n \to x^+$ as $n \to \infty$.*

If $d > 1$ then the same results hold true for `K-REGINN` in case of Assumption 5 holds true as well as $k_{\max} < \infty$ and $s \leq r$.

Proof. As this proof is relatively long, it is divided in four parts. In the first and second parts, the case $d > 1$ is analyzed and Assumption 5 is employed to prove inequality (4.31) below, for the case $\widehat{z}_{n,k} = z_{n,k}$ in the first part of the proof and for $\widehat{z}_{n,k} = x_n$ in the second one. In the third part, an inequality similar to (4.31) is established for the case $d = 1$. Finally, we prove in the fourth part of the proof that the residual converges to zero as n goes to infinity, which together with (4.31) will ensure the desired result.

If Algorithm 1 stops after a finite number of iterations then the current iterate is a solution of (3.6) and consequently of (3.1). Otherwise, $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, as we will prove now. Let $m, l \in \mathbb{N}$ with $m \leq l$. The trick of the whole proof is to use the triangle inequality in order to "cut" the norm $\|x_m - x_l\|$ in an appropriate vector x_z and then estimate each of the resulting norms using the properties of this vector.

Part 1: We consider first the case $d > 1$. In this case Assumption 5 holds true as well as $k_{\max} < \infty$ and $s \leq r$. Write $m = m_0 d + m_1$ and $l = l_0 d + l_1$ with $m_0, l_0 \in \mathbb{N}$ and $m_1, l_1 \in \{0, \ldots, d-1\}$. Of course $m_0 \leq l_0$. Choose $z_0 \in \{m_0, \ldots, l_0\}$ such that

$$\sum_{j=0}^{d-1} \left(\|y_j - F_j(x_{z_0 d+j})\| + \|x_{z_0 d+j+1} - x_{z_0 d+j}\|\right) \tag{4.20}$$
$$\leq \sum_{j=0}^{d-1} \left(\|y_j - F_j(x_{n_0 d+j})\| + \|x_{n_0 d+j+1} - x_{n_0 d+j}\|\right)$$

for all $n_0 \in \{m_0, \ldots, l_0\}$. Define $z := z_0 d + z_1$, where $z_1 = l_1$ if $z_0 = l_0$ and $z_1 = d-1$ otherwise. This setting guarantees $m \leq z \leq l$. As X is $s-$convex and the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded (see (4.11)), it follows from (2.23) that

$$\|x_m - x_l\|^s \leq 2^{s-1}\left(\|x_m - x_z\|^s + \|x_z - x_l\|^s\right) \lesssim \Delta_p(x_z, x_m) + \Delta_p(x_z, x_l).$$

Three points identity (2.21) implies now that

$$\|x_m - x_l\|^s \lesssim \beta_{m,z} + \beta_{l,z} + f(z,m,l) \tag{4.21}$$

with

$$\beta_{m,z} := \Delta_p\left(x^+, x_m\right) - \Delta_p\left(x^+, x_z\right)$$

and

$$f(z,m,l) := \left|\left\langle J_p(x_z) - J_p(x_m), x_z - x^+\right\rangle\right| + \left|\left\langle J_p(x_z) - J_p(x_l), x_z - x^+\right\rangle\right|.$$

By monotonicity (4.10), we conclude that $\Delta_p(x^+, x_n) \to \gamma \geq 0$ as $n \to \infty$. Thus, $\beta_{m,z}$ and $\beta_{l,z}$ converge to zero as $m \to \infty$ (which causes $z \to \infty$ and $l \to \infty$). Further,

$$f(z,m,l) \leq \sum_{n=m}^{l-1} \left|\left\langle J_p(x_{n+1}) - J_p(x_n), x_z - x^+\right\rangle\right|. \tag{4.22}$$

We want to show now that $f(z,m,l)$ is bounded from above by a term that converges to zero as $m \to \infty$, which together with (4.21) will prove that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Our first goal is to prove this property for the case $\widehat{z}_{n,k} = z_{n,k}$. As in the proof of Lemma 42,

$$\begin{aligned} &\left|\left\langle J_p(x_{n+1}) - J_p(x_n), x_z - x^+\right\rangle\right| \\ &\qquad\leq \sum_{k=0}^{k_n-1} \lambda_{n,k}\left(\|A_n s_{n,k} - b_n\|^{r-1}\|A_n e_z\| + \gamma_n\left|\left\langle J_p(s_{n,k} - \overline{x}_n), e_z\right\rangle\right|\right). \end{aligned} \tag{4.23}$$

Similarly to (4.19),

$$\gamma_n\left|\left\langle J_p(s_{n,k} - \overline{x}_n), e_z\right\rangle\right| \lesssim \|b_n\|^r. \tag{4.24}$$

We proceed applying Assumption 1(c) to estimate $\|A_n e_z\|$:

$$\begin{aligned} \|A_n e_z\| &\leq \left\|A_n\left(x^+ - x_n\right)\right\| + \|A_n(x_z - x_n)\| \\ &\leq \|b_n\| + \left\|b_n - A_n\left(x^+ - x_n\right)\right\| + \left\|F_{[n]}(x_z) - F_{[n]}(x_n)\right\| \\ &\quad + \left\|F_{[n]}(x_z) - F_{[n]}(x_n) - F'_{[n]}(x_n)(x_z - x_n)\right\| \\ &\leq (\eta+1)\left(\|b_n\| + \left\|F_{[n]}(x_z) - F_{[n]}(x_n)\right\|\right) \\ &\leq (\eta+1)\left(2\|b_n\| + \left\|y_{[n]} - F_{[n]}(x_z)\right\|\right). \end{aligned} \tag{4.25}$$

Note that in the last norm, the operator $F_{[n]}$ is applied in the "wrong" vector x_z. To estimate this norm, we use Assumption 1. Write $n = n_0 d + n_1$ with $n_0 \in \{m_0, \ldots, l_0\}$ and $n_1 \in \{0, \ldots, d-1\}$. Then,

$$\begin{aligned} \left\|y_{[n]} - F_{[n]}(x_z)\right\| &= \|y_{n_1} - F_{n_1}(x_{z_0 d + z_1})\| \\ &\leq \|y_{n_1} - F_{n_1}(x_{z_0 d + n_1})\| + \sum_{j=0}^{d-1}\|F_{n_1}(x_{z_0 d + j+1}) - F_{n_1}(x_{z_0 d + j})\| \\ &\leq \|y_{n_1} - F_{n_1}(x_{z_0 d + n_1})\| + \frac{1}{1-\eta}\sum_{j=0}^{d-1}\left\|F'_{n_1}(x_{z_0 d + j})(x_{z_0 d + j+1} - x_{z_0 d + j})\right\| \\ &\leq \left(1 + \frac{M}{1-\eta}\right)\sum_{j=0}^{d-1}\left(\|y_j - F_j(x_{z_0 d + j})\| + \|x_{z_0 d + j+1} - x_{z_0 d + j}\|\right). \end{aligned}$$

This inequality was the motivation for the definition of z. From (4.20),

$$\left\| y_{[n]} - F_{[n]}\left(x_z\right) \right\| \leq \left(1 + \frac{M}{1-\eta}\right) \sum_{j=0}^{d-1} \left(\left\| y_j - F_j\left(x_{n_0 d + j}\right) \right\| + \left\| x_{n_0 d+j+1} - x_{n_0 d+j} \right\| \right). \quad (4.26)$$

Inserting (4.26) in (4.25), (4.25) and (4.24) in (4.23), (4.23) in (4.22), and using the definition of k_n, we arrive at

$$f\left(z,m,l\right) \lesssim \sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \left\| A_n s_{n,k} - b_n \right\|^r + g\left(z,m,l\right) + h\left(z,m,l\right), \quad (4.27)$$

where

$$g\left(z,m,l\right) := \sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \left\| A_n s_{n,k} - b_n \right\|^{r-1} \sum_{j=0}^{d-1} \left\| y_j - F_j\left(x_{n_0 d+j}\right) \right\|,$$

$$h\left(z,m,l\right) := \sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \left\| A_n s_{n,k} - b_n \right\|^{r-1} \sum_{j=0}^{d-1} \left\| x_{n_0 d+j+1} - x_{n_0 d+j} \right\|.$$

The first term on the right-hand side of (4.27) can be estimated by (4.12). It remains to estimate g and h. At this point we need the boundedness of k_n. From $k_n \leq k_{\max}$,

$$\sum_{k=0}^{k_n-1} \left\| A_n s_{n,k} - b_n \right\|^v \lesssim \left(\sum_{k=0}^{k_n-1} \left\| A_n s_{n,k} - b_n \right\| \right)^v \lesssim \sum_{k=0}^{k_n-1} \left\| A_n s_{n,k} - b_n \right\|^v$$

for any $v > 0$. Then defining

$$w_n := \sum_{k=0}^{k_n-1} \left\| A_n s_{n,k} - b_n \right\|,$$

and making use of the inequality[2]

$$\sum_{i=0}^{d-1} a_i^{r-1} \sum_{j=0}^{d-1} b_j \leq d^2 \left(\sum_{i=0}^{d-1} a_i^r + \sum_{j=0}^{d-1} b_j^r \right) \text{ for every } a_i, b_j \geq 0,$$

we find applying Assumption 5 and definition of k_n (3.11),

$$\begin{aligned} g\left(z,m,l\right) &\lesssim \lambda_{\max} \sum_{n=m}^{l-1} w_n^{r-1} \sum_{j=0}^{d-1} \left\| y_j - F_j\left(x_{n_0 d+j}\right) \right\| \qquad (4.28) \\ &\leq \lambda_{\max} \sum_{n_0=m_0}^{l_0} \left(\sum_{n_1=0}^{d-1} w_{n_0 d+n_1}^{r-1} \sum_{j=0}^{d-1} \left\| y_j - F_j\left(x_{n_0 d+j}\right) \right\| \right) \\ &\lesssim \lambda_{\max} \sum_{n_0=m_0}^{l_0} \left(\sum_{n_1=0}^{d-1} w_{n_0 d+n_1}^{r} + \sum_{n_1=0}^{d-1} \left\| y_{n_1} - F_j\left(x_{n_0 d+n_1}\right) \right\|^r \right) \\ &\leq \lambda_{\max} \sum_{n=m_0}^{l_0 d+d-1} w_n^r + \sum_{n=m_0}^{l_0 d+d-1} \left\| b_n \right\|^r \\ &\lesssim \lambda_{\max} \sum_{n=m_0}^{l_0 d+d-1} \sum_{k=0}^{k_n-1} \left\| A_n s_{n,k} - b_n \right\|^r \leq \frac{\lambda_{\max}}{\lambda_{\min}} \sum_{n=m_0}^{l_0 d+d-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \left\| A_n s_{n,k} - b_n \right\|^r . \end{aligned}$$

[2] $a_i^{r-1} b_j \leq a_i^r$ if $b_j \leq a_i$, and $a_i^{r-1} b_j \leq b_j^r$ otherwise. Thus, $a_i^{r-1} b_j \leq a_i^r + b_j^r \leq \sum_{i=0}^{d-1} a_i^r + \sum_{j=0}^{d-1} b_j^r$ for $i, j = 0, ..., d-1$ and consequently $\sum_{i=0}^{d-1} a_i^{r-1} \sum_{j=0}^{d-1} b_j \leq d^2 \left(\sum_{i=0}^{d-1} a_i^r + \sum_{j=0}^{d-1} b_j^r \right)$.

Similarly,

$$h(z,m,l) \lesssim \sum_{n=m_0}^{l_0 d+d-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \|A_n s_{n,k} - b_n\|^r + \sum_{n=m_0}^{l_0 d+d-1} \|x_{n+1} - x_n\|^r . \tag{4.29}$$

Making use of the $s-$convexity of X once again,

$$\|x_{n+1} - x_n\|^s \lesssim \Delta_p(x_n, x_{n+1}) \overset{(4.17)}{\lesssim} \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right). \tag{4.30}$$

As $r \geq s$, we have for m, l large enough that

$$\begin{aligned}\sum_{n=m_0}^{l_0 d+d-1} \|x_{n+1} - x_n\|^r \leq & \sum_{n=m_0}^{l_0 d+d-1} \|x_{n+1} - x_n\|^s \\ & \lesssim \Delta_p\left(x^+, x_{m_0}\right) - \Delta_p\left(x^+, x_{l_0 d+d}\right) = \beta_{m_0, l_0 d+d}.\end{aligned}$$

Plugging this bound into (4.29), inserting then inequalities (4.29) and (4.28) in (4.27), (4.27) in (4.21), and using (4.12), we end up with

$$\|x_m - x_l\|^s \lesssim \beta_{m,z} + \beta_{l,z} + \beta_{m_0, l_0 d+d}. \tag{4.31}$$

Part 2: Now we consider the case $\widehat{z}_{n,k} = x_n$ (in this case $K_2 = 0$ and consequently $\gamma_n \equiv 0$ in Assumption 4) together with $d > 1$. Inequality (4.23) reads here

$$\left|\left\langle J_p(x_{n+1}) - J_p(x_n), x_z - x^+\right\rangle\right| \leq \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^{r-1} \|A_n e_z\| . \tag{4.32}$$

Proceeding like above, we find, similarly to (4.27),

$$f(z,m,l) \lesssim \sum_{n=m}^{l-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^r + g(z,m,l) + h(z,m,l)$$

with

$$\begin{aligned} g(z,m,l) &:= \sum_{n=m}^{l-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^{r-1} \sum_{j=0}^{d-1} \|y_j - F_j(x_{n_0 d+j})\| , \\ h(z,m,l) &:= \sum_{n=m}^{l-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^{r-1} \sum_{j=0}^{d-1} \|x_{n_0 d+j+1} - x_{n_0 d+j}\| . \end{aligned}$$

In the same way as done in (4.28) and (4.29) above, the inequalities

$$\begin{aligned} g(z,m,l) &\lesssim \sum_{n=m_0}^{l_0 d+l-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^r , \\ h(z,m,l) &\lesssim \sum_{n=m_0}^{l_0 d+l-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^r + \sum_{n=m_0}^{l_0} \|x_{n+1} - x_n\|^r \end{aligned}$$

can be proven. Using now (4.13) instead of (4.12) we arrive again at (4.31).

Part 3: The case $d = 1$ is considered now (therefore $k_{\max} = \infty$ is possible). This situation is easier because everything we need is to change the definition of x_z in (4.20) to the vector with the smallest residual in the outer iteration, i.e., choose $z \in \{m, \ldots, l\}$ such that $\|b_z\| \leq \|b_n\|$, for all $n \in \{m, \ldots, l\}$. Then, from (4.25),

$$\|A_n e_z\| \leq (\eta + 1)(2\|b_n\| + \|b_z\|) \leq 3(\eta + 1)\|b_n\| .$$

The inequality $\|b_n\| \leq \frac{1}{\mu}\|A_n s_{n,k} - b_n\|$ for $k = 0, ..., k_n - 1$, together with (4.32) (respectively (4.23) and (4.24)) and (4.22), results in

$$f(z, m, l) \lesssim \sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \|A_n s_{n,k} - b_n\|^r$$

in the case $\widehat{z}_{n,k} = z_{n,k}$ and

$$f(z, m, l) \lesssim \sum_{n=m}^{l-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^r$$

in the case $\widehat{z}_{n,k} = x_n$. Plugging now these inequalities into (4.21) and using again (4.12) (respectively (4.13)), we arrive at (4.31) with m_0 and $l_0 d + d$ replaced by m and l, respectively.

Part 4: In any case the right-hand side of inequality (4.31) converges to zero as $m \to \infty$ revealing $(x_n)_{n\in\mathbb{N}}$ to be a Cauchy sequence. As X is complete, it converges to some $x_\infty \in X$. Observe that $k_n \geq 1$ if $\|b_n\| \neq 0$ and since $\mu\|b_n\| \leq \|A_n s_{n,k} - b_n\|$ for all $k \leq k_n - 1$, it follows that for the case $\widehat{z}_{n,k} = z_{n,k}$,

$$\mu^{r+t}\sum_{n=0}^{\infty} \|b_n\|^{r+t} \leq \sum_{n=0}^{\infty} k_n (\mu\|b_n\|)^{r+t} \lesssim \sum_{n=0}^{\infty}\sum_{k=0}^{k_n-1} \lambda_{n,k}\|A_n s_{n,k} - b_n\|^r \overset{(4.12)}{<} \infty$$

Similarly, $\sum_{n=0}^{\infty} \|b_n\|^{r+t} < \infty$ for $\widehat{z}_{n,k} = x_n$. Then, $\left\|y_{[n]} - F_{[n]}(x_n)\right\| = \|b_n\| \to 0$ as $n \to \infty$ and since the F_j's are continuous for all $j = 0, \ldots, d-1$, we have $y_j = F_j(x_\infty)$. If (3.1) has only one solution in $B_\rho(x^+, \Delta_p)$ then $x_\infty = x^+$. ■

4.4 Regularization property

In this section we validate that, under appropriate conditions, K-REGINN is a regularization scheme for solving (3.1) with noisy data y^δ. Indeed, we show that the family $(x^\delta_{N(\delta)})_{\delta>0}$ of outputs of Algorithm 1 relative to the inputs $(y^\delta)_{\delta>0}$ converges strongly to solutions of (3.1) with exact data y as $\delta \to 0$.

To avoid possible wrong interpretations, we will not use the notation δ_j, $j = 0, \ldots, d-1$, as in (3.7) any more. Instead, when we write δ_i, we mean a positive number in a sequence of δ's as defined in (3.8), i.e., $\delta_i := \max\left\{(\delta_j)_i : j = 0, \ldots, d-1\right\} > 0$.

In order to prove the regularization property, a standard argument of three steps is frequently used. Firstly, one proves the monotonicity of the error: $\left\|x_n^\delta - x^+\right\| \leq \left\|x_{n-1}^\delta - x^+\right\|$ for $n = 1, ..., N(\delta)$. Secondly, the convergence in the noiseless situation is proven: for $\epsilon > 0$ fixed, $\|x_n - x^+\| < \epsilon/2$ provided n large enough. Thirdly, a stability result is necessary: $\left\|x_n^\delta - x_n\right\| < \epsilon/2$ provided δ small enough. Now, disregarding the details of pathological cases, $N(\delta) \to \infty$ as $\delta \to 0$. Thus, for $n \in \mathbb{N}$ sufficiently large but fixed and δ small enough to guarantee $N(\delta) \geq n$, it follows that

$$\left\|x^\delta_{N(\delta)} - x^+\right\| \leq \left\|x_n^\delta - x^+\right\| \leq \left\|x_n^\delta - x_n\right\| + \|x_n - x^+\| < \epsilon.$$

In this subsection, we apply a similar reasoning with the necessary modifications to fit it in with our framework. We follow ideas from [36] and [40]. In order to facilitate the comprehension, we summarize the main results:

1. From (3.11), it follows that k_n is an arbitrary number less than or equal k_{REG}, which means that the sequence $(x_n)_{n\in\mathbb{N}}$, generated from a run of K-REGINN using

data without noise, can change if the sequence $(k_{\max,n})_{n\in\mathbb{N}}$ is chosen differently. In Definition 44 below, we observe all the sequences which are possibly generated from a run of `K-REGINN` in the noiseless situation and collect all the $n-$th iterates in the set $\mathcal{X}_n$.

2. In Lemma 45, we prove that if the sequences $\left(x_n^{\delta_i}\right)_{n\in\mathbb{N}}$, with $(\delta_i)_{i\in\mathbb{N}} \subset \mathbb{R}$ being a zero-sequence, are generated by different runs of `K-REGINN` with the different noise levels δ_i, then for each $n \in \mathbb{N}$ fixed, the sequence of the $n-$th iterates $\left(x_n^{\delta_i}\right)_{i\in\mathbb{N}}$ splits into convergent subsequences, all of which converge to elements of $\mathcal{X}_n$ ($x_n^\delta \to \mathcal{X}_n$ as $\delta \to 0$).

3. Further, in Lemma 46 it is proven that for each $\epsilon > 0$, there exists a $M \in \mathbb{N}$ such that, for each $n \geq M$, there exists an element $\xi_n \in \mathcal{X}_n$ and a solution x_∞ of (3.1) satisfying $\|\xi_n - x_\infty\| < \epsilon$ (the sets $\mathcal{X}_n$ converge uniformly to the set of solutions of (3.1)).

4. Finally, in Theorem 47, we provide a proof that the sequence $\left(x_{N(\delta_i)}^{\delta_i}\right)_{i\in\mathbb{N}}$ of the final iterates of `K-REGINN` (generated with the different noise levels δ_i, where $(\delta_i)_{i\in\mathbb{N}} \subset \mathbb{R}$ is again a zero-sequence) splits into convergent subsequences, all of which converge strongly to solutions of (3.1) as $i \to \infty$.

In a first step we investigate the stability of the scheme, i.e., we study the behavior of the n-th iterate x_n^δ as δ approaches zero. The sets $\mathcal{X}_n$ defined below play an important role, see the Item 1 above.

Definition 44 *Let $\mathcal{X}_0 := \{x_0\}$ and define $\mathcal{X}_{n+1}$ from $\mathcal{X}_n$ by the following procedure: for each $\xi \in \mathcal{X}_n$, change x_n in definition of `K-REGINN` with ξ and change the respective inner iterate $z_{n,k}$ with $\sigma_{n,k}$, resulting from using ξ instead of x_n. More precisely, define $\sigma_{n,0} = \sigma_{n,0}(\xi) := \xi$ and $\sigma_{n,k+1} = \sigma_{n,k+1}(\xi)$ as*

$$J_p\left(\sigma_{n,k+1}\right) := J_p\left(\widehat{\sigma}_{n,k}\right) - \lambda_{n,k}^{\xi}\left(F'_{[n]}\left(\xi\right)^* j_r\left(v_{n,k}^{\xi}\right) + \gamma_n^{\xi} J_p\left(\left(\sigma_{n,k} - \xi\right) - \overline{x}_n^{\xi}\right)\right),$$

where $\widehat{\sigma}_{n,k}$, $\lambda_{n,k}^{\xi}$, $v_{n,k}^{\xi}$, γ_n^{ξ} and $\overline{x}_n^{\xi}$ replace $\widehat{z}_{n,k}$, $\lambda_{n,k}$, $v_{n,k}$, γ_n and $\overline{x}_n$ in Assumption 4, page 62, respectively, when the vectors $z_{n,k}$ and x_n are replaced by $\sigma_{n,k}$ and ξ respectively.

Let $\widetilde{b}_n := y_{[n]} - F_{[n]}\left(\xi\right)$ and define $k_{REG}(\xi) := 0$ in case of $\widetilde{b}_n = 0$ and

$$k_{REG}\left(\xi\right) := \min\left\{k \in \mathbb{N} : \left\|F'_{[n]}\left(\xi\right)\left(\sigma_{n,k} - \xi\right) - \widetilde{b}_n\right\| < \mu\left\|\widetilde{b}_n\right\|\right\} \tag{4.33}$$

otherwise. Then $\sigma_{n,k}(\xi) \in \mathcal{X}_{n+1}$ for $k = 1, \ldots, k_{REG}(\xi)$ in case $k_{REG}\left(\xi\right) \geq 1$ and only for $k = 0$ in case $k_{REG}\left(\xi\right) = 0$. We call $\xi \in \mathcal{X}_n$ the PREDECESSOR *of the vectors $\sigma_{n,k}(\xi) \in \mathcal{X}_{n+1}$ and these ones* SUCCESSORS *of ξ.*

Note that x_n is a possible outer iterate of `K-REGINN` if and only if $x_n \in \mathcal{X}_n$. Moreover, $\mathcal{X}_n$ is finite for each $n \in \mathbb{N}$ and from inequality (4.15) follows that, if $\xi_{n+1} \in \mathcal{X}_{n+1}$ is as successor of $\xi_n \in \mathcal{X}_n$,

$$\Delta_p\left(x^\dagger, \xi_{n+1}\right) \leq \Delta_p\left(x^\dagger, \xi_n\right)$$

whenever $x^\dagger$ is a solution of (3.1) in $B_\rho\left(x^+, \Delta_p\right)$. We emphasize that the sets $\mathcal{X}_n$, $n \in \mathbb{N}_0$, are defined with respect to exact data y.

The following stability property of inner iteration of `K-REGINN` is crucial to prove convergence with noisy data. This assumption is a variation of [36, Assumption (3.12)].

Assumption 6 *Let $(\delta_i)_{i\in N} \subset \mathbb{R}$ be a zero-sequence. Assume that the sequence $\left(z_{n,k}^{\delta_i}\right)_{i\in\mathbb{N}}$ converges to $\sigma_{n,k}(\xi)$ for $k = 0, ..., k_{REG}(\xi)$ whenever the sequence $\left(x_n^{\delta_i}\right)_{i\in\mathbb{N}}$ converges to $\xi \in \mathcal{X}_n$. This is,*

$$\lim_{i\to\infty} x_n^{\delta_i} = \xi \in \mathcal{X}_n \Longrightarrow \lim_{i\to\infty} z_{n,k}^{\delta_i} = \sigma_{n,k}(\xi) \text{ for } k = 0, ..., k_{REG}(\xi).$$

Assumption 6 is verified for the following methods:

- All the dual gradient methods, defined by iteration (3.29) in Subsection 3.1.2 and satisfying $\lambda_{n,k}^{\delta} \leq \lambda_{DE}^{\delta}$, where $\lambda_{n,k}^{\delta}$ depends continuously on δ. In particular, the DE method itself, the MSD and LW methods. For these methods, we assume that X is uniformly smooth and $s-$convex with $p \leq s$. Additionally, we suppose that the spaces Y_j, $j = 0, ..., d-1$ are uniformly smooth;
- The Bregman variation of the Iterated-Tikhonov method (3.62), assuming that X is uniformly smooth and $s-$convex with $p \leq s \leq r$;
- The Bregman variation of Tikhonov-Phillips method (3.58). We assume that X is uniformly smooth and uniformly convex;
- The mixed gradient-Tikhonov methods, defined by iteration (3.76). Here the uniform smoothness and the $s-$convexity of X, with $p \leq s$, as well as the uniform smoothness of Y_j, $j = 0, ..., d-1$ is assumed.

As these proofs are long, we transfer them to Appendix A.

Based on Assumption 6, we prove the next lemma, which basically adapts ideas of [36, Lemma 3.11].

Lemma 45 *Let all assumptions of Theorem 38 hold true and assume additionally Assumption 6 and that X is $s-$convex with $p \leq s$. If $\delta_i \to 0$ as $i \to \infty$ then for $n \leq N(\delta_i)$ with $\delta_i > 0$ sufficiently small, the sequence $\left(x_n^{\delta_i}\right)_{i\in\mathbb{N}}$ splits into convergent subsequences, all of which converge to elements of $\mathcal{X}_n$.*

Proof. We prove the statement by induction. For $n = 0$, $x_0^{\delta_i} = x_0 \to x_0 \in \mathcal{X}_0$ as $i \to \infty$. Now, suppose that for some $n \in \mathbb{N}$ with $n < N(\delta_i)$ for i large enough, $\left(x_n^{\delta_i}\right)_{i\in\mathbb{N}}$ splits into convergent subsequences, all of which converge to elements of $\mathcal{X}_n$. To simplify the notation, let $\left(x_n^{\delta_i}\right)_{i\in\mathbb{N}}$ itself be a subsequence which converges to an element of $\mathcal{X}_n$, say

$$\lim_{i\to\infty} x_n^{\delta_i} = \xi \in \mathcal{X}_n. \tag{4.34}$$

We have to prove that the sequence $\left(x_{n+1}^{\delta_i}\right)_{i\in\mathbb{N}}$ splits into convergent subsequences, each one converging to an element of $\mathcal{X}_{n+1}$. From Assumption 6 and (4.34),

$$\lim_{i\to\infty} z_{n,k}^{\delta_i} = \sigma_{n,k}(\xi) \text{ for } k = 0, ..., k_{REG}(\xi). \tag{4.35}$$

As the functions F_j and F_j' are continuous for $j = 0, ..., d-1$, it follows from (3.7), (4.35) and (4.34) that

$$\lim_{i\to\infty} \left\| b_n^{\delta_i} \right\| = \left\| \widetilde{b}_n \right\| \text{ and} \tag{4.36}$$
$$\lim_{i\to\infty} \left\| F_{[n]}'\left(x_n^{\delta_i}\right) s_{n,k}^{\delta_i} - b_n^{\delta_i} \right\| = \left\| F_{[n]}'(\xi)(\sigma_{n,k} - \xi) - \widetilde{b}_n \right\|$$

for all $k \leq k_{REG}(\xi)$. Now, we have to differ two cases.

Case 1: $\widetilde{b}_n \neq 0$. From definition of $k_{REG}(\xi)$ (see (4.33)),

$$\left\| F'_{[n]}(\xi)\left(\sigma_{n,k_{REG}(\xi)} - \xi\right) - \widetilde{b}_n \right\| < \mu \left\| \widetilde{b}_n \right\|$$

and we conclude, making use of (4.36) that for i large enough

$$\left\| F'_{[n]}\left(x_n^{\delta_i}\right) s_{n,k_{REG}(\xi)}^{\delta_i} - b_n^{\delta_i} \right\| < \mu \left\| b_n^{\delta_i} \right\|,$$

which implies in view of (3.10) that $k_n^{\delta_i} \leq k_{REG}^{\delta_i} \leq k_{REG}(\xi)$. Hence $k_n^{\delta_i} \in \{0, \ldots, k_{REG}(\xi)\}$ for i large enough. This means that for i large enough, the sequence $\left(k_n^{\delta_i}\right)_{i\in\mathbb{N}}$ splits into at most $k_{REG}(\xi)+1$ constant subsequences $\left(k_n^{\delta_{i_\ell}}\right)_{\ell\in\mathbb{N}}$ satisfying $k_n^{\delta_{i_\ell}} = \overline{k} \in \{0, \ldots, k_{REG}(\xi)\}$. Accordingly,

$$\lim_{\ell\to\infty} x_{n+1}^{\delta_{i_\ell}} = \lim_{\ell\to\infty} z_{n,k_n^{\delta_{i_\ell}}}^{\delta_{i_\ell}} = \lim_{\ell\to\infty} z_{n,\overline{k}}^{\delta_{i_\ell}} \overset{(4.35)}{=} \sigma_{n,\overline{k}} \in \mathcal{X}_{n+1}.$$

Case 2: $\widetilde{b}_n = 0$. In this case, $y_{[n]} = F_{[n]}(\xi)$ and $\sigma_{n,0} = \xi \in \mathcal{X}_{n+1}$ (see Definition 44). As the sequence $\left(x_n^{\delta_i}\right)_{n\leq N(\delta_i)}$ is uniformly bounded (see (4.11)) and X is $s-$convex, we conclude that (2.23) applies. We prove now that $x_{n+1}^{\delta_i} \to \xi$ as $i \to \infty$. Assume the contrary, then there exist an $\epsilon > 0$ and a subsequence $(\delta_{i_\ell})_{\ell\in\mathbb{N}}$ such that $\epsilon < \left\| \xi - x_{n+1}^{\delta_{i_\ell}} \right\|^s$. It follows that

$$\epsilon < \left\| \xi - x_{n+1}^{\delta_{i_\ell}} \right\|^s \overset{(2.23)}{\leq} C\Delta_p\left(\xi, x_{n+1}^{\delta_{i_\ell}}\right) \overset{(4.15)}{\leq} C\Delta_p\left(\xi, x_n^{\delta_{i_\ell}}\right) \to C\Delta_p(\xi,\xi) = 0$$

as $\ell \to \infty$, contradicting $\epsilon > 0$. ■

In the second step towards establishing the regularization property we provide a kind of uniform convergence of the set sequence $(\mathcal{X}_n)_n$ to solutions of (3.1). For the rigorous formulation in Lemma 46 below we need to introduce further notation: Let $l \in \mathbb{N}$ and set $\xi_0^{(l)} := x_0$. Now define $\xi_{n+1}^{(l)} := \sigma_{n,k_n^{(l)}}(\xi_n^{(l)})$ by choosing $k_n^{(l)} \in \{1, \ldots, k_{REG}(\xi_n^{(l)})\}$ in case of $k_{REG}(\xi_n^{(l)}) \geq 1$ and $k_n^{(l)} = 0$ otherwise. Then $\xi_{n+1}^{(l)}$ is a successor of $\xi_n^{(l)}$. Of course $\xi_n^{(l)} \in \mathcal{X}_n$ for all $n \in \mathbb{N}$ and reciprocally, each element in $\mathcal{X}_n$ can be written as $\xi_n^{(l)}$ for some $l \in \mathbb{N}$.

Observe that $(\xi_n^{(l)})_{n\in\mathbb{N}}$ represents a sequence generated by `K-REGINN` in the case of exact data is given and with the inner iteration stopped with an arbitrary stop index $k_n^{(l)}$ less than or equal $k_{REG}(\xi_n^{(l)})$. Due to this fact, we call the sequence $(k_n^{(l)})_{n\in\mathbb{N}}$ a *stop rule*. The sequence $(\xi_n^{(l)})_{n\in\mathbb{N}}$ is therefore one of many possible sequences which can be generated from a run of `K-REGINN` with initial vector x_0. Since the results of Theorem 43 hold true for all these sequences, it applies in particular to $(\xi_n^{(l)})_{n\in\mathbb{N}}$. Thus, the limit

$$x_\infty^{(l)} := \lim_{n\to\infty} \xi_n^{(l)} \tag{4.37}$$

exists and is a solution of (3.1) in $B_\rho(x^+, \Delta_p)$.

The following result was first presented in [22] in the context of Hilbert spaces and later generalized to the Banach spaces framework in Proposition 19 of [40], where only the case of a unique solution of (3.1) was analyzed. The present version, appeared first in [39].

Lemma 46 *Let all assumptions of Theorem 43 hold true and let $(\xi_n^{(l)})_{n\in\mathbb{N}}$ denote the sequence generated by the stop rule $(k_n^{(l)})_{n\in\mathbb{N}}$. Then, for each $\epsilon > 0$ there exists an $M = M(\epsilon) \in \mathbb{N}$ such that*

$$\left\| \xi_n^{(l)} - x_\infty^{(l)} \right\| < \epsilon \text{ for all } n \geq M \text{ and all } l \in \mathbb{N}.$$

In particular, if x^+ is the unique solution of (3.1) in $B_\rho(x^+, \Delta_p)$ then $\|\xi_n^{(l)} - x^+\| < \epsilon$ for all $n \geq M$ and all $l \in \mathbb{N}$.

Proof. Assume the statement is not true. Then, there exist an $\epsilon > 0$ and sequences $(n_j)_j, (l_j)_j \subset \mathbb{N}$ with $(n_j)_j$ strictly increasing such that

$$\left\| \xi_{n_j}^{(l_j)} - x_\infty^{(l_j)} \right\| \geq \epsilon \text{ for all } j \in \mathbb{N}$$

where $(\xi_n^{(l_j)})_n$ represents the sequence generated by the stop rule $(k_n^{(l_j)})_n$. We stress the fact that the iterates $\xi_{n_j}^{(l_j)}$ must be generated by infinitely many different sequences of stop rules (otherwise, as $\xi_{n_j}^{(l)} \to x_\infty^{(l)}$ as $j \to \infty$ for each l and as the l_j's attain only a finite number of values, we would have $\|\xi_{n_j}^{(l_j)} - x_\infty^{(l_j)}\| < \epsilon$ for n_j large enough). Next we reorder the numbers l_j (excluding some iterates if necessary) such that

$$\left\| \xi_{n_l}^{(l)} - x_\infty^{(l)} \right\| \geq \epsilon \text{ for all } l \in \mathbb{N}. \tag{4.38}$$

Set $\widehat{\xi}_0 := x_0$ $(= \xi_0^{(l)}$ for all $l \in \mathbb{N})$. As $k_{REG}(\widehat{\xi}_0) < \infty$ and $k_0^{(l)} \in \{0, \ldots, k_{REG}(\xi_0^{(l)})\} = \{0, \ldots, k_{REG}(\widehat{\xi}_0)\}$ for all $l \in \mathbb{N}$, there exists a number $\widehat{k}_0$ in this set such that $\widehat{k}_0 = k_0^{(l)}$ for infinitely many $l \in \mathbb{N}$. Let $\mathcal{L}_0 \subset \mathbb{N}$ be the set of those indices l. Fix now $\widehat{k}_0$ as the stop index of the first inner iteration, i.e., $\widehat{\xi}_1 := \sigma_{0,\widehat{k}_0}(\widehat{\xi}_0)$, see Definition 44. Then $\widehat{\xi}_1 = \xi_1^{(l)}$ for all $l \in \mathcal{L}_0$ and as $k_{REG}(\widehat{\xi}_1) < \infty$, we conclude like before, that there exists a number $\widehat{k}_1 \in \{0, \ldots, k_{REG}(\widehat{\xi}_1)\}$ such that $\widehat{k}_1 = k_1^{(l)}$ for infinitely many $l \in \mathcal{L}_0 \backslash \{1\}$. Those l's are collected in $\mathcal{L}_1$. Proceeding by induction, we find a sequence $(\widehat{\xi}_n)_n$, generated by the stop rule $(\widehat{k}_n)_n$ as well as a sequence of unbounded sets $(\mathcal{L}_n)_n$ satisfying $\mathcal{L}_n \subset \mathbb{N} \backslash \{1, \ldots, n\}$ with $\mathcal{L}_{n+1} \subset \mathcal{L}_n$ and

$$\xi_{n+1}^{(l)} = \widehat{\xi}_{n+1} \text{ for all } l \in \mathcal{L}_n,\ n \in \mathbb{N}_0. \tag{4.39}$$

In view of (4.37) the limit $\widehat{x}_\infty := \lim_{n\to\infty} \widehat{\xi}_n$ exists in $B_\rho(x^+, \Delta_p)$ and solves (3.1). It follows that the sequence $(\widehat{\xi}_n)_n$ is bounded and since X is $s-$convex, inequality (2.23) holds. Additionally, there exists $M = M(\epsilon) \in \mathbb{N}$ such that

$$\Delta_p\left(\widehat{x}_\infty, \widehat{\xi}_n\right) < \frac{\epsilon^s}{2^s C} \text{ for all } n > M, \tag{4.40}$$

where $C > 0$ is the constant from (2.23). We can additionally suppose that $\widehat{\xi}_n \in B_\rho(x^+, \Delta_p)$ for all $n > M$. In fact, as $\lim_{n\to\infty} \widehat{\xi}_n = \widehat{x}_\infty$ and the mappings J_p and $\Delta_p(\widehat{x}_\infty, \cdot)$ are continuous, we have that $\Delta_p\left(\widehat{x}_\infty, \widehat{\xi}_n\right)$ and $\left\langle J_p(\widehat{\xi}_n) - J_p(\widehat{x}_\infty), \widehat{x}_\infty - x^+\right\rangle$ converge to zero as $n \to \infty$. From three points identity (2.21),

$$\Delta_p\left(x^+, \widehat{\xi}_n\right) = \Delta_p\left(\widehat{x}_\infty, \widehat{\xi}_n\right) + \Delta_p\left(x^+, \widehat{x}_\infty\right) + \left\langle J_p(\widehat{\xi}_n) - J_p(\widehat{x}_\infty), \widehat{x}_\infty - x^+\right\rangle$$

and as $\Delta_p(x^+, \widehat{x}_\infty) < \rho$, we conclude that $\Delta_p\left(x^+, \widehat{\xi}_n\right) < \rho$ for n large enough.

Now, for $l_0 \in \mathcal{L}_M$ fixed,

$$\Delta_p\left(\widehat{x}_\infty, \xi_{M+1}^{(l_0)}\right) \overset{(4.39)}{=} \Delta_p\left(\widehat{x}_\infty, \widehat{\xi}_{M+1}\right) \overset{(4.40)}{<} \frac{\epsilon^s}{2^s C}.$$

As $\widehat{x}_\infty$ is a solution of (3.1) and $\xi_{M+1}^{(l_0)} = \widehat{\xi}_{M+1} \in B_\rho(x^+, \Delta_p)$, inequality (4.10) applies and the errors $\Delta_p\left(\widehat{x}_\infty, \xi_n^{(l_0)}\right)$ are monotonically decreasing in n for all $n \geq M+1$. In particular, $n_{l_0} \geq l_0 \geq M+1$ (because $l_0 \in \mathcal{L}_M \subset \mathbb{N} \backslash \{1, \ldots, M\}$). Then

$$\Delta_p\left(\widehat{x}_\infty, \xi_{n_{l_0}}^{(l_0)}\right) \leq \Delta_p\left(\widehat{x}_\infty, \xi_{M+1}^{(l_0)}\right) < \frac{\epsilon^s}{2^s C}.$$

Since $\xi_n^{(l_0)} \to x_\infty^{(l_0)}$ as $n \to \infty$, we conclude that

$$\Delta_p\left(\widehat{x}_\infty, x_\infty^{(l_0)}\right) = \lim_{n\to\infty} \Delta_p\left(\widehat{x}_\infty, \xi_n^{(l_0)}\right) \overset{(4.10)}{\leq} \Delta_p\left(\widehat{x}_\infty, \xi_{n_{l_0}}^{(l_0)}\right) < \frac{\epsilon^s}{2^s C}.$$

From the $s-$convexity of X,

$$\begin{aligned}\left\|\xi_{n_{l_0}}^{(l_0)} - x_\infty^{(l_0)}\right\|^s &\leq 2^{s-1}\left(\left\|\xi_{n_{l_0}}^{(l_0)} - \widehat{x}_\infty\right\|^s + \left\|\widehat{x}_\infty - x_\infty^{(l_0)}\right\|^s\right) \\ &\leq 2^{s-1} C\left(\Delta_p\left(\widehat{x}_\infty, \xi_{n_{l_0}}^{(l_0)}\right) + \Delta_p\left(\widehat{x}_\infty, x_\infty^{(l_0)}\right)\right) < \epsilon^s,\end{aligned}$$

contradicting (4.38). ■

We are now in position to prove our main result. The result of the next theorem was first established in our work [39].

Theorem 47 (Regularization property) *Assume all the hypotheses of Theorem 38, Assumptions 4, page 62, and 6, page 70, and suppose that X is $s-$convex with $p \leq s$. If $d > 1$, assume additionally that Assumption 5, page 63, holds true, as well as $k_{\max} < \infty$ and $s \leq r$. If $\delta_i \to 0$ as $i \to \infty$, then the sequence $\left(x_{N(\delta_i)}^{\delta_i}\right)_{i\in\mathbb{N}}$ splits into convergent subsequences, all of which converge strongly to solutions of (3.1) as $i \to \infty$. In particular, if x^+ is the unique solution of (3.1) in $B_\rho(x^+, \Delta_p)$ then*

$$\lim_{i\to\infty}\left\|x_{N(\delta_i)}^{\delta_i} - x^+\right\| = 0.$$

Proof. If $N(\delta_i) \leq I$ as $i \to \infty$ for some $I \in \mathbb{N}$, then the sequence $(x_{N(\delta_i)}^{\delta_i})_{i\in\mathbb{N}}$ splits into subsequences of the form $(x_n^{\delta_{i_\ell}})_{\ell\in\mathbb{N}}$ where n is a fixed number less than or equal to I. According to Lemma 45, each of these subsequences splits into convergent subsequences. Hence each limit of such a subsequence must be a solution of (3.1) due to the discrepancy principle (3.13). In fact, if $x_n^{\delta_i} \to a$ as $i \to \infty$, then using (3.7),

$$\|y_j - F_j(a)\| = \lim_{i\to\infty}\left\|y_j - F_j\left(x_n^{\delta_i}\right)\right\| \leq \lim_{i\to\infty}(1+\tau)\delta_i = 0, \quad j = 0, \ldots, d-1.$$

Suppose now that $N(\delta_i) \to \infty$ as $i \to \infty$ and let $\epsilon > 0$ be given. As the Bregman distance is a continuous function in both arguments, there exists $\gamma = \gamma(\epsilon) > 0$ such that

$$\Delta_p\left(x, x_n^{\delta_i}\right) < \frac{\epsilon^s}{C} \text{ whenever } \left\|x - x_n^{\delta_i}\right\| < \gamma, \tag{4.41}$$

where $C > 0$ is the constant from (2.23). From Lemma 46, there is an $M \in \mathbb{N}$ such that, for each $\xi_M^{(l)} \in \mathcal{X}_M$, there exists a solution $x_\infty^{(l)}$ of (3.1) satisfying

$$\left\|x_\infty^{(l)} - \xi_M^{(l)}\right\| < \frac{\gamma}{2}.$$

According to Lemma 45, the sequence $x_M^{\delta_i}$ splits into convergent subsequences, each one converging to an element of $\mathcal{X}_M$. Let $(x_M^{\delta_{i_j}})_{j\in\mathbb{N}}$ be a generic convergent subsequence, which converges to an element of $\mathcal{X}_M$, say

$$\lim_{\ell\to\infty} x_M^{\delta_{i_\ell}} = \xi_M^{(l_0)} \in \mathcal{X}_M.$$

We prove now that the subsequence $(x_{N(\delta_{i_\ell})}^{\delta_{i_\ell}})_{\ell\in\mathbb{N}}$ converges to the solution $x_\infty^{(l_0)}$. In fact, since $x_M^{\delta_{i_\ell}} \to \xi_M^{(l_0)}$ as $\ell \to \infty$, there exists a $L_1 = L_1(\epsilon)$ such that

$$\left\|\xi_M^{(l_0)} - x_M^{\delta_{i_\ell}}\right\| < \frac{\gamma}{2} \text{ for all } \ell \geq L_1.$$

As $N(\delta_{i_\ell}) \to \infty$ as $\ell \to \infty$, we have $N\left(\delta_{i_\ell}\right) \geq M$ for all $\ell \geq L$ where $L \geq L_1$ is a sufficiently large number. Then, for all $\ell \geq L$,

$$\left\| x_\infty^{(l_0)} - x_M^{\delta_{i_\ell}} \right\| \leq \left\| x_\infty^{(l_0)} - \xi_M^{(l_0)} \right\| + \left\| \xi_M^{(l_0)} - x_M^{\delta_{i_\ell}} \right\| < \gamma.$$

Finally, the $s-$convexity of X leads to

$$\left\| x_{N(\delta_{i_\ell})}^{\delta_{i_\ell}} - x_\infty^{(l_0)} \right\|^s \overset{(2.23)}{\leq} C\Delta_p\left(x_\infty^{(l_0)}, x_{N(\delta_{i_\ell})}^{\delta_{i_\ell}} \right) \overset{(4.10)}{\leq} C\Delta_p\left(x_\infty^{(l_0)}, x_M^{\delta_{i_\ell}} \right) \overset{(4.41)}{<} \epsilon^s.$$

■

Chapter 5

Numerical Experiments

To test the performance of `K-REGINN`, we have chosen a severely ill-posed problem, namely, the *Electrical Impedance Tomography* (in short EIT). In this non-invasive problem, one aims to reconstruct specific features of the interior of an object collecting information on its boundary. The procedure consists in applying different electric current configurations on the boundary of a bounded set and then measuring the resulting potentials on the boundary as well. The objective is to access information of the interior of this set in order to reconstruct the electrical conductivity.

This idea was originally introduced by Calderón in his famous paper [8]. The problem proposed by him is currently known as the *Continuum Model* of EIT (EIT-CM), where electric current is applied on the whole boundary and the corresponding potential is also read in all the point of the boundary. The Calderón's method is more of theoretical than practical interest because in real situations is actually impossible to apply or record this information in the whole boundary. This method is however important in the mathematical point of view. More details and some numerical experiments using `K-REGINN` to solve this problem are presented in Section 5.1 below.

To modify the EIT-CM in order to adjust it to a more concrete and realistic framework, many attempts have been made and different methods have been suggested. One of the most promising and currently considered a very realistic framework is the so-called *Complete Electrode Model* (EIT-CEM), which was first presented by Somersalo et al in [50]. In the EIT-CEM, electrodes are attached on the boundary of an object and the electric current is injected on this object through these electrodes. The resulting voltage is then read in the same electrodes. The inverse problem of reconstructing the electrical conductivity in the entire set from the application of different configurations of currents and the measure of the respective voltages on its boundary is considered in Section 5.2, where we additionally present some numerical experiments and discuss the results.

Concerning practical situations, EIT has a vast range of potential applications in several fields, including non-invasive medical imaging, non-destructive testing for locating resistivity anomalies, monitoring of oil and gas mixtures in oil pipelines, geophysics, environmental sciences, among others. See [5, 23] and the references therein.

Before starting, we would like to point out that both EIT variations are appropriated and naturally suitable to the use of Kaczmarz methods. The reason is simple: the electrical conductivity is reconstructed from a set of individual measurements. Each of those measurements can be regarded as an individual operator in a system of equations to be solved. The details will be explained later in a convenient moment.

5.1 EIT - Continuum Model

Let $\Omega \subset \mathbb{R}^2$ be a simply connected Lipschitz domain. Applying some electric currents $f\colon \partial\Omega \to \mathbb{R}$ on the boundary of Ω and recording the voltages $g\colon \partial\Omega \to \mathbb{R}$ on the same set, we aim to find the electrical conductivity $\gamma\colon \Omega \to \mathbb{R}$ in whole of Ω. The set Ω is assumed to have no electrical sources or drains, which means that the electric flux $\gamma\nabla u$ is divergence free, where $u\colon \Omega \to \mathbb{R}$ represents the potential distribution in Ω. This condition is written as

$$\operatorname{div}(\gamma\nabla u) = 0 \text{ in } \Omega. \tag{5.1}$$

The electric flux is assumed to be completely transferred to the boundary, which means that

$$\gamma\frac{\partial u}{\partial \nu} = g \text{ on } \partial\Omega, \tag{5.2}$$

where $\partial u/\partial\nu$ is the outward normal derivative of u. In order to prove existence and uniqueness of a solution u, we write the weak formulation of (5.1) and (5.2) : Given $g \in H_\Diamond^{-1/2}(\partial\Omega)$ and $\gamma \in L_+^\infty(\Omega)$, find $u \in H_\Diamond^1(\Omega)$ such that

$$\int_\Omega \gamma\nabla u\nabla\varphi = \int_{\partial\Omega} g\varphi \text{ for all } \varphi \in H_\Diamond^1(\Omega). \tag{5.3}$$

The symbol $\Diamond$ means that the integral of the function over the boundary of Ω is zero:

$$H_\Diamond^{1/2}(\partial\Omega) := \left\{ v \in H^{1/2}(\partial\Omega) : \int_{\partial\Omega} v = 0 \right\}.$$

If the function is defined in Ω, this integral is understood in the sense of the trace theorem: for $u \in H^1(\Omega)$, its trace $f = u|_{\partial\Omega}$ belongs to $H^{1/2}(\partial\Omega)$, and we define

$$H_\Diamond^1(\Omega) := \left\{ u \in H^1(\Omega) : \int_{\partial\Omega} f = 0 \right\}.$$

Finally, the set $H_\Diamond^{-1/2}(\partial\Omega)$ is defined as the dual space of $H_\Diamond^{1/2}(\partial\Omega)$ and

$$L_+^\infty(\Omega) := \{ v \in L^\infty(\Omega) : v \geq C \text{ a.e. in } \Omega \},$$

where $C > 0$ is a constant.

The condition $\int_{\partial\Omega} g = 0$ ($g \in H_\Diamond^{-1/2}(\partial\Omega)$) is interpreted as the law of conservation of charge. Using it in combination with the Lemma of Lax Milgram, one can prove that there exists a vector $u \in H^1(\Omega)$ satisfying (5.3) and that it is unique up to a constant. The condition $\int_{\partial\Omega} f = 0$ ($u \in H_\Diamond^1(\Omega)$) is therefore required to guarantee uniqueness of solutions in (5.3). It can be interpreted as the grounding of potential. Further, the lower bound $\gamma \geq C$ a.e. in Ω ($\gamma \in L_+^\infty(\Omega)$) ensure that the electric flux can flow through the entire set Ω.

With the existence and uniqueness of solutions of the variational problem (5.3), the Neumann-to-Dirichlet (NtD) operator Λ_γ, which associates the current with the respective voltage,

$$\Lambda_\gamma\colon H_\Diamond^{-1/2}(\partial\Omega) \to H_\Diamond^{1/2}(\partial\Omega), \; g \longmapsto f$$

is well-defined. Moreover, it can be proven to be an invertible linear bounded operator. Its bounded linear inverse is called the Dirichlet-to-Neumann operator (DtN).

The *forward operator* associated with EIT-CM problem is now defined as the nonlinear function

$$F\colon \mathrm{D}(F) \subset L^\infty(\Omega) \to \mathcal{L}\left(H_\Diamond^{-1/2}(\partial\Omega), H_\Diamond^{1/2}(\partial\Omega)\right), \; \gamma \longmapsto \Lambda_\gamma, \tag{5.4}$$

with $\mathrm{D}(F) := L_+^\infty(\Omega)$. The EIT-CM inverse problem consists therefore in recovering γ from a partial knowledge of the NtD operator Λ_γ, which is a nonlinear and highly ill-posed problem [1].

In 2006, Astala and Päivärinta [2] proved the injectivity of F in (5.4), which means that the EIT-CM is uniquely solvable. In practice however, one cannot expect to have full knowledge of the NtD operator. The best it can be done is to apply a finite number of currents $\Im := (g_0, ..., g_{d-1})$ and then recover the corresponding voltages, $f_j = \Lambda_\gamma g_j$, $j = 0, ..., d-1$. This fact suggests that the operator

$$F_\Im \colon \mathrm{D}(F) \subset L^\infty(\Omega) \to \left(H_\Diamond^{1/2}(\partial\Omega)\right)^d, \ \gamma \longmapsto (f_0, ..., f_{d-1}),$$

is a natural substitute for (5.4) in practical applications.

Observe that by defining the operators

$$F_j \colon \mathrm{D}(F) \subset L^\infty(\Omega) \to H_\Diamond^{1/2}(\partial\Omega), \ \gamma \longmapsto f_j, \ j = 0, ..., d-1, \tag{5.5}$$

we immediately see that $F_\Im = (F_0, ..., F_{d-1})^\top$ and $Y = Y_0 \times ... \times Y_{d-1}$, with $Y := \left(H_\Diamond^{1/2}(\partial\Omega)\right)^d$ and $Y_j := H_\Diamond^{1/2}(\partial\Omega)$ for $j = 0, ..., d-1$. This is exactly the structure presented in (3.6), which makes this problem suitable to the application of a Kaczmarz method. Moreover, each of the operators in (5.5) is Fréchet-differentiable as we will discuss now.

5.1.1 Fréchet-differentiability of the forward operator

The F-differentiability of the forward operators F_j in (5.5) is a well-known result, see e.g. [35], and we do not intend to prove it here. The goal of this subsection is only to give a roughly explanation of how the vectors $F_j'(\gamma)\sigma$ and $F_j'(\gamma)^*\eta$ for $\gamma \in \operatorname{int}(\mathrm{D}(F))$, $\sigma \in L^\infty(\Omega)$ and $\eta \in H_\Diamond^{1/2}(\partial\Omega)$ can be calculated, which is important for our numerical implementations in the next subsection.

Fix $g_j \in H_\Diamond^{-1/2}(\partial\Omega)$ and define $G_j(\gamma) := u_\gamma$, where $u_\gamma \in H_\Diamond^1(\Omega)$ is the unique solution of (5.3) with $g := g_j$, i.e.,

$$\int_\Omega \gamma \nabla u_\gamma \nabla \varphi = \int_{\partial\Omega} g_j \varphi \text{ for all } \varphi \in H_\Diamond^1(\Omega). \tag{5.6}$$

Thus $G_j(\gamma)|_{\partial\Omega} = F_j(\gamma)$. Further, let $t \in \mathbb{R}\setminus\{0\}$ be small enough to satisfy $\gamma + t\sigma \in \mathrm{D}(F)$ and define $u_{\gamma+t\sigma} := G_j(\gamma + t\sigma)$, which is the unique vector in $H_\Diamond^1(\Omega)$ satisfying the equation

$$\int_\Omega (\gamma + t\sigma) \nabla u_{\gamma+t\sigma} \nabla \varphi = \int_{\partial\Omega} g_j \varphi \text{ for all } \varphi \in H_\Diamond^1(\Omega). \tag{5.7}$$

Subtracting (5.7) from (5.6) we obtain

$$\int_\Omega \gamma \nabla (u_{\gamma+t\sigma} - u_\gamma) \nabla \varphi = -t \int_\Omega \sigma \nabla u_{\gamma+t\sigma} \nabla \varphi \text{ for all } \varphi \in H_\Diamond^1(\Omega).$$

Dividing both sides of above equality by t and letting $t \to 0$ we obtain

$$\int_\Omega \gamma \nabla w_\sigma \nabla \varphi = -\int_\Omega \sigma \nabla u_\gamma \nabla \varphi \text{ for all } \varphi \in H_\Diamond^1(\Omega), \tag{5.8}$$

where

$$w_\sigma := \lim_{t\to 0} \frac{u_{\gamma+t\sigma} - u_\gamma}{t} = \lim_{t\to 0} \frac{G_j(\gamma + t\sigma) - G_j(\gamma)}{t} = G_j'(\gamma)\sigma$$

is the Fréchet derivative of G_j evaluated in γ and applied in the vector σ. As the trace operator is linear and continuous,

$$F_j'(\gamma)\,\sigma = G_j'(\gamma)\,\sigma\big|_{\partial\Omega} = w_\sigma|_{\partial\Omega}\,.$$

Therefore, in order to calculate $F_j'(\gamma)\,\sigma$, one needs first to find $u_\gamma \in H^1_\lozenge(\Omega)$ in (5.6), use it to find $w_\sigma \in H^1_\lozenge(\Omega)$ in (5.8) and finally evaluate its trace $w_\sigma|_{\partial\Omega} \in H^{1/2}_\lozenge(\partial\Omega)$.

To calculate the vector $F_j'(\gamma)^*\,\eta$, we use the following procedure: let $\vartheta \in L^\infty(\Omega)$ be fixed and let $\psi_\eta \in H^1_\lozenge(\Omega)$ be the unique solution of (5.3) for $g := \eta$, i.e.,

$$\int_\Omega \gamma\nabla\psi_\eta\nabla\varphi = \int_{\partial\Omega}\eta\varphi \text{ for all } \varphi \in H^1_\lozenge(\Omega)\,. \tag{5.9}$$

Now, let $w_\vartheta \in H^1_\lozenge(\Omega)$ denote the unique solution of (5.8) for $\sigma := \vartheta$, which implies that $F_j'(\gamma)\,\vartheta = w_\vartheta|_{\partial\Omega}$. Consequently,

$$\begin{aligned}\left\langle F_j'(\gamma)^*\,\eta,\vartheta\right\rangle = \left\langle \eta, F_j'(\gamma)\,\vartheta\right\rangle &= \int_{\partial\Omega}\eta w_\vartheta \overset{(5.9)}{=} \int_\Omega \gamma\nabla\psi_\eta\nabla w_\vartheta\\ &= \int_\Omega\gamma\nabla w_\vartheta\nabla\psi_\eta \overset{(5.8)}{=} -\int_\Omega\vartheta\nabla u_\gamma\nabla\psi_\eta = \left\langle -\nabla u_\gamma\nabla\psi_\eta,\vartheta\right\rangle.\end{aligned}$$

Thus,

$$F_j'(\gamma)^*\,\eta = -\nabla u_\gamma\nabla\psi_\eta, \tag{5.10}$$

with $u_\gamma, \psi_\eta \in H^1_\lozenge(\Omega)$ being given in (5.6) and (5.9) respectively.

Unfortunately, the space $L^\infty(\Omega)$ has too poor smoothness/convexity properties to be included in the convergence analysis of K-REGINN in Chapter 4 (it is not uniformly smooth, for example). But, as Ω is bounded, $L^\infty(\Omega) \subset L^p(\Omega)$ for $1 < p < \infty$, and since these spaces have the necessary properties (remember that for $1 < p < \infty$, the Lebesgue spaces L^p are $p\vee 2-$convex and $p\wedge 2-$smooth (see Example 14)), a possible and immediate solution would be redefine the operators F_j in different spaces:

$$F_j\colon \mathrm{D}(F) \subset L^p(\Omega) \to H^{1/2}_\lozenge(\partial\Omega)\,, 1 < p < \infty.$$

The duality mapping $J_p\colon L^p(\Omega) \to L^{p^*}(\Omega)$ can be now easily calculated via (2.14).

Using this strategy however, a new problem becomes evident: $\mathrm{D}(F)$ has no interior points in the L^p-topology[1], which means that differentiability or even the continuity of F_j are compromised. To overcome this technical obstacle, we suggest restricting the searched-for conductivity space X to a finite dimensional space $V \subset L^\infty(\Omega)$, that is, redefine the functions F_j's as

$$F_j\colon V_+ \subset V_p \to L^2(\partial\Omega_j)\,,\ \gamma \longmapsto f_j,\ V_+ := \mathrm{D}(F)\cap V, \tag{5.11}$$

where $\partial\Omega_j$ is the part of the boundary where the experiments are actually taken. The subscript index p in $V_p := (V, \|\cdot\|_{L^p})$ highlights the fact that the L^p-topology is used in[2] $V = \operatorname{span}\{v_1, ..., v_M\}$. This is a reasonable framework because from a computational point of view, the best one can do is to find the coefficients of the conductivity vector in a specific basis of a finite dimensional subspace of $L^\infty(\Omega)$. Moreover, using a finite number of experiments, only a finite number of degrees of freedom of the conductivity can be restored.

Since in finite dimensional spaces all the norms are equivalent, the F-derivative of F_j remain the same[3].

[1] For $\gamma \in \mathrm{D}(F)$ fixed, the ball $\left\{\widetilde{\gamma} \in L^\infty(\Omega) : \|\gamma - \widetilde{\gamma}\|_{L^p(\Omega)} < \rho\right\}$ is not entirely contained in $\mathrm{D}(F)$ for any $\rho > 0$.

[2] The vectors $v_i \in L^\infty(\Omega)\,, i = 1, ..., M$ are naturally assumed to be linearly independent.

[3] The change of the data space $H^{1/2}_\lozenge(\partial\Omega)$ with $L^2(\partial\Omega_j)$ in (5.11) does not alter the derivative of F_j either, since $H^{1/2}(\partial\Omega) \hookrightarrow L^2(\partial\Omega)$, see e.g. [14, Theo. 2.72].

5.1.2 Computational implementation

In order to improve the reconstructions in our numerical experiments, we propose the use of a *weight-function* as done in [52]. To formalize the results and properly adjust the ideas to our scheme, the next proposition needs to be proven.

Proposition 48 *Let X and Y be Banach spaces defined over the same field $\Bbbk$. If there exists a linear, isometric and surjective operator $T\colon X \to Y$, then the operator $\left(T^{-1}\right)^* : X^* \to Y^*$ is well-defined and shares the same properties.*

Proof. Suppose that $T\colon X \to Y$ has the required properties. From its isometry follows that T is injective and hence bijective. Thus $T^{-1}\colon Y \to X$ is well-defined and it is clearly linear, invertible and isometric too. From the isometry of T^{-1} follows now the boundedness of this operator, which implies that its Banach-adjoint $\left(T^{-1}\right)^* : X^* \to Y^*$ is well-defined. Further, $\left(T^{-1}\right)^*$ is a linear operator. It remains to prove that this operator is isometric and surjective. Let therefore $x^* \in X^*$ be given. As T^{-1} is invertible and isometric,

$$\left\| \left(T^{-1}\right)^* x^* \right\|_{Y^*} = \sup_{\|y\|_Y = 1} \left| \left\langle \left(T^{-1}\right)^* x^*, y \right\rangle \right| = \sup_{\|T^{-1}y\|_X = 1} \left| \left\langle x^*, T^{-1}y \right\rangle \right| = \|x^*\|_{X^*},$$

which proves that $\left(T^{-1}\right)^*$ is isometric. Lastly, to prove surjectivity, let $y^* \in Y^*$ be given and define the operator $x^*\colon X \to \Bbbk$ as

$$\langle x^*, x \rangle := \langle y^*, Tx \rangle, \; x \in X.$$

Then it is clear that $x^* \in X^*$. We will prove that $\left(T^{-1}\right)^* x^* = y^*$. In fact, for each $y_0 \in Y$, define $x_0 := T^{-1}y_0$. Then

$$\langle y^*, y_0 \rangle = \langle y^*, Tx_0 \rangle = \langle x^*, x_0 \rangle = \left\langle x^*, T^{-1}y_0 \right\rangle = \left\langle \left(T^{-1}\right)^* x^*, y_0 \right\rangle.$$

■

Note that if T has the required properties of above proposition, then it is a linear, bounded, isometric and invertible operator. Further, its inverse is also a bounded operator, because it is isometric too. This means that T is an isomorphismus of Banach spaces and the same applies to $\left(T^{-1}\right)^*$. Moreover, it obviously holds, for all $x^* \in X^*$ and $x \in X$,

$$\langle x^*, x \rangle_{X^* \times X} = \left\langle \left(T^{-1}\right)^* x^*, Tx \right\rangle_{Y^* \times Y},$$

which implies that $\left(T^{-1}\right)^*$ is actually an isomorphismus of dual spaces. The result implies in particular that

$$x^* \in J_\varphi(x) \Longleftrightarrow \left(T^{-1}\right)^* x^* \in J_\varphi(Tx),$$

this is, $\left(T^{-1}\right)^* J_\varphi(x) = J_\varphi(Tx)$. Then, for all $x, y \in X$,

$$\langle J_\varphi(Tx), Ty \rangle_{Y^* \times Y} = \left\langle \left(T^{-1}\right)^* J_\varphi(x), Ty \right\rangle_{Y^* \times Y} = \langle J_\varphi(x), y \rangle_{X^* \times X}.$$

Each property of X which only depends either on the norm $\|\cdot\|_X$ or on the duality pairing $\langle \cdot, \cdot \rangle_{X^* \times X}$ of its dual space X^* is therefore preserved. In particular, smoothness and convexity properties of X and Y are exactly the same. For instance, the equivalence

$$\frac{1}{p}\|x-y\|^p \leq \frac{1}{p}\|x\|^p - \langle J_p(x), y \rangle + \frac{\overline{C}_p}{p}\|y\|^p$$

if and only if

$$\frac{1}{p}\|Tx-Ty\|^p \leq \frac{1}{p}\|Tx\|^p - \langle J_p(Tx), Ty \rangle + \frac{\overline{C}_p}{p}\|Ty\|^p,$$

implies that X is $p-$smooth if and only if Y is $p-$smooth. Furthermore, the constant $\overline{C}_p$ remains unchanged.

The motivation for proving Proposition 48 is to define the *weighted Banach space* $L^p_\omega(\Omega)$. Let $\omega\colon \Omega \to \mathbb{R}$ be a positive function and define the vector space

$$L^p_\omega(\Omega) := \left\{ f\colon \Omega \to \mathbb{R} : \int_\Omega |f|^p \omega < \infty \right\}.$$

Observe that $f \in L^p_\omega(\Omega)$ if and only if $f\omega^{1/p} \in L^p(\Omega)$. We can therefore define the norm

$$\|f\|_{L^p_\omega(\Omega)} := \left\| f\omega^{1/p} \right\|_{L^p(\Omega)}, \; f \in L^p_\omega(\Omega),$$

which transforms $L^p_\omega(\Omega)$ in a normed space. If $\omega_{\min} \leq \omega(x) \leq \omega_{\max}$ for all $x \in \Omega$, where $0 < \omega_{\min} \leq \omega_{\max}$ are constants independent of x, then

$$\frac{1}{\omega_{\max}} \|f\|^p_{L^p_\omega(\Omega)} \leq \|f\|^p_{L^p(\Omega)} \leq \frac{1}{\omega_{\min}} \|f\|^p_{L^p_\omega(\Omega)}.$$

Thus, the two vector spaces have equivalent norms and the same elements. Further, the completeness of $L^p(\Omega)$ is transmitted to $L^p_\omega(\Omega)$ and the operator $T\colon L^p_\omega(\Omega) \to L^p(\Omega)$, $f \longmapsto f\omega^{1/p}$ is well-defined and satisfies all the hypotheses of Proposition 48. Therefore the Banach space $L^p_\omega(\Omega)$ inherits the properties of $L^p(\Omega)$. In particular, $(L^p_\omega(\Omega))^* = L^{p^*}_\omega(\Omega)$ and for all $f, g \in L^p_\omega(\Omega)$,

$$\begin{aligned}
\langle J_p(f), g\rangle_{L^{p^*}_\omega \times L^p_\omega} &= \langle J_p(Tf), Tg\rangle_{L^{p^*} \times L^p} = \left\langle |Tf|^{p-1} \operatorname{sgn}(Tf), Tg \right\rangle_{L^{p^*} \times L^p} \\
&= \left\langle |f|^{p-1} \operatorname{sgn}(f)\, \omega^{1/p^*}, g\omega^{1/p} \right\rangle_{L^{p^*} \times L^p} = \int_\Omega |f|^{p-1} \operatorname{sgn}(f)\, g\omega \\
&= \left\langle |f|^{p-1} \operatorname{sgn}(f), g \right\rangle_{L^{p^*}_\omega \times L^p_\omega}.
\end{aligned}$$

Hence, the duality mapping J_p in $L^p_\omega(\Omega)$ is still given by $J_p(f) = |f|^{p-1}\operatorname{sgn}(f)$.

The adjoint operator $F'_j(\gamma)^*$ changes slightly in the new space because for the new adjoint operator $\overline{F}'_j(\gamma)^*$ we want to have for each $\gamma \in \operatorname{int}(\mathrm{D}(F))$, $\sigma \in L^p(\Omega)$ and $\eta \in L^2(\partial\Omega_j)$,

$$\left\langle \overline{F}'_j(\gamma)^* \eta, \sigma \right\rangle_{L^{p^*}_\omega(\Omega) \times L^p_\omega(\Omega)} \stackrel{!}{=} \langle \eta, F'_j(\gamma)\sigma \rangle_{L^2(\partial\Omega_j)} = \langle F'_j(\gamma)^* \eta, \sigma \rangle_{L^{p^*}(\Omega) \times L^p(\Omega)},$$

that is

$$\int_\Omega \overline{F}'_j(\gamma)^* \eta\sigma\omega = \int_\Omega F'_j(\gamma)^* \eta\sigma.$$

Thus (see (5.10)),

$$\overline{F}'_j(\gamma)^* = \omega^{-1} F'_j(\gamma)^* = -\omega^{-1} \nabla u_\gamma \nabla \psi_\eta. \tag{5.12}$$

We keep using the old notation F_j instead of $\overline{F}_j$ for the new operator.

The idea of using the function ω is to give different weights for different regions of Ω. With this procedure we hope to increase stability in some regions where is important to have it. We come back to this subject later to introduce an appropriated weight-function for our framework. Before doing that however, some preliminaries are needed.

For our numerical experiments in this section, we observe only the Kaczmarz version of `REGINN` and let the comparison between Kaczmarz and non-Kaczmarz methods for the EIT-CEM problem, which will be analyzed in next section.

In order to perform the experiments, we have chosen Ω as the unit square $(0,1)\times(0,1)$ and $d=4m$ $(m\in\mathbb{N})$ independent currents

$$g_j := \begin{cases} \cos(2j_0\pi x)\cos(2j_0\pi y) & : (x,y)\in\Gamma_{j_1} \\ 0 & : \text{ otherwise} \end{cases},$$

where $j = 4(j_0-1)+(j_1-1)$, $j_0 = 1,...,m$ and $j_1 = 1,...,4$. The sets Γ_{j_1} represent the faces of Ω : $\Gamma_1 := [0,1]\times\{1\}$, $\Gamma_2 := \{1\}\times[0,1]$, $\Gamma_3 := [0,1]\times\{0\}$ and $\Gamma_4 := \{0\}\times[0,1]$. The voltages $f_j = \Lambda_\gamma g_j$ are measured in $\partial\Omega_j = \partial\Omega\backslash\Gamma_{j_1}$, which means that we do not read the voltages where we apply the currents.

For implementing `K-REGINN` numerically, the computational evaluation of the vectors $F_j(\gamma)$, $F_j'(\gamma)\sigma$ and $F_j'(\gamma)^*\eta$ for $\gamma\in\operatorname{int}(V_+)$, $\sigma\in V_p$ and $\eta\in L^2(\partial\Omega)$ are necessary. Since an analytical solution is in general not available, we apply the Finite Element Method (FEM), constructing a Delaunay triangulation

$$\Upsilon := \{T_i : i = 1,...,M\} \tag{5.13}$$

for Ω, provided by MATLAB's pde toolbox with $M = 2778$. The same triangulation is then used to solve the elliptic problems (5.6), (5.8) and (5.9) and to reconstruct the conductivity. The elements of the basis used to define the finite dimensional space V are defined as

$$v_i(x) := \chi_{T_i}(x) = \begin{cases} 1 & : x\in T_i \\ 0 & : x\notin T_i \end{cases},\ x\in\Omega, \tag{5.14}$$

which means that we are looking for piecewise constant conductivities. This choice of V guarantees injectivity of $F_j'(\gamma)$. Moreover, F_j satisfies the Tangential Cone Condition (Assumption $1(c)$, page 31), see [35, Subsection 3.1].

We want to test the performance of `K-REGINN` to reconstruct sparse conductivities in the $\{v_1,...v_M\}$ basis, and for this reason, we define the exact solution as

$$\gamma^+(x) := \gamma_0(x) + \sum_{i=1}^{M}\alpha_i v_i(x),\ \alpha_i := \begin{cases} 0.9 & : x\in B \\ 0 & : \text{ otherwise} \end{cases},\ x\in\Omega \tag{5.15}$$

where $\gamma_0(x)\equiv 0.1$ represents a known background and it is also the first iterate. The set B represents two open balls inclusions with radii equal 0.15 and centers $(0.35, 0.35)$ and $(0.65, 0.65)$. Figure 5.1 shows γ^+ and the triangulation Υ of Ω.

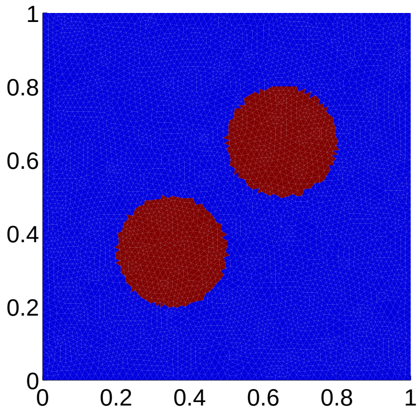
To properly compare the results in different spaces, we always calculate the error in the L^2-norm. The *relative iteration error* of the $n-$th iterate γ_n is therefore defined by

$$e_n := 100\frac{\|\gamma_n-\gamma^+\|_{L^2(\Omega)}}{\|\gamma^+\|_{L^2(\Omega)}}, \tag{5.16}$$

which implies that the *initial error* is $e_0\approx 87,4\%$. The (final) relative iteration error $e_{N(\delta)}$ is denominated the *reconstruction error*. The corresponding data $f_j = \Lambda_{\gamma^+}g_j$ have been synthetically computed using again the FEM, but with a different and much more refined mesh than Υ to avoid an inverse crime. The generated data is then corrupted by artificial random noise of relative noise level δ, that is, the perturbed data are

$$f_j^\delta = f_j + \delta\|f_j\|_{L^2(\partial\Omega_j)}\,per_j,\ j = 0,...,d-1, \tag{5.17}$$

where per_j is an uniformly distributed random variable with $\|per_j\|_{L^2(\partial\Omega_j)} = 1$. in contrast to the previous chapter, δ denotes here a relative noise level. The other input parameter of Algorithm 1 are chosen as $\tau = 1.8$, $\mu = 0.8$ and $k_{\max} = 50$.



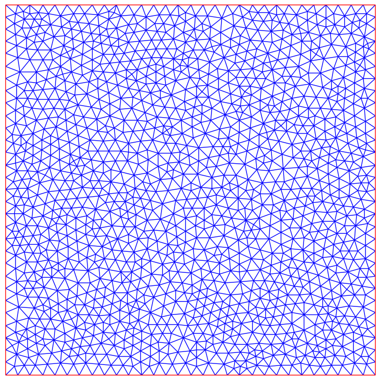


Figure 5.1: Left: the searched-for conductivity γ^+, defined in (5.15) and modelled by a resistive background (in blue) and two balls-like conductive inclusions (in dark red). Right: the triangulation Υ of Ω.

An appropriate choice of the weight-function ω is crucial for the quality of the reconstructions. Based on the explanations from [52], we have defined ω as a piecewise constant function: $\omega := \sum_{i=1}^{M} \omega_i \chi_{T_i}$, with

$$\omega_i := \frac{\sqrt{\sum_{j=0}^{d-1} \left\| F_j'(\gamma_0)\, \chi_{T_i} \right\|_{L^2(\partial\Omega_j)}^2}}{|T_i|}, \tag{5.18}$$

where $|T_i|$ represents the area of the triangle T_i. In spite of being only a heuristic choice for the EIT-CM, this definition of ω has a special meaning for the complete electrode model in Hilbert spaces, see (5.43) and Remark 50 in next section.

Since $Y_j = L^2(\partial\Omega_j)$ is a Hilbert space, we choose $r = 2$ because in this case, $j_2(f) = f$ for all $f \in L^2(\partial\Omega_j)$. Since we are interested in studying sparse conductivities, we follow ideas from [12] and always choose $1 < p \leq 2$. For the first test, we have used $p = 1.1$. The goal of the first experiment is to test the quality of the weight-function ω. Figure 5.2 illustrates the results obtained using $d = 8$ and different relative noise levels δ. It compares the influence of the Banach spaces $X = V_{p,\omega} := \left(V, \|\cdot\|_{L^p_\omega(\Omega)}\right)$, weighted with ω and the standard space $X = V_p$ in the reconstruction $\gamma_{N(\delta)}$. The dual gradient DE method with $C_1 = C_2 = 0.1$ (see (3.29) and (3.39)) is employed as inner iteration of `K-REGINN` to generate the results. Below each picture it is shown the number of outer iterations $N = N(\delta)$, the reconstruction error e_N and the number k_{all}, which represents the sum of all inner iterations performed until convergence. All the pictures are in the same scale of colors and below all of them, a color bar is exhibited. Observe that for all noise levels, the reconstructions in first row, obtained with the weighted-norm $\|\cdot\|_{L^{1.1}_\omega}$, are both qualitatively and quantitatively superior than those found using the $L^{1.1}-$norm and displayed in the second row. Moreover, in the weighted space, Algorithm 1 requires less iterations until convergence. Further, it is clear that the use of the weight-function ω brings more stability to the solutions in the sense that it reduces the oscillation near the boundary, which is constantly present in the reconstructions in the non-weighted space $L^{1.1}$. This figure also makes perceptible in both spaces, the behavior $N(\delta) \to \infty$ as $\delta \to 0$ and the regularization property, $\gamma_{N(\delta)} \to \gamma^+$ as $\delta \to 0$, proved in Theorem 47.

Due to the improvement provided by the use of ω, this weight-function is used in all the remaining experiments of this subsection. For this reason, we skip the dependence of V_p on ω, i.e., from now on $V_p = \left(V, \|\cdot\|_{L^p_\omega(\Omega)}\right)$.

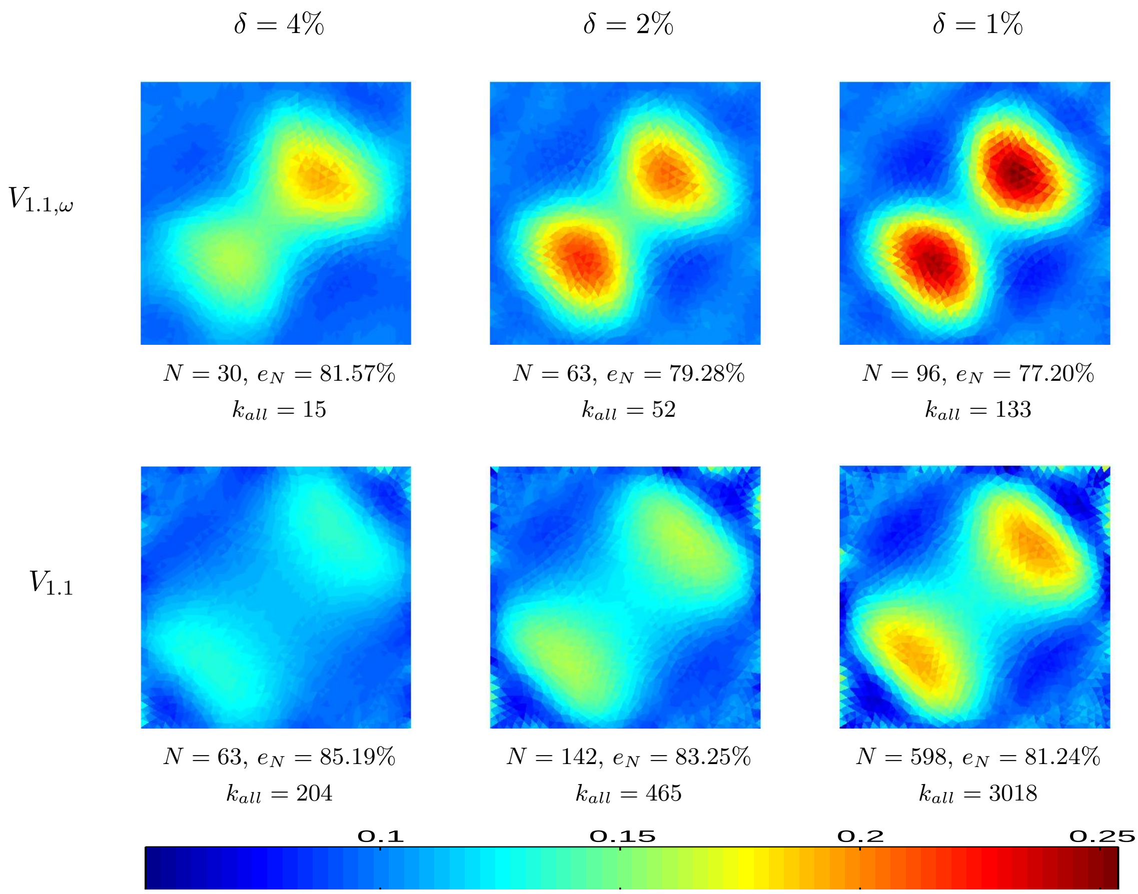


Figure 5.2: Reconstructed conductivity $\gamma_{N(\delta)}$ in two different Banach spaces and with different noise levels. The first and second rows display the results obtained using respectively the weighted-norm $\|\cdot\|_{L^{1.1}_{\omega}}$ and the standard norm $\|\cdot\|_{L^{1.1}}$.

The purpose of the second test is to check the results of Theorem 43. We fix all the same parameters of last experiment (inclusively $p = 1.1$ and $d = 8$) and analyze the behavior of the Bregman variation of TP method (3.58) in the inner iteration of `K-REGINN` with maximal number of inner iterations is fixed to $k_{\max} = 1$. This configuration results in a variation of the Levenberg-Marquardt method. The regularization parameter in (3.57) is chosen as $\alpha_0 = 0.01$. Now, in order to carry out the inner iteration, either the solution of the nonlinear equation (3.58) needs to be found or the minimization of the functional (3.60) must be executed. A fixed-point iteration for $z_{n,k+1}$ in (3.58) could be employed to find this vector. Convergence is guaranteed if the spaces X and Y have enough regularity and the constant α_0 is large enough, see e.g. [39, Appendix A]. However, these conditions are very restrictive: convergence is only ensured for large values of α_0 and this constant needs to be very large for small values of p, which in a practical point of view, seems to render this method unfeasible. In the other hand, minimizing the functional (3.60) , even in Hilbert spaces, is not a trivial task and it is far from simple in more general Banach spaces. In order to achieve this goal, we have utilized the DE method ((3.29) and (3.39)) with $C_1 = C_2 = 0.01$, which showed itself robust enough to reach a satisfactory precision in the minimization of (3.60). As Theorem 43 refers to the noiseless situation, no artificial noise is added in the generated data ($\delta = 0$). Table 5.1 presents the results. It reinforces the fact that in the noiseless situation, both the residual b_n and the relative iteration error e_n converge to zero as the number of outer iterations n grows to infinity.

n	0	1	10	100	1000	10000
e_n	87.40	86.27	80.50	74.57	69.29	60.17
$\|\|b_n\|\|$	0.0715	0.0690	0.0291	0.0030	0.0014	0.0008

Table 5.1: Levenberg-Marquardt method employed to confirm the results from Theorem 43: the relative iteration error e_n as well as the nonlinear residual b_n approaches zero as the outer iteration index n grows to infinity.

In the next experiment we aim to confirm the results of Theorem 38, which states that the iteration error in the Bregman distance is monotonically non-increasing in the outer iteration of `K-REGINN`, this is, $\Delta_p(\gamma^+, \gamma_n) \leq \Delta_p(\gamma^+, \gamma_{n-1})$ for $n = 1, ..., N(\delta)$. Figure 5.3 exhibits the evolution of the iteration error versus the outer iteration index n for three different dual gradient methods (3.29) engaged as inner iteration: the DE, MSD and LW methods with $C_1 = C_2 = 0.1$ in (3.39), $K_1 = K_2 = 0.1$ in (3.45) and $\lambda_{LW} = 0.1$ (see (3.43)). The noise level is set to $\delta = 1\%$, $k_{\max} = 50$ and $\tau = 1.5$. The other parameters remain the same as in the last experiment. See that the (final) iteration errors are comparable for all methods. This behavior is somehow expected because the quality of the approximations in the inner iteration are not only dependent on the utilized method itself, but it is also controlled by the stop criteria of inner iteration (3.5), which remains the same for all methods. Among these three methods however, the DE method is by far the method which results in the fastest convergence while the LW method seems to be the slowest. See that although the iteration error is strictly decreasing in most iterates, it remains constant in some of them. This behavior is in accordance with the results of Theorem 38, which states that the iteration error in the Bregman distance remains the same in the iterates where the discrepancy principle 3.9 is satisfied and it is strictly decreasing otherwise.

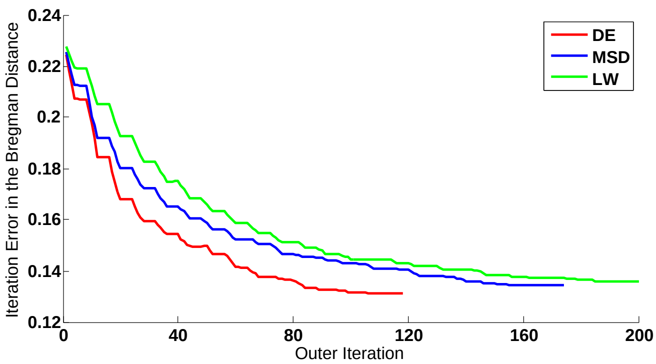


Figure 5.3: The decreasing error behavior of the outer iteration of `K-REGINN` (shown in Theorem 38) observed with three different dual gradient methods in the inner iteration.

In order to observe the influence of different Banach spaces in the reconstructions of γ^+, we apply the LW method with the same configuration of last experiment to generate the Figure 5.4. The reconstruction error $e_{N(\delta)}$ is displayed in the vertical axis and confronted with the noise level δ. The Banach spaces V_p, with $p = 2$, $p = 1.1$ and $p = 1.01$ are tested in this situation and the results are compared using a discrete set of four different noise levels: $\delta = 4\%$, $\delta = 2\%$, $\delta = 1\%$ and $\delta = 0,5\%$. The regularization property $\gamma_{N(\delta)} \to \gamma^+$ as $\delta \to 0$, is somehow observed in all the tested Banach spaces. Note however that the reconstruction errors are higher for $p = 2$ in all noise levels and become lower as p approaches 1. This quantitative improvement in the reconstructions for small values of p is expected and can be credited to the L^p Banach space norm for suitable values of p. in contrast to the Hilbert space L^2, which has the tendency of producing over-smoothed reconstructions and

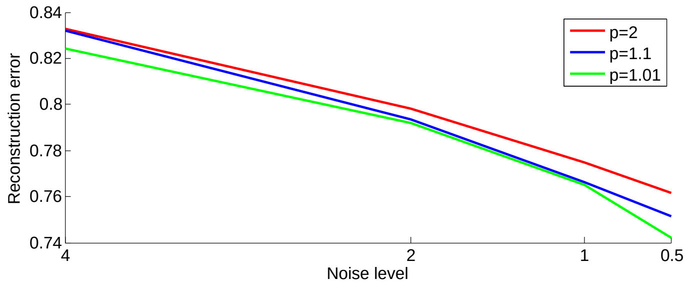


Figure 5.4: Reconstruction error confronted with the noise level and observed in three different Banach spaces norms: L^2 (in red) $L^{1.1}$ (in blue) and $L^{1.01}$ (in green).

consequently destroying the sparsity properties of the solution γ^+, the L^p-norm with p assuming values close to 1 preserves these sparsity features and it is therefore more capable to reconstruct sparse solutions, see e.g., [12].

One of the main advantages of using the Kaczmarz methods is the fact that this kind of algorithms does not work with all the equations of (3.6) in all cycles. Only the equations whose the current iterate is not a good enough approximation[4] are used to perform an iteration. The remaining equations whose the current iterate is already considered good enough, are not used. This procedure only updates the current iterate in the necessary directions, accelerating the method until convergence and rendering better reconstructions. This behavior is evident in Figure 5.5, where the number of active equations is compared on each cycle of `K-REGINN`. The dual gradient MSD method ((3.29) and (3.45)) is employed as inner iteration with the same parameters of last experiment but with the fixed noise level $\delta = 1\%$, $p = 1.1$ and $d = 16$. Observe that after some few initial cycles, the number of active equations drops to relative small levels, where only the relevant equations are worked until termination. The average of active equations in this example is roughly half of the whole number.

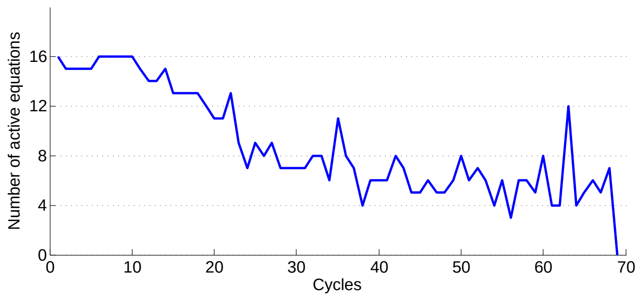


Figure 5.5: Behavior of Kaczmarz methods: the average of active equations becomes lower with the time and the current iterate is corrected using only the relevant information.

The regularization property proved in Theorem 47 is once more evident in Figure 5.6, where different numbers of electric currents are applied on the boundary of Ω (different values of d, or equivalently, different numbers of equations in (3.6)) are tested. We want to illustrate both behaviors in this experiment: more information improves the reconstructions independently on the noise levels and the results become better whenever the noise levels are reduced. Each row in this Figure presents the reconstructions for a different value of $d : d = 4$, $d = 8$ and $d = 12$. The columns exhibit and compare the different noise levels $\delta = 2\%$, $\delta = 1\%$ and $\delta = 0.5\%$. The pictures have been generated using the Bregman variation of the IT method (3.62) with the same parameters of the last experiment but

[4] Called the *active equations*, they are those equations, whose inequality (3.9) is not satisfied.

with $k_{\max} = 10$. Further, the regularization parameter in (3.68) is defined as the constant $\alpha_n \equiv 0.1$ and the dual gradient DE method ((3.29) and (3.39)) with $C_1 = C_2 = 0.01$ is used to find the minimizer of the associated functional (3.63). All the pictures are in the same scale of colors and below all of them, a color bar shows the values represented by each of those colors. A clear improvement in the reconstructions is seen moving towards the rightmost column (where the noise levels are lower) and moving downwards (where more information is available).

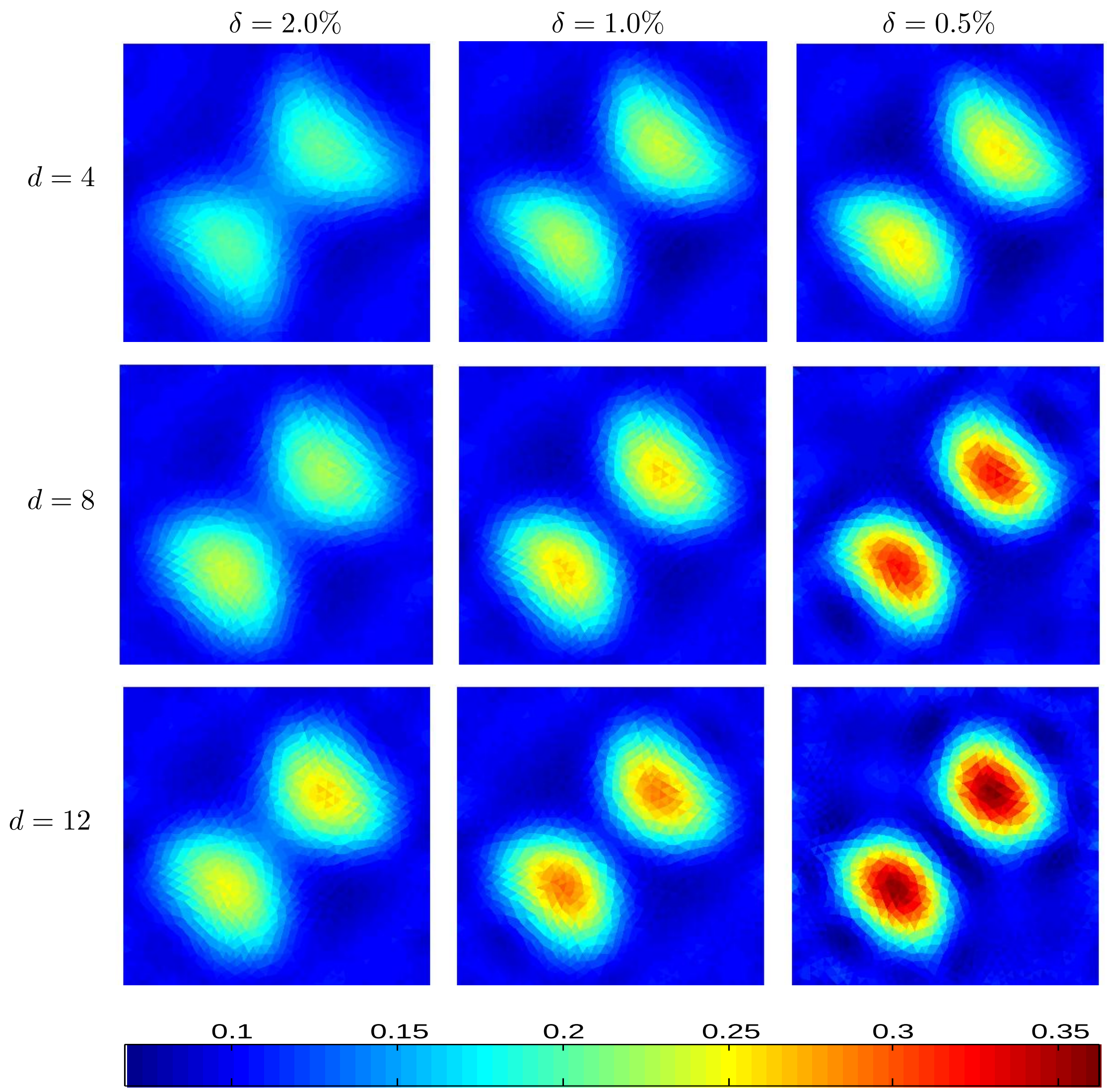
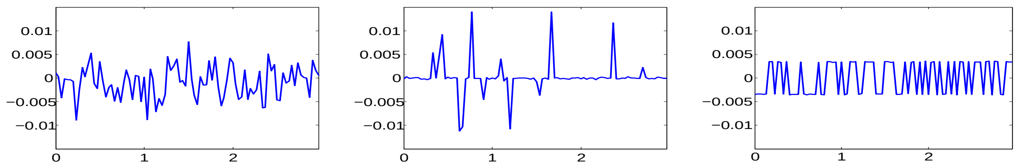


Figure 5.6: Reconstruction $\gamma_{N(\delta)}$ in the space $V_{1.1}$ obtained with the noise levels $\delta = 4\%$, $\delta = 2\%$ and $\delta = 1\%$ and with different number of equations: $d = 4$, $d = 8$ and $d = 12$.

The norm of the noise can vary if it is regarded as a vector in different spaces. For instance, the so-called *impulsive noise*, which consists of standard uniform noise superimposed with some few highly inconsistent data points called *outliers*, has a small L^r−norm for r small and becomes larger if r increases. In contrast, the so-called *Gaussian noise*, which is an uniformly distributed noise, is less sensitive to the chosen norm and has more similar values in different L^r−norms. In the first row of Figure 5.7, three completely different kinds of noise are presented. The above referred Gaussian an impulsive noises are the first and second picture respectively. Below each kind noise, the value of the corresponding L^r−norm is shown for different values of r. The noises are scaled such that the (relative)

Noise



$r = 2$	6.0e−3	6.0e−3	6.0e−3
$r = 1.01$	8.1e−3	3.9e−3	10.2e−3
$r = 50$	8.4e−3	13.2e−3	3.5e−3

Figure 5.7: Norm of three different kinds of noise, measured in the Banach space norms $L^2(\partial\Omega_1)$ (first row), $L^{1.01}(\partial\Omega_1)$ (second row) and $L^{50}(\partial\Omega_1)$ (third row).

L^2-noise is $\delta = 1\%$ in all cases.

The same noises shown in the top of Figure 5.7 are used to compute the reconstructions presented in Figure 5.8. The pictures exposed in this new figure are organized in columns and rows in the same way as in the last figure. Thus, each column and each row of Figure 5.8 represents respectively the same kind of noise and space of Figure 5.7. We have used $d = 4$, $\tau = 1.8$ and $k_{\max} = 50$. The dual gradient DE method have been engaged again as inner iteration of `K-REGINN` with and $C_1 = C_2 = 0.1$ in (3.39). The parameter p is fixed in 2, which means that X is a Hilbert space in this experiment. On the other hand, we have chosen different values for r expecting to obtain better reconstructions in those spaces where the noise, measured in the corresponding norm, has smaller values.

Since the set $\partial\Omega_j$ is an uni-dimensional Lipschitz manifold and $H^{1/2}((a,b))$ is continuously embedded in $L^r((a,b))$ for $(a,b) \subset \mathbb{R}$ and $1 < r < \infty$, see e.g. [14, Theo. 2.72], it follows that $H^{1/2}(\partial\Omega_j) \hookrightarrow L^r(\partial\Omega_j)$ [18] and we have permission to set $Y_j = L^r(\partial\Omega_j)$ for an arbitrary $r \in (1,\infty)$. Since $\|g_j\|_{H^{1/2}(\partial\Omega_j)} \lesssim \|g_j\|_{L^r(\partial\Omega_j)}$ for $1 < r < \infty$, the new operators F_j are still Fréchet differentiable and have the same derivatives as before. We want to point out however, that the standard choice J_r for the duality mapping is in principle not allowed if $r < 2$. Indeed, we do not have any convergence results for the Kaczmarz methods in this case. Theorem 47 actually requires that $s \leq r$ for the case $d > 1$, but as X is a Hilbert space, it follows that $s = 2$, and accordingly, the index r of the duality mapping should be larger than or equal 2. This technical problem could be overcome for instance, using the normalized duality mapping J_2 in the L^r space, which can be realized via (2.14). However, this seems not to be the best solution because the reconstructions found with this framework are not as good as the ones found when the duality mapping J_r is used in the space L^r (see also Remark 25). Seemingly, the duality mapping J_r is the right one for the space L^r and for this reason, we have chosen to use it, even without having the complete convergence proof in this situation.

As expected, all the reconstructions in the L^2 space (first row of Figure 5.8) are similar because all the noises have the same L^2-norm. The results shown in the first column are also similar because the Gaussian noise has similar norms in the three different investigated spaces. But, a slightly superior quality of the picture in the first row can be observed. This behavior can be explained looking again to the first column of Figure 5.7, which shows that the smallest value for the Gaussian noise is obtained with the use of the L^2-norm. The most interesting results however, are shown in the two last rows intersected with the two last columns. Comparing the results of Figure 5.8 with those shown in Figure 5.7, we clearly see that an improvement is achieved when the kind of noise matches with the used space, in the sense that the resulting noise-norm has a small value (this is the case in the intersection between the second row with the second column and the third row with the third column). On the other hand, the contrary effect occurs if the combination of kind of

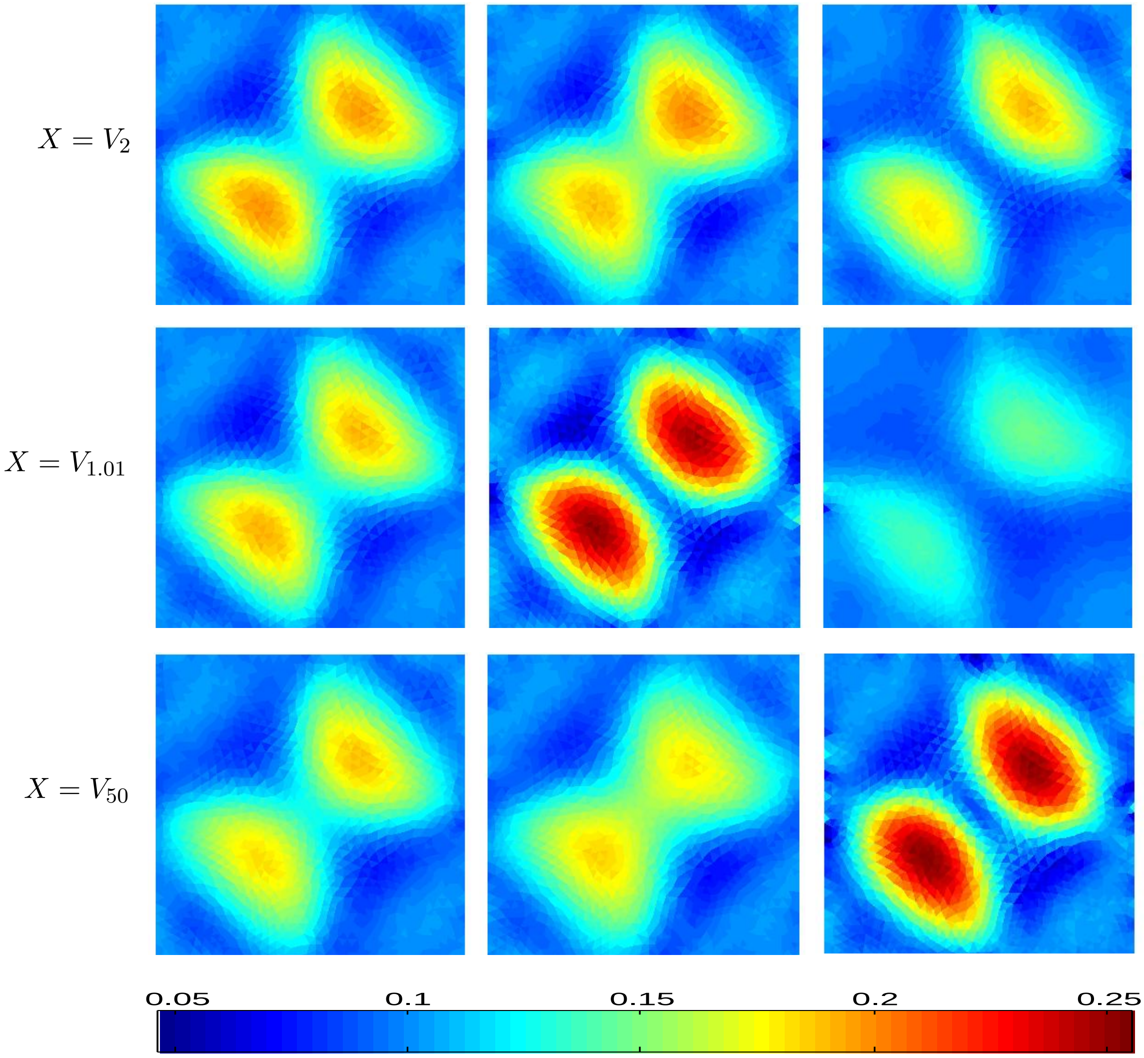


Figure 5.8: Reconstruction $\gamma_{N(\delta)}$ obtained in different Banach spaces norm (compared by rows) and with different kinds of noise (compared by columns). All the noises are scaled and have the same L^2-norm.

noise and used space results in a large noise-norm (see the second row intersected with the third column and the intersection between the third row and the second column).

To finish this section we explain how to take advantage of the weight-function ω to incorporate a priori information about the solution γ^+. If the inclusions B are expected to be located in a specific region A of Ω, one way to get better reconstructions is giving less weight for the points within this region and "penalizing" the distance to this set: let $A \subset \Omega$ be a closed subset in Ω where the inclusions are expect to be located. Define the new weight-function

$$\varpi(x) := \omega(x) h(x), \ x \in \Omega,$$

where

$$h(x) := (c_0 + \text{dist}(x, A)). \tag{5.19}$$

The function $\text{dist}(\cdot, A) : \Omega \to \mathbb{R}$ measures the distance between a point of Ω and the set A :

$$\text{dist}(x, A) := \inf\{\|x - y\| : y \in A\}, \ x \in \Omega.$$

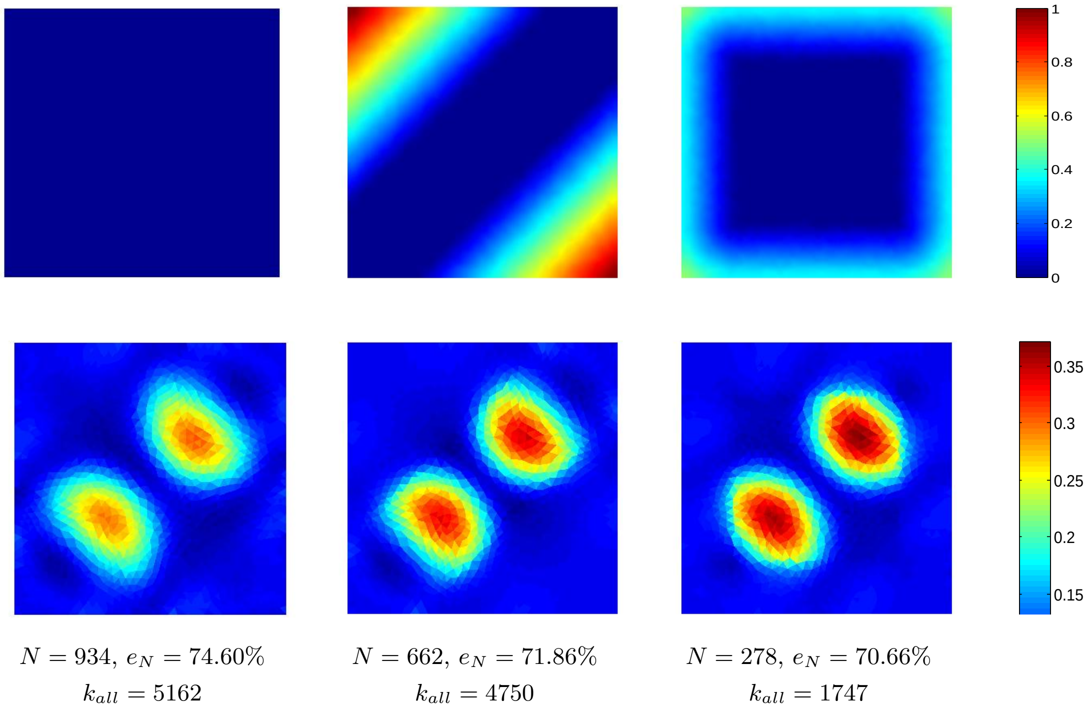


Figure 5.9: Reconstructed solution $\gamma_{N(\delta)}$ for different weight-functions. The first row shows the function h from (5.19).

The small constant $c_0 > 0$ is used to avoid zero-weights and to define the "contrast" between the points which belong and the ones which do not belong to the set A. This new weight-function is still bounded from above and from below: $c_0\omega_{\min} \leq \varpi \leq (c_0 + |\Omega|)\,\omega_{\max}$, where $0 < \omega_{\min} \leq \omega_{\max}$ are respectively lower and upper bounds for ω.

Figure 5.9 collects the results obtained using different sets A and the same constant $c_0 = 0.03$. The first row shows the function h, defined in (5.19), while the second one exhibits the respective reconstructions with the number of outer iteration N, the overall number of inner iteration k_{all} and the reconstruction error e_N being highlighted below each picture. At the end of each row, a color bar is displayed. The first column presents the result for the set $A = \Omega$, which means that the new weight $\varpi(x) = c_0\omega(x)$ is just a multiple of the original weight-function ω. For the reconstruction in the second column the strip-like set

$$A = \{(x, y) \in \Omega : y - 0.25 \leq x \leq y + 0.25\}$$

have been used, and the surface delimited by a square

$$A = [0.2, 0.8] \times [0.2, 0.8]$$

is the chosen set in the third column. The pictures have been generated using the DE method with $C_1 = C_2 = 0.1$ and the following configuration: $d = 8$, $p = 1.1$, $r = 2$, $\tau = 1.5$ and $\delta = 0.5\%$. The other parameters are the same of last experiment.

Note that better results are found in both the second and third columns, where more accurate information about the location of the inclusions is provided. Further, the overall number of outer and inner iterations are lower in both cases.

5.2 EIT - Complete Electrode Model

In this new procedure, introduced by Somersalo et al. in [50] to build a more realistic framework for the EIT problem, electric currents are injected in the simply connected Lipschitz domain $\Omega \subset \mathbb{R}^2$ via $L \in \mathbb{N}$ electrodes attached to its boundary $\partial\Omega$ and the resulting voltages are measured in the same electrodes with the goal of restoring the electrical conductivity $\gamma\colon \Omega \to \mathbb{R}$. For the correct translation in an appropriate mathematical model, we suppose that the electrodes $e_1, ..., e_L$ are identified with the part of the surface of Ω they contact. Thus, $e_i \subset \partial\Omega$ is open and has positive measure $|e_i| > 0$ for $i = 1, ..., L$. Additionally, the electrodes are connected and separated: $\overline{e}_i \cap \overline{e}_j = \varnothing$ for $i \neq j$. Similarly to the EIT-CM model, we suppose that Ω has no sources or drains and thus (5.1) is satisfied, where $u\colon \Omega \to \mathbb{R}$ represents again the potential distribution in Ω. The electrodes are modelled as perfect conductors, which means that the electric current in the electrode e_i is a constant $I_i \in \mathbb{R}$, which agrees with the total electric flux over the same electrode:

$$\int_{e_i} \gamma \frac{\partial u}{\partial \nu} \mathrm{dS} = I_i, \ i = 1, ..., L.$$

The vector $I := (I_1, ..., I_L)^\top \in \mathbb{R}^L$, which collects in a single vector the electric currents of all electrodes, is called a *current pattern.* As the electric flux does not flow in the electrodes gaps we have

$$\gamma \frac{\partial u}{\partial \nu} = 0 \text{ in } \partial\Omega \backslash \cup_{i=1}^{L} e_i.$$

At the contact of the electrodes with $\partial\Omega$, there is an electro-chemical effect which gives rise to a thin, highly resistive layer characterized here by the quantities $z_i > 0$, $i = 1, ..., L$ and called the *contact impedances.* The product of the contact impedance z_i times the electric flux $\gamma\,(\partial u / \partial \nu)$ results in the drop of the voltage in the electrode e_i. Thus

$$u + z_i \gamma \frac{\partial u}{\partial \nu} = U_i, \text{ on } e_i, \ i = 1, ..., L,$$

where U_i is the measured voltage in the $i-$th electrode e_i. We denote by $U := (U_1, ..., U_L)^\top \in \mathbb{R}^L$, the vector of the voltages associated with the current pattern I.

In [50, Prop. 3.1], it is demonstrated that if the set Ω, the electrical conductivity γ and the electrodes e_i have enough regularity, the above conditions can be equivalently replaced by the variational problem

$$B\left((u, U), (v, V)\right) = \sum_{i=1}^{L} I_i V_i \text{ for all } (v, V) \in H, \tag{5.20}$$

where $H := H^1(\Omega) \oplus \mathbb{R}^L$ and the bi-linear form $B\colon H \times H \to \mathbb{R}$ is defined by

$$B\left((u, U), (v, V)\right) := \int_\Omega \gamma \nabla u \nabla v + \sum_{i=1}^{L} \frac{1}{z_i} \int_{e_i} (u - U_i)(v - V_i)\,\mathrm{dS}. \tag{5.21}$$

Aiming now to apply the Lemma of Lax Milgram [32] in (5.20) in order to prove the existence of a vector (u, U) for fixed contact impedances $z_i > 0$ and a fixed current pattern $I \in \mathbb{R}^L$, it was suggested in [50, Theo. 3.3] changing the space H with $\widetilde{H} = H/\mathbb{R}$, which is essentially the same space but with the difference that two vectors from H which differ only by a constant function are regarded as the same vector in $\widetilde{H}$, i.e.,

$$(v, V) \sim \left(\widetilde{v}, \widetilde{V}\right) \Leftrightarrow v - \widetilde{v} = \text{const} = V_1 - \widetilde{V}_1 = ... = V_L - \widetilde{V}_L. \tag{5.22}$$

In this new vector space, the bi-linear form B in (5.21) is coercive whenever $\gamma \in L^\infty_+(\Omega)$. Now defining $f\colon \widetilde{H} \to \mathbb{R}$ as

$$f\left((v, V)\right) := \sum_{i=1}^{L} I_i V_i,$$

we see that, in order to have a well-defined function, the equality $f((v,V)) = f\left(\left(\widetilde{v},\widetilde{V}\right)\right)$ must be verified whenever the right-hand side of (5.22) holds true. Thus, the equality

$$0 = f((v,V)) - f\left(\left(\widetilde{v},\widetilde{V}\right)\right) = \sum_{i=1}^{L} I_i V_i - \sum_{i=1}^{L} I_i \left(V_i + \text{const}\right) = \text{const} \sum_{i=1}^{L} I_i,$$

must be verified for an arbitrary constant. We thus assume that $I \in \mathbb{R}^L_\Diamond$, where

$$\mathbb{R}^L_\Diamond := \left\{ I \in \mathbb{R}^L : \sum_{i=1}^{L} I_i = 0 \right\}.$$

In this case, f is a well-defined, bounded linear functional and the Lemma of Lax Milgram applies. It follows that there exists a unique solution of (5.20) if H is replaced with $\widetilde{H}$. Therefore, there exist infinite solutions in H, differing from each other only by a constant. To obtain uniqueness in H, we suppose that $U \in \mathbb{R}^L_\Diamond$. The conditions $I, U \in \mathbb{R}^L_\Diamond$ can be understood as the law of the conservation of the charge and the grounding of the potential respectively.

The above reasoning leads to the conclusion that the following related variational problem is well-defined: fixed a current pattern $I \in \mathbb{R}^L_\Diamond$ and positive contact impedances $z_1, ..., z_L$, find the unique pair $(u,U) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ satisfying

$$B\left((u,U),(v,V)\right) = \sum_{i=1}^{L} I_i V_i \text{ for all } (v,V) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond, \tag{5.23}$$

where $B\colon \widehat{H} \times \widehat{H} \to \mathbb{R}$, with $\widehat{H} := H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ is defined in (5.21).

Remark 49 *The variational problem* (5.23) *has a unique solution even if the fixed current pattern I belongs to*[5] $\mathbb{R}^L \backslash \mathbb{R}^L_\Diamond$*. Indeed, defining $a := \sum_{i=1}^{L} I_i$ and changing I with $I - C$, where $C := (a/L, ..., a/L)^\top$, we see that $I - C \in \mathbb{R}^L_\Diamond$ and for all $V \in \mathbb{R}^L_\Diamond$,*

$$\sum_{i=1}^{L} (I_i - C_i) V_i = \sum_{i=1}^{L} I_i V_i - \frac{a}{L} \underbrace{\sum_{i=1}^{L} V_i}_{=0} = \sum_{i=1}^{L} I_i V_i.$$

Since a unique solution of (5.23) *exists with the current pattern $I - C$, the same occurs with I.*

Using (5.23), it is not difficult to prove that the Neumann-to-Dirichlet (NtD) operator

$$\Lambda_\gamma\colon \mathbb{R}^L \to \mathbb{R}^L,\ I \mapsto U, \tag{5.24}$$

which associates a current pattern $I \in \mathbb{R}^L$ with the respective voltage vector $U \in \mathbb{R}^L_\Diamond \subset \mathbb{R}^L$ is linear. The Complete Electrode Model of the EIT (in short EIT-CEM) forward problem is now defined by the function

$$F\colon \mathrm{D}(F) \subset L^\infty(\Omega) \to \mathcal{L}\left(\mathbb{R}^L, \mathbb{R}^L\right),\ \gamma \mapsto \Lambda_\gamma, \tag{5.25}$$

with $\mathrm{D}(F) := L^\infty_+(\Omega)$. Recovering γ from a partial knowledge of Λ_γ is the associated inverse problem to be solved.

[5] The current pattern vector $I \in \mathbb{R}^L \backslash \mathbb{R}^L_\Diamond$ does not satisfy the principle of the conservation of the charge. Therefore, it has no physical meaning.

As can be seen from Remark 49, the NtD operator (5.24) is not injective. In fact, the null-space of Λ_γ is given by

$$\mathrm{N}\left(\Lambda_\gamma\right) = \left\{I \in \mathbb{R}^L : I_1 = ... = I_L = \text{const}\right\} \neq 0.$$

But, since it is a linear operator, the knowledge of the vectors $U^j = \Lambda_\gamma I^j$, $j = 1, ..., L$, where $\left\{I^1, ..., I^L\right\}$ is a basis for $\mathbb{R}^L$, implies in the knowledge of the NtD operator Λ_γ itself[6].

In practical situations, one fixes $\ell \in \mathbb{N}$ (not necessarily linearly independent) current patterns $I^j \in \mathbb{R}^L_\Diamond$, $j = 1, ..., \ell$, which for notational reasons we put together in a single vector

$$\Im := \left(I^1, ..., I^\ell\right)^\top \in \mathbb{R}^{\ell L}, \tag{5.26}$$

and reads in the EIT-CEM experiment a noisy version of

$$\Gamma_\gamma := \left(U^1, ..., U^\ell\right)^\top \in \mathbb{R}^{\ell L}, \tag{5.27}$$

where $\left(u^j, U^j\right) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ is the unique solution of (5.23) associated with the current pattern $I = I^j$, that is, $U^j := \Lambda_\gamma I^j$, $j = 1, ..., \ell$. Observe that the noisy versions of U^j belong to the space $\mathbb{R}^L$ but not necessarily to $\mathbb{R}^L_\Diamond$.

We now reformulate (5.25) as

$$F_\Im \colon \mathrm{D}(F) \subset L^\infty(\Omega) \to \mathbb{R}^{\ell L},\ \gamma \mapsto \Gamma_\gamma, \tag{5.28}$$

and regard this operator as the EIT-CEM forward problem for the rest of this section. The determination of an approximation for γ from partial knowledge of Γ_γ is therefore the associated inverse problem we want to solve.

Similarly to the EIT-CM, presented in the last section, the space $Y = \mathbb{R}^{\ell L}$ factorizes into the spaces $Y = Y_1 \times ... \times Y_\ell$, with $Y_j := \mathbb{R}^L$, $j = 1, ..., \ell$ and accordingly, $F_\Im = (F_1, ..., F_\ell)^\top$ with

$$F_j \colon \mathrm{D}(F) \subset L^\infty(\Omega) \to \mathbb{R}^L,\ \gamma \mapsto U^j,\ j = 1, ..., \ell, \tag{5.29}$$

which replaces the problem (5.28) with a more suitable version for an application of a Kaczmarz method.

5.2.1 Fréchet-differentiability of the forward operator

The Fréchet differentiability of the operators F_j in (5.29) can be proven similarly to the EIT-CM, see [28] and [34, Theo 4.1]. Further, for $\gamma \in \mathrm{int}(\mathrm{D}(F))$ and $\eta \in L^\infty(\Omega)$, the F-derivative $F_j'(\gamma)\eta =: W^j$ is the second element of the pair $\left(w^j, W^j\right) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$, which is the unique solution of the following variational problem:

$$B\left(\left(w^j, W^j\right), (v, V)\right) = -\int_\Omega \eta \nabla u^j \nabla v \mathrm{d}x \text{ for all } (v, V) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond, \tag{5.30}$$

where $\left(u^j, U^j\right) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ is the unique solution of (5.23) with the current pattern $I = I^j$, this is, $U^j = F_j(\gamma)$.

Proceeding similarly to (5.10), one can prove that for each $Z \in \mathbb{R}^L$, the adjoint operator of the F-derivative of F_j satisfies

$$F_j'(\gamma)^* Z = -\nabla u^j \nabla y, \tag{5.31}$$

where $(y, Y) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ is the unique solution of (5.23) with the current pattern $I = Z$.

[6] Since $\dim(\mathrm{N}(\Lambda_\gamma)) \neq 0$ and Λ_γ is symmetric [50], less linearly independent measures are actually needed.

The F-derivative of the forward operator (5.28) is now given by $F'_{\Im}(\gamma) = (F'_1(\gamma), ..., F'_\ell(\gamma))^\top$, i.e., for each $\eta \in L^\infty(\Omega)$,

$$F'_{\Im}(\gamma)\eta = \begin{pmatrix} F'_1(\gamma)\eta \\ \vdots \\ F'_\ell(\gamma)\eta \end{pmatrix} \in \mathbb{R}^{\ell L}. \tag{5.32}$$

Since only a finite number of experiments can be made, only a finite number of degrees of freedom of the conductivity can be restored. Thus, it makes fully sense to restrict the searched-for conductivities to a finite dimensional space. Proceeding similarly to the last section, we define the triangulation $\Upsilon := \{T_i : i = 1, ..., M\}$ of Ω as in (5.13) and analogously to (5.11) and (5.14), we restrict the space X to the finite dimensional space V and the domain of definition of F to

$$V_+ := V \cap L^\infty_+(\Omega) \subset V_p, \tag{5.33}$$

where $V := \operatorname{span}\{\chi_{T_1}, ..., \chi_{T_M}\} \subset L^\infty(\Omega)$ and $V_p := \left(V, \|\cdot\|_{L^p(\Omega)}\right)$. The discrete version of the operators defined in (5.28) and (5.29) are

$$F\colon V_+ \subset V_p \to \mathbb{R}^{\ell L},\ \gamma \mapsto \Gamma_\gamma, \tag{5.34}$$

and

$$F_j\colon V_+ \subset V_p \to \mathbb{R}^L,\ \gamma \mapsto U^j,\ j = 1, ..., \ell, \tag{5.35}$$

respectively. Identifying the arbitrary vector $\gamma = \sum_{i=1}^M \alpha_i \chi_{T_i}$ of $V_+ \subset L^\infty(\Omega)$ with the vector of its coordinates $(\alpha_1, ..., \alpha_M)^\top \in \mathbb{R}^M$, we see that the functions in (5.34) and in (5.35) above can now be seen as nonlinear operators from (a subset of) $\mathbb{R}^M$ to $\mathbb{R}^{\ell L}$ and from (a subset of) $\mathbb{R}^M$ to $\mathbb{R}^L$ respectively. Their derivatives $F'(\gamma)$ and $F'_j(\gamma)$, evaluated in a vector $\gamma \in \operatorname{int}(V_+)$, can accordingly be regarded as two matrices (called the *Jacobian* matrices) with dimension $M \times \ell L$ and $M \times L$ respectively. In this framework, the adjoint operators of these derivatives are simply the transpose Jacobian matrices.

We now explain how to calculate the Jacobian matrix efficiently: let $\gamma \in \operatorname{int}(V_+)$ be given and observe that

$$F'_j(\gamma) = \left(F'_j(\gamma)\chi_{T_1}, ..., F'_j(\gamma)\chi_{T_M}\right) \in \mathbb{R}^{L\times M},\ j = 1, ..., \ell.$$

Further, using (5.30), the vector $F'_j(\gamma)\chi_{T_i} =: W_i^j$ can be evaluated as a part of $\left(w_i^j, W_i^j\right) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$, the unique solution of

$$B\left(\left(w_i^j, W_i^j\right), (v, V)\right) = -\int_\Omega \chi_{T_i} \nabla u^j \nabla v \mathrm{dx} \text{ for all } (v, V) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond, \tag{5.36}$$

where $\left(u^j, U^j\right) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ is the unique solution of (5.23) for the current pattern $I = I^j$. This relative straightforward procedure for calculating the Jacobian matrix is however highly expensive. In fact, it is necessary to solve the problem (5.23) for $I = I^j$ in order to obtain the vector u^j and then solve the problem (5.36) for $i = 1, ..., M$ in order to calculate $F'_j(\gamma)$. Similarly, ℓ solutions of (5.23) and $M\ell$ solutions of (5.36) are required for the evaluation of $F'(\gamma)$. But, using a simple trick, as explained in [42], we are able to strongly reduce these numbers. For each $i \in \{1, ..., L\}$, let $\left(v^i, V^i\right) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$ be the unique solution of (5.23) with the current pattern $I = J^i := (\delta_{i,j})_{j=1}^L$, which is the vector having the value 1 in the $i-$th coordinate and zero elsewhere. This is,

$$B\left(\left(v^i, V^i\right), (z, Z)\right) = \left\langle J^i, Z\right\rangle \text{ for all } (z, Z) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond. \tag{5.37}$$

Thus

$$
\begin{aligned}
F_j'(\gamma)\,\chi_{T_i} &= W_i^j = \left(\left\langle J^1, W_i^j\right\rangle, ..., \left\langle J^L, W_i^j\right\rangle\right)^\top && (5.38)\\
&\overset{(5.37)}{=} \left(B\left(\left(v^1, V^1\right), \left(w_i^j, W_i^j\right)\right), ..., B\left(\left(v^L, V^L\right), \left(w_i^j, W_i^j\right)\right)\right)^\top\\
&= \left(B\left(\left(w_i^j, W_i^j\right), \left(v^1, V^1\right)\right), ..., B\left(\left(w_i^j, W_i^j\right), \left(v^L, V^L\right)\right)\right)^\top\\
&\overset{(5.36)}{=} \left(-\int_\Omega \chi_{T_i} \nabla u^j \nabla v^1 \mathrm{dx}, ..., -\int_\Omega \chi_{T_i} \nabla u^j \nabla v^L \mathrm{dx}\right)^\top\\
&= \left(-\int_{T_i} \nabla u^j \nabla v^1 \mathrm{dx}, ..., -\int_{T_i} \nabla u^j \nabla v^L \mathrm{dx}\right)^\top.
\end{aligned}
$$

Thus, in order to calculate the Jacobian matrix, which represents the derivative of (5.34), everything one needs to do is to solve the problem (5.23) for each of the ℓ current patterns in (5.26), solve (5.37) for $i \in \{1, ..., L\}$ and assemble the results as explained in (5.38) and (5.32). This procedure demands the solution of only $\ell + L$ variational problems and since the number of electrodes L is normally much smaller than the number of triangles M in Υ, a tremendous saving in the computational effort can be achieved with this strategy. If we further observe that in each inner iteration of REGINN, the vectors u^j, $j = 1, ..., \ell$ have already been calculated in the outer iteration in order to evaluate $F(\gamma)$, then the calculation of the Jacobian actually requires the additional calculation of L solutions of (5.37). Analogously, the evaluation of the Jacobian $F_j'(\gamma)$ for the Kaczmarz version (5.35) demands L solutions of (5.37) if the information $F_j(\gamma)$ is already available.

All the methods presented in Section 3.1 demand the calculation of at least one derivative applied in a vector ($A_n s$, $s \in X$) or of the calculation of the adjoint of the derivative applied in a vector ($A_n^* b$, $b \in Y^*$) to calculate the first vector $s_{n,1}$ in each inner iteration of REGINN. Thus, if one wants to calculate $s_{n,1}$ for the problem (5.34) without executing the calculation of the Jacobian matrix, we conclude in view of (5.32) that at least L solutions of (5.30) or (5.31) need to be found. Considering that the process of obtaining a solution of a variational problem is the unique relevant computational effort in a run of Algorithm 1, and recalling that L solutions of (5.37) is everything one needs in order to calculate the Jacobian for this problem, we conclude that it is always worth to evaluate this matrix.

The situation is a little different for the Kaczmarz version of REGINN, i.e., if the problem (5.35) is analyzed instead of (5.34). If the Jacobian matrix $F_j'(\gamma)$ is not available and a dual gradient method (3.29) or a mixed gradient-Tikhonov method (3.76) is employed as inner iteration for instance, the solution of two variational problems are required to compute each single vector: one for the calculation of the derivative applied in $s_{n,k}$ and a second one for the adjoint of the derivative applied in $j_r\left(A_n s_{n,k} - b_n^\delta\right)$. This results in a total of $2k_n$ required solutions for the entire execution of an inner iteration. In the first iteration however, the calculation of $A_n s_{n,0}$ can be avoided because $s_{n,0} = 0$. But, if the inner iteration is not terminated by the maximal number $k_{\max,n}$, the additional derivative $A_n s_{n,k_n}$ needs to be calculated only to verify the stop criteria (3.5) in the last iteration, which leave us with the same number $2k_n$. The calculation of the last derivative is not necessary however, if the maximal number of inner iterations is reached. In this case, the inner iteration calculates the solution of $2k_{\max,n} - 1$ variational problems. Since the evaluation of the Jacobian matrix demands a total of L solutions of (5.37), we conclude that it is *not* worth to calculate this matrix if $2k_{\max,n} - 1 \leq L$ or $2k_n \leq L$ for the cases when the maximal number of inner iterations is reached or not reached respectively. But, as an a priori estimation for the number k_n is not available, it is difficult to decide whether the Jacobian should be calculated. Note however, that if $k_{\max} \leq L/2$, then both inequalities are satisfied and therefore the calculation of the Jacobian is more expensive. See a detailed comparison in the Tables 5.2 and 5.3 in the final of Subsection 5.2.2 below.

5.2.2 Computational implementation

To implement our experiments, we define Ω as the circle centered in the origin and with radius equals $4/\pi$. To reconstruct the conductivity, we use the same triangulation Υ of Ω defined in the last subsection, which is also used to calculate the Jacobian matrix of the forward operator. Since an analytical solution of (5.23) is in general not available, the FEM is used to find an approximate solution. For this approximate solution however, a different and more refined triangulation Θ of Ω is used[7]. Further, we fix $d = 1$ (non-Kaczmarz method), which means that we analyze the problem defined in (5.34).

In [35], Lechleiter and Rieder have proven that once Υ is fixed, there exists a number $L_{\min} \in \mathbb{N}$ dependent on Υ, such that if the number of electrodes satisfies $L \geq L_{\min}$ then the Fréchet-derivative of the (discrete) forward operator (5.34) is injective and satisfies the Tangential Cone Condition (Assumption 1(c), page 31) in a small ball centered on an arbitrary element $\gamma \in \operatorname{int}(\mathrm{D}(F))$. Further, the same result remains true if the vector $(u, U) \in H^1(\Omega) \oplus \mathbb{R}^L_\Diamond$, which is the exact solution of (5.23), is changed by its FEM-approximation (u_Θ, U_Θ) in the mesh Θ whenever this triangulation is sufficiently fine, see [35, Theo. 4.5 and 4.9]. In particular, the inequality $\dim \Upsilon \leq L(L-1)/2$ is a necessary condition for the injectivity of F'.

Once fixed the meshes, the forward operator (5.34) is well-defined and acts between finite dimensional spaces and since all the norms are equivalent in such spaces, the derivative of the forward operator does not change if the norms in X and Y are modified. As explained in the last subsection, we choose the norm in X to be the L^p-norm and denote this space by $X := V_p$, see (5.33) above. We now transform the vector space $Y = \mathbb{R}^{\ell L}$ in a Banach space, equipping it with the norm

$$\|U\|_r := \left(\sum_{i=1}^{\ell L} |U_i|^r\right)^{1/r}, \ U \in \mathbb{R}^{\ell L}, \tag{5.39}$$

where $r > 1$ is a fixed number. A similar calculation to that made in Example 17 shows that the duality mapping is given by

$$(J_r(U))_i = |U_i|^{r-1} \operatorname{sgn}(U_i), \ i = 1, ..., \ell L, \tag{5.40}$$

where

$$J_r(U) := \left(J_r(U)_1, ... J_r(U)_{\ell L}\right)^\top, \ U \in Y.$$

In our experiments, we have chosen Υ and Θ with $\dim \Upsilon = 1200$ and $\dim \Theta = 14340$. We highlight the fact that the use of the relatively large dimension of the triangulation Υ implies that the inequality $\dim \Upsilon \leq L(L-1)/2$, necessary for the injectivity of F', is only verified for $L \geq 50$. However, most experiments we have performed in this subsection have a moderate number of electrodes, which is in general much smaller than 50. The result is that a solution of an under-determined system

$$A_n s = b_n \tag{5.41}$$

needs to be approximated in each inner iteration of `K-REGINN`. Among all the possible solutions we could approximate, we then pick up a specific one which is the most suitable to our interests. We follow ideas from [52] and select an appropriate weight-function $\omega \colon \Omega \to \mathbb{R}$, changing the L^p-norm in X with L^p_ω, as done in the last section. In [52], Winkler and Rieder have proposed choosing a different weight-function $\omega = \omega_n$ for each (outer) iteration. They argument that an usual choice for the solution of (5.41) in the Hilbert spaces is

$$s := \underset{\overline{s} \in N(A_n)^\perp}{\arg\min} \|A_n \overline{s} - b_n\|,$$

[7] For a full explanation of how to employ the FEM in order to find a solution of (5.23) we recommend [42].

where the condition $\overline{s} \in N(A_n)^{\perp}$ resolves under-determinedness. The author's idea is to define a suitable piecewise constant and positive weight

$$\omega_n := \sum_{i=1}^{M} \omega_{n,i}\chi_{T_i},$$

such that the weighted inner product $\langle \cdot, \cdot \rangle_{\omega_n} := \langle \cdot, \omega_n \cdot \rangle_{L^2(\Omega)}$ substitutes the original L^2 inner product and then redefine

$$s := \underset{\overline{s} \in N(A_n)^{\perp_{\omega_n}}}{\arg\min} \ \|A_n \overline{s} - b_n\|, \tag{5.42}$$

where $N(A_n)^{\perp_{\omega_n}}$ represents the orthogonal complement of $N(A_n)$ with the new inner product $\langle \cdot, \cdot \rangle_{\omega_n}$. They have suggested an strategy which updates indistinguishable coefficients of the current (outer) iterate $\gamma_n := \sum_{i=1}^{M} \gamma_{n,i}\chi_{T_i}$ with the same amount as following: the definition of the coefficients

$$\omega_{n,i} := \frac{\|S_i\|_2}{|T_i|}\gamma_{n,i}^{-1}, \text{ for } i = 1, ..., M, \tag{5.43}$$

where $|T_i|$ is the area of the triangle T_i and

$$\|S_i\|_2 := \sqrt{\sum_{j=1}^{\ell} \left\| F_j'(\gamma_n)\chi_{T_i} \right\|_2^2}$$

is the $2-$norm of the $i-$th column of the Jacobian matrix $A_n = F'(\gamma_n)$, results in a weight-function ω_n which provides a solution $s = \sum_{i=1}^{M} s_i\chi_{T_i}$ of (5.42) whose coefficients s_i and s_j are proportional to the local updates $\gamma_{n,i}$ and $\gamma_{n,j}$ whenever the columns S_i and S_j of the Jacobian matrix A_n are linearly dependent. More precisely, in [52, Theo. 1], it is proven that if the coefficients of ω_n are defined like in[8] (5.43), then the equality

$$\frac{s_i}{\gamma_{n,i}} = \operatorname{sgn}(\beta)\frac{s_j}{\gamma_{n,j}}$$

is guaranteed whenever the condition

$$S_j = \beta S_i, \ \beta \in \mathbb{R}\setminus\{0\}$$

holds true.

Remark 50 *Since γ_0 was chosen as a constant in the last section, the weight function ω used in that section is just given by the first weight ($n = 0$ in Definition* (5.43) *above) times a constant, this is, $\omega = \gamma_0\omega_0$, see* (5.18).

Note that the use of the weighted L^2 inner product $\langle f, g \rangle_{\omega_n} = \int_\Omega fg\omega_n$ results in the weighted space $X_n = V_{2,\omega_n} = \left(V, \|\cdot\|_{L^2_{\omega_n}(\Omega)}\right)$. Unfortunately, our framework does not allow the use of different spaces X in different iterations and for this reason we proceed like in the last section and fix the same weight-function for all iterations. We therefore define

$$\omega := \sum_{i=1}^{M} \omega_i\chi_{T_i}$$

[8] The authors of [52] actually proposed the use of a slightly different weight-function $\overline{\omega}_n$, whose coefficients $\overline{\omega}_{n,i}$ correspond to $|T_i|\,\omega_{n,i}$ in (5.43). But, since they have used the Euclidean inner product in $\mathbb{R}^M$ in place of the L^2 inner product, both choices result in equivalent approaches.

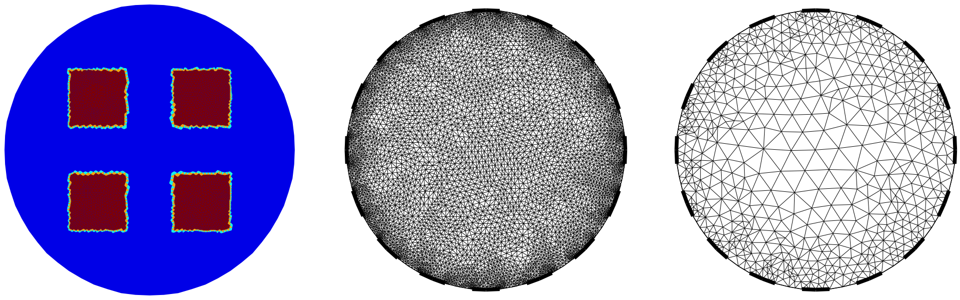


Figure 5.10: Left: example of a conductivity γ^+, modeled by a background (in blue) and a sparsely distributed inclusion (in dark red). Middle and right: Triangulations Θ and Υ respectively.

with

$$\omega_i := \frac{\sqrt{\sum_{j=1}^{\ell} \left\| F_j'(\gamma_0)\, \chi_{T_i} \right\|_2^2}}{|T_i|} \gamma_{0,i}^{-1}. \tag{5.44}$$

Even not having any results for Banach spaces, we use this weight-function in all the experiments. In particular, the change of the space $X = V_p$ with $X = V_{p,\omega}$ results in the new norm

$$\|f\|_{L^p_\omega(\Omega)} = \left\| f\omega^{1/p} \right\|_{L^p(\Omega)}, \ f \in X.$$

Further, this modification in X implies a slight modification in the way the adjoint operator of the derivative is calculated. In this new space, the adjoint A_n^* of the Jacobian matrix is not only the transpose of the matrix A_n, but the transpose of the weighted matrix $A_{n,\omega}$, defined multiplying each element in the $i-$th row of the matrix A_n by ω_i. Indeed, for all $u \in \mathbb{R}^M$ and $v \in \mathbb{R}^{\ell L}$,

$$\left\langle u, (A_{n,\omega})^\top v \right\rangle_Y = \langle A_{n,\omega} u, v \rangle_{V_p} = \langle \omega A_n u, v \rangle_{V_p} = \langle A_n u, v \rangle_{V_{p,\omega}} = \langle u, A_n^* v \rangle_Y ,$$

resulting in $A_n^* = (A_{n,\omega})^\top$.

We point out that no additional computational effort is required to calculate the weights in (5.44) since all the needed information have already been calculated in order to evaluate the Jacobian matrix $A_0 = F'(\gamma_0)$, necessary to perform the first inner iteration.

Since the weighted-space $V_{p,\omega}$ is always used in our experiments, we skip the dependence of V_p on ω, denoting $V_p = \left(V, \|\cdot\|_{L^p_\omega(\Omega)} \right)$ from now on.

In each experiment we have performed, all the electrodes have the same measure, are uniformly distributed on the boundary of Ω and cover 50% of $\partial\Omega$. We have fixed $L = \ell = 16$ for the first experiments and the current patterns in (5.26) are defined as the vectors $I^i := (0, ..., 0, 1, -1, 0, ..., 0)^\top$, with 1 in the $i-$th coordinate, -1 in the immediately following one and zero elsewhere. The contact impedances are the known constant $z_j = 0.04/\pi$ for $j = 1, ..., L$, which is 1% of the radius of Ω.

The exact solution γ^+ we are looking for has the form shown in (5.15). However, different inclusions $B \subset \Omega$ are considered in different experiments. The first iterate $\gamma_0 \equiv 0.1$ matches the background again and the relative iteration error e_n as well as the error reconstruction is defined in the same way as (5.16). In all the tests we have realized, the data Γ_γ in (5.27) has been synthetically generated using a very fine mesh with more than 230000 triangles.

Figure 5.10 illustrates an example of γ^+, where the inclusion B is modelled by a four squares-like inclusion. The triangulations Θ and Υ, used respectively to solve the elliptical problem (5.23) and to reconstruct the conductivity, are also exhibited.

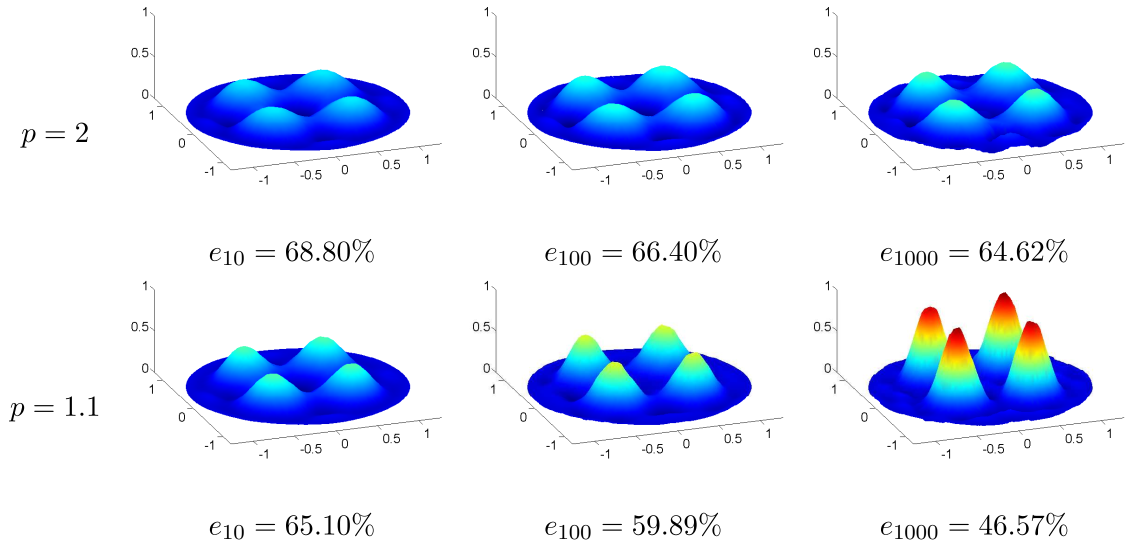


Figure 5.11: Convergence in the noiseless situation observed in two different Banach space norms. The first and second rows show the reconstructions obtained using the Hilbert space $X = V_2$ and the Banach space $X = V_{1.1}$ respectively.

To confirm the important noiseless convergence result, proved in Theorem 43, we add no artificial noise in the generated data ($\delta = 0$), and see what happens to the approximation γ_n as n grows to infinity. For this first experiment, the IT method (3.62) has been employed as inner iteration of `REGINN` and the results, which have been built in the spaces $X = V_2$ ($p = 2$) and $X = V_{1.1}$ ($p = 1.1$), are respectively displayed in the first and second row of Figure 5.11. The Euclidean norm $\|\cdot\|_2$ (Definition (5.39) with $r = 2$) is used to transform the vector space Y in a Hilbert space. With this configuration, the duality mapping in (5.40) is just the identity operator. The parameter α_n in (3.68) is given by the constant value 0.1 and the dual gradient DE method (see (3.29) and (3.39)) with $C_1 = C_2 = 0.1$ is applied to find the minimizer of the functional (3.63), which is necessary to realize the inner iteration. The other parameters of `REGINN` are $\mu = 0.8$ and $k_{\max} = 10$. Each picture displayed in Figure 5.11 is actually a linear interpolation of the coefficients of the piecewise constant reconstructions. All these pictures are in the same scale of colors and below each of them, the relative error iteration e_n is shown. The first column illustrates the iterate γ_{10} while the second and third one refers to γ_{100} and γ_{1000} respectively. As expected, the reconstructions in the Hilbert space $X = V_2$ are over-smoothed, with a relative large oscillation in the background and too low inclusions. In the other hand, the second row clearly shows more thin and high inclusions with larger slopes and a lower oscillation levels in the background, which is a typical behavior of solutions restored in the Banach space norms L^p for small values of p, see e.g. [12].

In the next experiment we use artificially generated noise to contaminate the data Γ_γ in (5.27) with the relative noise level δ :

$$\Gamma_\gamma^\delta = \Gamma_\gamma + \delta \left\|\Gamma_\gamma\right\|_2 per,$$

where the perturbation vector $per \in \mathbb{R}^{L^2}$ is an uniformly distributed random variable with $\|per\|_2 = 1$. We fix $\delta = 0.1\%$ and compare the performance of `REGINN` for reconstructing sparsely located inclusions when different norms in X are employed to this task. We use $p = 2$ and $p = 1.01$. The parameter $k_{\max}$ is now set to 500 and a mixed gradient-Tikhonov method is used as inner iteration, combining the Tikhonov-Phillips method with the dual gradient DE method ($\overline{x}_n = 0$ in (3.76) with $\lambda_{n,k} = \overline{\lambda}_{DE}$ being defined in (3.72) with $C_1 = C_2 = 0.1$). The constant τ has the value 1.5 and the other parameters are the same as

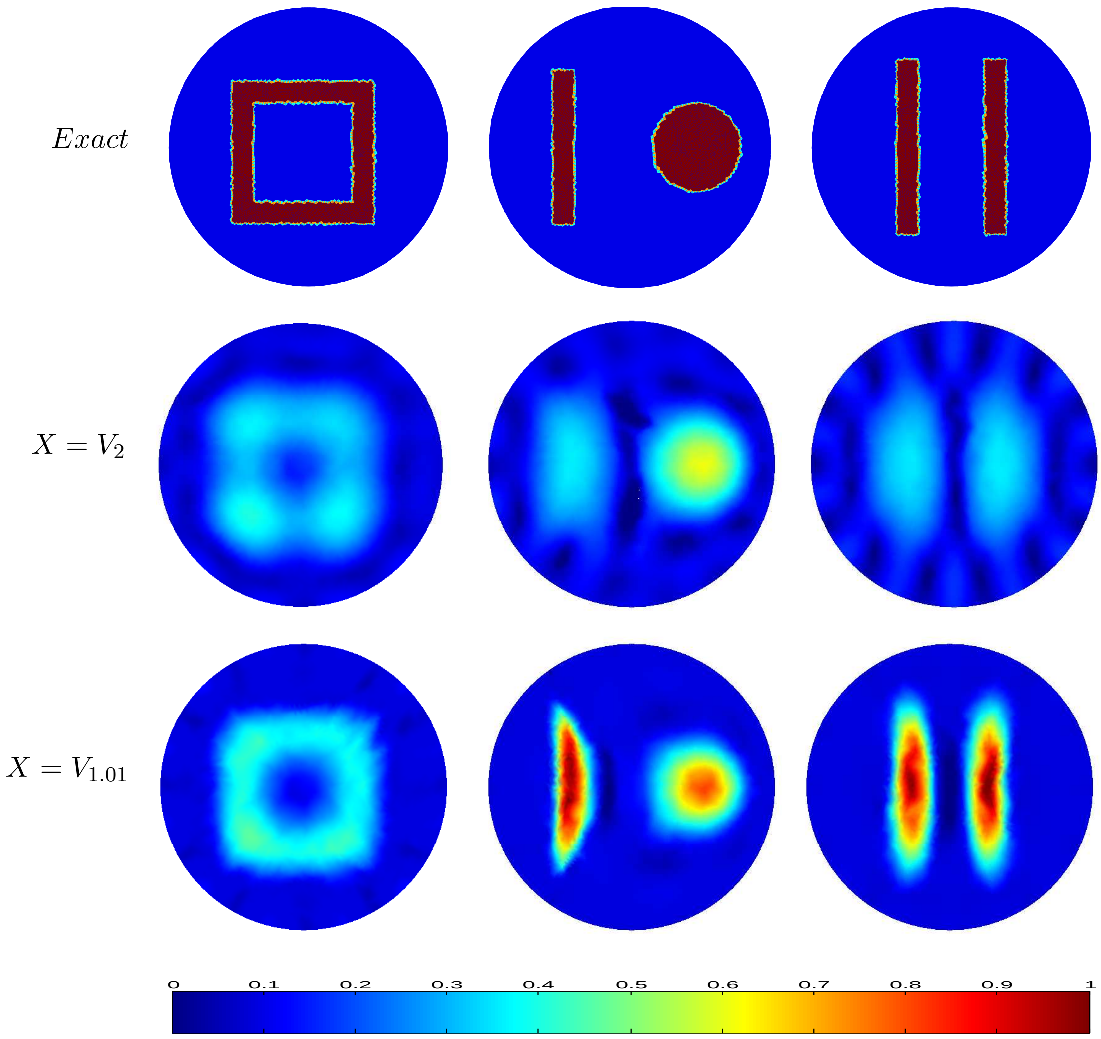


Figure 5.12: Conductivity with sparsely located inclusions (first row) and the respective reconstructed solutions: using the Hilbert space norm L^2 (second row) and the Banach space norm $L^{1.01}$ (third row).

in the last experiment. Figure 5.12 presents the searched-for solutions γ^+ in the first row and exhibits in the second and third rows, the respective V_2- and $V_{1.01}-$reconstructions. All the pictures are in the same scale of colors and after the third row, a color bar is displayed. We clearly see a lower level of oscillation in the background for the reconstructions in the third row. Additionally, a doubtless improvement in the reconstructions is achieved if the Hilbert space V_2 is replaced with the Banach space $V_{1.01}$, resulting in sharper inclusions with more accurate values, forms and locations.

In order to examine the behavior of `REGINN` for reconstructing a sparse conductivity together with impulsive noise[9], we choose different norms in both X and Y. The values $p = 2$ and $p = 1.01$ as well as $r = 2$ and $r = 1.01$ have been designated to perform this task. Since $d = 1$ (non-Kaczmarz method), inequality $s \leq r$ in Theorem 47 is an unnecessary condition and therefore, the index r of the duality mapping in Y can be freely chosen. Figure 5.13 illustrates the searched-for conductivity γ^+, modelled by a ring-like inclusion,

[9] Impulsive noise in the context of the Complete Electrode Model corresponds to a low level error in most electrodes in the measure of the voltage for a fixed current pattern, but with very high error levels in some few of them.

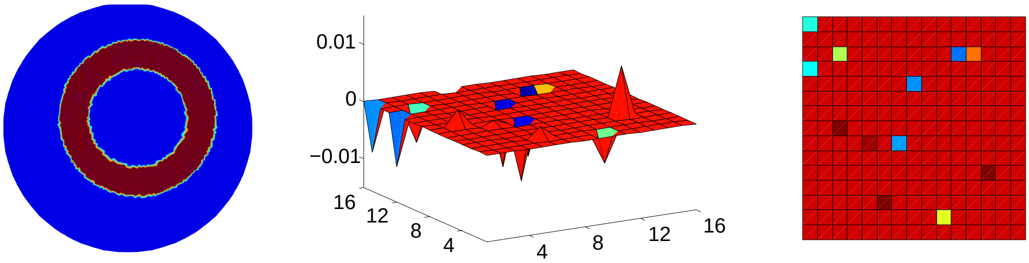


Figure 5.13: Left: the searched-for sparsely distributed conductivity γ^+. Middle and right: impulsive noise $\Gamma_\gamma^\delta - \Gamma_\gamma$.

followed by the noise $\Gamma_\gamma^\delta - \Gamma_\gamma$ (see (5.27)) displayed in two different angles[10] and with noise level $\delta = 1\%$.

Figure 5.14 complements Figure 5.13 and shows the reconstructions in the above referred framework. The dual gradient DE method ((3.29) and (3.39) with $C_1 = C_2 = 0.1$) is used as inner iteration of `REGINN` to generate the results. The remaining parameters of `REGINN` are the same as in the last experiment. In the first and second rows we present the results found using the Hilbert space $X = V_2$ and the Banach space $X = V_{1.01}$ respectively. The first column shows the reconstructions when $r = 2$ in the norm (5.39) is used, while the second column displays the results for $r = 1.01$. Below each picture, the number of outer iterations $N = N(\delta)$ as well as the reconstruction error e_N and the overall number of inner iterations k_{all} are shown. Further, a color bar is presented on the right. It is clear that the pictures in the last column have a superior quality than the ones in the first column, which means that the norm $\|\cdot\|_{1.01}$ is more appropriated to deal with the impulsive noise than the standard Euclidean norm $\|\cdot\|_2$. The price to be paid for better reconstructions is a significantly increase in the overall number of inner and outer iterations. Observe however, that a small value for p produces an additional effect of less variation in the background in the same time that it demands less computational effort until convergence. Figure 5.14 makes clear that the combination $p = r = 1.01$ results in a very satisfactory framework for this specific situation.

To finish this chapter, we confront the non-Kaczmarz version of `REGINN` (5.34) (where the case $d = 1$ is considered) with its corresponding Kaczmarz version (5.35), where $d = \ell$ is used (i.e., each single current pattern in (5.26) results in a different equation used by `K-REGINN`). The goal is to compare the reconstruction errors and the computational effort necessary to perform an entire run of Algorithm 1. For the case $d = 1$, it is only considered the case where the Jacobian is calculated because this is always the most advantageous situation (see the discussion at the end of last subsection). Further, the computational effort necessary to perform all the iterations using or not using the calculation of the Jacobian matrix is compared for the Kaczmarz version. Different numbers of electrodes, and respectively of current patterns, are used: $L = \ell$ ($= d$ for the Kaczmarz version) with $L = 8$, $L = 16$, $L = 32$ and $L = 64$. Trying to be as fair as possible in the comparisons, we have not used any weight-function in the L^p-norms of X. The constant τ has been chosen as the smallest constant such that Algorithm 1 terminates in all cases[11]. We found the values $\tau = 1.3$ for the non-Kaczmarz version and $\tau = 3$ for the Kaczmarz version as the optimal values. The conductivity function γ^+ is the one shown in the intersection of

[10]The vector-noise $\Gamma_\gamma^\delta - \Gamma_\gamma \in \mathbb{R}^{L^2}$ is actually displayed in Figure 5.13 in the matrix form, where the $j-$th column corresponds to the vector $U^{j,\delta} - U^j \in \mathbb{R}^L$ and generated from the current pattern I^j in (5.26).

[11]The constant η in the TCC can be significantly different for the problems (5.34) and (5.35) if the number of experiments $\ell \in \mathbb{N}$ is large. Since the constant τ depends on η, see (4.9), it can be different for the Kaczmarz and non-Kaczmarz versions.

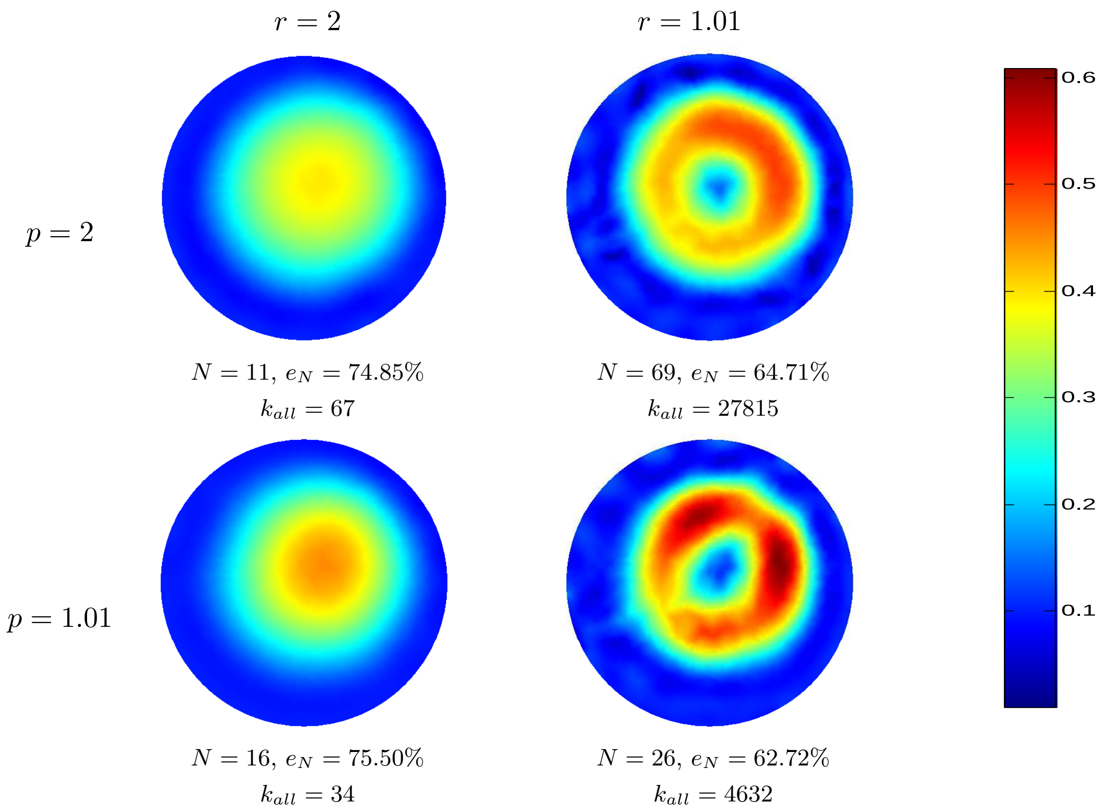


Figure 5.14: Conductivity function γ^+ from Figure 5.13, reconstructed in different Banach spaces with 1% of impulsive noise. In the first column $Y = \left(\mathbb{R}^{L^2}, \|\cdot\|_2\right)$ and $Y = \left(\mathbb{R}^{L^2}, \|\cdot\|_{1.01}\right)$ in the second one. The first and second rows correspond to the spaces $X = V_2$ and $X = V_{1.01}$ respectively.

the first row and the first column of Figure 5.12. For the inner iteration, the dual gradient DE method with $C_1 = C_2 = 0.1$ is used again. Further, all the tests have been made with the same noise level $\delta = 0.1\%$ and with $p = 1.1$. Since a large constant $k_{\max}$ represents an extra advantage for the case where the Jacobian matrix is calculated in comparison with the case where this matrix is not calculated, this constant must be carefully chosen. We fixed it with the value 50, which we consider a good number. All the other parameters are the same as in the last experiment. The results of this experiment are collected in the Tables 5.2 and 5.3. In the first table, the number of outer iterations $N(\delta)$ as well as the overall number of inner iterations k_{all} and the reconstruction error $e_{N(\delta)}$ are displayed for each electrode configuration.

As already discussed at the end of last subsection, the regular version of `REGINN` requires

L		08	16	32	64
$N(\delta)$	Regular	39	149	109	561
	Kaczmarz	558	28174	33111	202499
k_{all}	Regular	3700	26294	14788	18643
	Kaczmarz	2003	103227	93392	535148
$e_{N(\delta)}$	Regular	79.20	70.38	67.10	65.48
	Kaczmarz	79.08	69.16	65.78	64.32

Table 5.2: Comparison between the Regular version of `REGINN` and its Kaczmarz version with different numbers L of electrodes.

L		08	16	32	64
AI		331	14954	18093	77693
	Regular	616	4752	6944	71744
CE	Kaczmarz Jac	3206	267438	612087	5174851
	Kaczmarz	4564	234628	219895	1272795

Table 5.3: Computational effort CE (overall number of required solutions of variational problems for an entire run of K-REGINN). The row denoted by "Regular" represents the case $d = 1$ (non-Kaczmarz version) with the evaluation of the Jacobian matrix. The rows denoted by "Regular Jac" and "Regular" represent the Kaczmarz versions ($d = L$) when the Jacobian matrix is respectively evaluated and not evaluated.

the solution of L variational problems in each outer iteration in order to calculate $F(\gamma_n)$. Each inner iteration in turn, requires the solution of L additional problems for the evaluation of the Jacobian matrix. Since the last outer iteration is the unique iteration which does not perform an inner iteration, we conclude that the overall number of variational problems which need to be solved in an entire run of Algorithm 1 is $(2N - 1)L$. For the Kaczmarz version of REGINN, 1 solution of a variational problem is required in each outer iteration in order to calculate $F_{[n]}(\gamma_n)$. If the Jacobian is calculated, L additional solutions are necessary in each *active iteration*[12], which results in a total of $N + L.AI$, where the number AI represents the overall number of active iterations. On the other hand, each vector calculated in the inner iteration of K-REGINN demands the solution of approximately 2 variational problem if the Jacobian matrix is not calculated, which results in a total of $N + 2k_{all}$ required solutions.

Turning again to the results, we see that a small improvement in the reconstruction error $e_{N(\delta)}$ is shown in the Table 5.2 if the Kaczmarz version of REGINN is considered. The price to be paid is however high, as Table 5.3 makes evident. This table compares the computational effort necessary to achieve the results. It shows the overall number of active iterations AI and the "computational effort" CE for each electrode configuration[13]. Further, the regular version of REGINN (where $d = 1$ and the Jacobian matrix is calculated) and the Kaczmarz versions (where $d = L$) with and without the calculation of the Jacobian matrix are compared. It is clear that the regular version of REGINN always works less than its Kaczmarz versions in all situations. Comparing only the Kaczmarz versions, we see that the calculation of the Jacobian matrix demands less computational effort only for a small number of electrodes. If a relative large number of electrodes is used, the computational effort necessary to calculate the Jacobian matrix becomes higher than the direct calculation of the solutions of (5.23) and (5.30) and the evaluation of this matrix is not worth any longer.

[12]An active iteration is an outer iteration where the discrepancy principle (3.9) is not satisfied and consequently, the inner iteration needs to be performed.

[13]The number CE actually represents the overall number of required solutions of variational problems used by Algorithm 1 until termination.

Chapter 6

Conclusions and Final Considerations

We consider the convergence analysis of Chapter 4 the main accomplishment of this work. In this chapter, a Kaczmarz version of the inexact-Newton method `REGINN` [43] is analyzed in Banach spaces and the proofs are carried out considering an inner iteration defined in a relatively general way, which fits in with various methods. The result is a convergence analysis of `K-REGINN`, valid at the same time for several different regularization methods in the inner iteration. This analysis is a generalization to Banach spaces and to Kaczmarz methods of ideas previously discussed in [36]. In order to properly develop this general convergence analysis however, strong restrictions needed to be imposed. We think it is possible to weaken these hypotheses, especially those required on the solution space X. For instance, in the cases where the non-Kaczmarz version ($d = 1$ in (3.6)) is observed, the first inequality in (4.30) becomes unnecessary to prove convergence in the noiseless situation. But, this inequality seems to be exactly the crucial point, where the $s-$convexity of X is most needed. In the remaining cases, the $s-$convexity is apparently preventable and could be handled with similar techniques to those engaged in [47]. Assuming the condition $d = 1$ therefore, the uniform smoothness and $s-$convexity of X could be weaken to only smoothness and uniform convexity. Of course without the $s-$convexity, the verification of Assumption 3, page 58, for the methods investigated in Section 3.1 would be much more arduous.

In our opinion, the uniform smoothness of data space Y, required to show the stability property of the dual gradient methods in Appendix A, could also be avoided using similar arguments to those employed in [47]. However, in order to perform this modification, the proof of the stability property should be made simultaneously with the proof of the regularization property, which means that, at least in principle, Assumption 6, page 70, could not be proven separately, and consequently the general convergence analysis could be harmed.

The Tikhonov-Phillips method (3.58) has a peculiar iteration form, which is somewhat different from the other methods of Section 3.1. This characteristic complicates the convergence analysis of Chapter 4 substantially, forcing the recurrent separation in two cases: $\widehat{z}_{n,k} = z_{n,k}$ and $\widehat{z}_{n,k} = x_n$, see Assumptions 3, and 4, page 62. But looking on the bright side, none geometrical property on the space Y is required for this method in order to prove Assumption 6, see the stability proof in Appendix A. This suggests that the space L^1 could be used as data space in further numerical experiments, building up a more appropriate framework to deal with impulsive noise for example. If $Y = L^1$ however, although the functional (3.60) remains strictly convex, it is not differentiable any longer and the process of finding its minimizer, necessary to perform the inner iteration, becomes much more challenging.

At the end of Subsection 3.1.2, the possibility of obtaining the optimal step-size of the dual gradient methods is discussed and a lower bound for this step-size is shown. Though an explicit formula for the optimal step-size is hard to achieve, sharper bounds could possibly be determined. An upper bound like $\lambda_{opt} \lesssim \lambda_{DE}$ for instance, would be especially interesting and could facilitate the verification of Assumption 3, including the gradient method associated with the optimal step-size in our convergence analysis of Chapter 4.

A similar situation is observed with the dual gradient Steepest Descent method. Finding an explicit formula for λ_{SD} in (3.44) involves the differentiation of the duality mapping, which makes difficult the determination of this step-size. To overcome this technical obstacle, we have replaced this method with its similar version, the Modified Steepest Descent method (3.45) , which already has an explicit step-size satisfying the desired inequality $\lambda_{MSD} \leq \lambda_{DE}$, necessary to include it in our convergence analysis. Although an obvious formula for λ_{SD} is not available, it is not unconditionally necessary to prove Assumption 3 since only the inequality $\lambda_{SD} \leq \lambda_{DE}$ needs to be verified. Once proven, Assumption 3 would imply the convergence of Steepest Descent method. The implicit defined step-size λ_{SD} could be numerically approximated with the help of an optimization algorithm in the real-line.

In addition to several classical regularization techniques, which have been adapted from Hilbert to more general Banach spaces in this work, a few novel approaches are presented. Among the methods which have been firstly introduced in this thesis, we highlight the mixed gradient-Tikhonov methods (Subsection 3.1.4) and the dual gradient Decreasing Error method ((3.29) and (3.39)). Both algorithms have shown themselves useful methods to reconstruct stable solutions of ill-posed problems as can be seen in Figures 5.12 and 5.14 for example. The additional (regularization) term of the mixed gradient-Tikhonov methods confers further stability to regular gradient methods and results in a more steady inner iteration. The Decreasing Error method in turn, exhibits the advantage of playing the role of the gradient method with the fastest decreasing error in the case of a linear problem in a Hilbert space been observed. This behavior seems to be somehow transmitted to nonlinear problems in Banach spaces as Figure 5.3 shows.

The nonlinearity of the duality mapping is possibly the biggest complication in the convergence analysis of `K-REGINN` in general Banach spaces. The extra effort to demonstrate the theorems in these more complicated spaces is however counterbalanced with the improvements provided by the use of more convenient norms. The use of Banach spaces can be therefore lucrative in order to achieve better reconstructions of inverse problems in some specific situations, especially if sparsity constraints on the searched-for solution or on the data-noise are present. This enhancement of the quality becomes evident in the numerical experiments performed in Chapter 5, however, many questions concerning the mathematical modelling of EIT problem with L^p-norms remain still open. In its original (infinite dimensional) version, differentiability or even continuity of the forward operator with the Lebesgue spaces L^p are in principle not valid for any $p \in (1, \infty)$ and legitimizing the use of these spaces seems not to be a trivial issue. For a proper adaption, further research is therefore needed.

A superficial examination on the computational implementations of Subsection 5.2.2 and particularly on Tables 5.2 and 5.3 could erroneously lead to the conclusion that the Kaczmarz version of `REGINN` is not competitive. These results however, are in large part due to the manipulation described in (5.38) , which reduces considerably the computational effort necessary to calculate the Jacobian matrix, especially for the non-Kaczmarz version of `REGINN`. But without this strategy, the regular `REGINN` would be much more "expensive" and its Kaczmarz version could become more advantageous. Thus, to have a more accurate answer about the real utility of the Kaczmarz version of `REGINN`, supplementary numerical experiments, preferably with a different problem, should be performed.

To finish this work, we want to compare all the algorithms introduced in Section 3.1. The primal gradient methods from Subsection 3.1.1 are the easiest to implement and are from far the methods with least prerequisites on space X. But they require strong hypotheses on Y and only termination of (regular) `REGINN` can be shown if such a method is used in the inner iteration. The dual gradient methods from Subsection 3.1.2 on the other hand, can be employed as inner iteration of `REGINN` to show convergence in the noiseless case and the regularization property when only noisy data is available. These results hold true even for the Kaczmarz version of `REGINN` if the Landweber method is considered. However, the iteration (3.29) needs to be performed using the current (outer) iterate x_n, which makes this version of the gradient methods a little more complicated. Additionally, the requirements on the spaces X and Y constitute a considerable disadvantage. Among all the methods introduced in Section 3.1, these are the ones which demand the most restrictive Banach spaces to work: uniformly smoothness of X and Y and $p-$convexity of X are required.

Tikhonov methods own the convenient property of working without any restrictions on space Y, as shown in Subsection 3.1.3 and Appendix A. Moreover, convergence and the regularization property of `K-REGINN` can be proven whenever either the Iterated-Tikhonov or the Tikhonov-Phillips method is the chosen inner iteration. However, the obligation of solving an optimization problem in order to generate a new vector in the inner iteration represents a big obstacle: the more poor convexity/smoothness properties the space Y possesses, the more difficult this optimization problem becomes. To combine the advantages of dual gradient and Tikhonov methods, we have introduced in Subsection 3.1.4 the mixed gradient-Tikhonov methods. Though this kind of algorithm necessitates the same (strong) requirements on the spaces X and Y as the dual gradient methods, it confers extra stability to the inner iteration by incorporating a regularization parameter α_n on its iteration, see (3.76) . Further, it does not request the solution of any optimization problem, although similarly to Tikhonov methods, a suitable regularization parameter α_n needs to be determined, which in the particular case of mixed methods, depends on a priori information about the norm of the solution, see (3.71) .

Appendix A

Stability Property of the Regularizing Sequences

This appendix provides for different methods, the proof of the stability property given in Assumption 6, page 70. For the ease of the presentation, we assume all the hypotheses of Theorem 47 except for Assumption 6 itself. For some methods, other hypotheses are necessary to complete the proof and these additional hypotheses will be required at the specific points where they are needed.

Let $(\delta_i)_{i\in N} \subset \mathbb{R}$ be a zero-sequence and suppose that $x_n^{\delta_i} \to \xi \in \mathcal{X}_n$ as $i \to \infty$. We proceed giving a proof by induction: for $k = 0$, $z_{n,0}^{\delta_i} = x_n^{\delta_i} \to \xi = \sigma_{n,0}(\xi)$ as $i \to \infty$. Assume now that $\lim_{i\to\infty} z_{n,k}^{\delta_i} = \sigma_{n,k}(\xi)$ for $k < k_{REG}(\xi)$. Our task is now to prove that $z_{n,k+1}^{\delta_i} \to \sigma_{n,k+1}$ as $i \to \infty$.

- Dual gradient methods DE, LW and MSD

  For these methods we have $\widehat{z}_{n,k}^{\delta_i} = z_{n,k}^{\delta_i}$ and $v_{n,k}^{\delta_i} = A_n^{\delta_i} s_{n,k}^{\delta_i} - b_n^{\delta_i}$. Accordingly, $\widehat{\sigma}_{n,k} = \sigma_{n,k}$.

  This proof is a slightly different version of that given in [40, Lemma 10]. Assume that the Banach spaces Y_j, $j = 0, ..., d-1$ are uniformly smooth. As the functions F_j and F_j' are continuous, using the induction hypothesis and (3.7) it is clear that the vector

$$v_{n,k}^{\delta_i} = A_n^{\delta_i} s_{n,k}^{\delta_i} - b_n^{\delta_i} = F_{[n]}'\left(x_n^{\delta_i}\right)\left(z_{n,k}^{\delta_i} - x_n^{\delta_i}\right) - \left(y_{[n]}^{\delta_i} - F_{[n]}\left(x_n^{\delta_i}\right)\right)$$

  converges to

$$v_{n,k}^{\xi} = F_{[n]}'(\xi)\left(\sigma_{n,k}(\xi) - \xi\right) - \left(y_{[n]} - F_{[n]}(\xi)\right)$$

  as $i \to \infty$. Similarly, $\lambda_{n,k}^{\delta_i} \to \lambda_{n,k}^{\xi}$ as $i \to \infty$ (see (3.39), (3.43) and (3.45)). As the spaces Y_j's are uniformly smooth, the selection $j_r : Y_j \to Y_j^*$ is unique and continuous and since the mappings J_p, F_j and F_j' are continuous too,

$$J_p\left(z_{n,k+1}^{\delta_i}\right) = J_p\left(z_{n,k}^{\delta_i}\right) - \lambda_{n,k}^{\delta_i} F_{[n]}'\left(x_n^{\delta_i}\right)^* j_r\left(A_n^{\delta_i} s_{n,k}^{\delta_i} - b_n^{\delta_i}\right) \tag{A.1}$$

  converges to

$$J_p\left(\sigma_{n,k+1}\right) = J_p\left(\sigma_{n,k}\right) - \lambda_{n,k}^{\xi} F_{[n]}'(\xi)^* j_r\left(v_{n,k}^{\xi}\right), \tag{A.2}$$

  as $i \to \infty$. This means that $\lim_{i\to\infty} z_{n,k+1}^{\delta_i} = \sigma_{n,k+1}$ because the duality mapping $J_p^{-1} = J_{p^*}^*$ is also continuous.

- Bregman variation of Iterated-Tikhonov method (3.62)

  Here $\widehat{z}_{n,k}^{\delta_i} = z_{n,k}^{\delta_i}$ and $v_{n,k}^{\delta_i} = A_n^{\delta_i} s_{n,k+1}^{\delta_i} - b_n^{\delta_i}$.

  This proof was first presented in [39] in its current version. It is an adaptation from [26, Lemma 3.4], which in turn uses ideas from [16]. The stability proof in this case is more complicated because we do not assume any condition on the Banach spaces Y_j's.

  See that the vectors $z_{n,k+1}^{\delta_i}$ and $\sigma_{n,k+1}$ respectively minimize the functionals

$$T_{n,k}^{\delta_i}(z) := \frac{1}{r}\left\|F'_{[n]}\left(x_n^{\delta_i}\right)\left(z - x_n^{\delta_i}\right) - b_n^{\delta_i}\right\|^r + \alpha_n \Delta_p\left(z, z_{n,k}^{\delta_i}\right)$$

and

$$W_{n,k}(z) := \frac{1}{r}\left\|F'_{[n]}(\xi)(z-\xi) - \widetilde{b}_n\right\|^r + \alpha_n \Delta_p(z, \sigma_{n,k}),$$

where $\widetilde{b}_n := y_{[n]} - F_{[n]}(\xi)$. As the family $\left(z_{n,k+1}^{\delta_i}\right)_{\delta_i > 0}$ is uniformly bounded (see (4.16)) and X is reflexive, there exists, by picking a subsequence if necessary, some $\overline{z} \in X$ such that $z_{n,k+1}^{\delta_i} \rightharpoonup \overline{z}$ as $i \to \infty$. We first prove that $\overline{z} = \sigma_{n,k+1}$ and later that $z_{n,k+1}^{\delta_i} \to \overline{z}$ as $i \to \infty$. For all $g \in Y_{[n]}^*$,

$$\left\langle g, F'_{[n]}\left(x_n^{\delta_i}\right) s_{n,k+1}^{\delta_i}\right\rangle = \left\langle g, \left(F'_{[n]}\left(x_n^{\delta_i}\right) - F'_{[n]}(\xi)\right) s_{n,k+1}^{\delta_i}\right\rangle + \left\langle g, F'_{[n]}(\xi)\, s_{n,k+1}^{\delta_i}\right\rangle.$$

But as $s_{n,k+1}^{\delta_i} = z_{n,k+1}^{\delta_i} - x_n^{\delta_i} \rightharpoonup \overline{z} - \xi =: \overline{s}$ as $i \to \infty$ and $F'_{[n]}(\xi)^* g \in X^*$,

$$\left\langle g, F'_{[n]}(\xi)\, s_{n,k+1}^{\delta_i}\right\rangle = \left\langle F'_{[n]}(\xi)^* g, s_{n,k+1}^{\delta_i}\right\rangle \to \left\langle F'_{[n]}(\xi)^* g, \overline{s}\right\rangle = \left\langle g, F'_{[n]}(\xi)\,\overline{s}\right\rangle.$$

Now, as $F'_{[n]}$ is continuous and $x_n^{\delta_i} \to \xi$,

$$\begin{aligned}&\left|\left\langle g, \left(F'_{[n]}\left(x_n^{\delta_i}\right) - F'_{[n]}(\xi)\right) s_{n,k+1}^{\delta_i}\right\rangle\right| \\ &\qquad\qquad \le \|g\|_{Y_{[n]}^*}\left\|F'_{[n]}\left(x_n^{\delta_i}\right) - F'_{[n]}(\xi)\right\|_{\mathcal{L}(X, Y_{[n]})}\left\|s_{n,k+1}^{\delta_i}\right\|_X \to 0\end{aligned}$$

as $i \to \infty$ because $\left\|s_{n,k+1}^{\delta_i}\right\| \le \left\|z_{n,k+1}^{\delta_i}\right\| + \left\|x_n^{\delta_i}\right\|$ is uniformly bounded (see (4.11) and (4.16)). Then,

$$\left\langle g, F'_{[n]}\left(x_n^{\delta_i}\right) s_{n,k+1}^{\delta_i}\right\rangle \to \left\langle g, F'_{[n]}(\xi)\,\overline{s}\right\rangle \tag{A.3}$$

and as $g \in Y_{[n]}^*$ is arbitrary,

$$F'_{[n]}\left(x_n^{\delta_i}\right) s_{n,k+1}^{\delta_i} \rightharpoonup F'_{[n]}(\xi)\,\overline{s}.$$

From (3.7) we conclude that

$$b_n^{\delta_i} - F'_{[n]}\left(x_n^{\delta_i}\right) s_{n,k+1}^{\delta_i} \rightharpoonup \widetilde{b}_n - F'_{[n]}(\xi)\,\overline{s},$$

and then

$$\left\|\widetilde{b}_n - F'_{[n]}(\xi)\,\overline{s}\right\| \le \liminf \left\|b_n^{\delta_i} - F'_{[n]}\left(x_n^{\delta_i}\right) s_{n,k+1}^{\delta_i}\right\|. \tag{A.4}$$

Now, as J_p is continuous, we have similarly to $(A.3)$,

$$\begin{aligned}\left\langle J_p\left(z_{n,k}^{\delta_i}\right), z_{n,k+1}^{\delta_i}\right\rangle &= \left\langle J_p\left(z_{n,k}^{\delta_i}\right) - J_p(\sigma_{n,k}), z_{n,k+1}^{\delta_i}\right\rangle + \left\langle J_p(\sigma_{n,k}), z_{n,k+1}^{\delta_i}\right\rangle \\ &\to \left\langle J_p(\sigma_{n,k}), \overline{z}\right\rangle\end{aligned}$$

which in turn implies

$$\begin{aligned}\Delta_p\left(\overline{z},\sigma_{n,k}\right) &= \frac{1}{p}\left\|\overline{z}\right\|^p + \frac{1}{p^*}\left\|\sigma_{n,k}\right\|^p - \left\langle J_p\left(\sigma_{n,k}\right),\overline{z}\right\rangle \\ &\leq \liminf\left(\frac{1}{p}\left\|z_{n,k+1}^{\delta_i}\right\|^p + \frac{1}{p^*}\left\|z_{n,k}^{\delta_i}\right\|^p - \left\langle J_p\left(z_{n,k}^{\delta_i}\right), z_{n,k+1}^{\delta_i}\right\rangle\right) \\ &= \liminf \Delta_p\left(z_{n,k+1}^{\delta_i}, z_{n,k}^{\delta_i}\right).\end{aligned} \tag{A.5}$$

From $x_n^{\delta_i} \to \xi$, $(A.4)$, $(A.5)$ and due to the minimality property of $z_{n,k+1}^{\delta_i}$,

$$\begin{aligned}W_{n,k}\left(\overline{z}\right) &\leq \liminf T_{n,k}^{\delta_i}\left(z_{n,k+1}^{\delta_i}\right) \leq \liminf T_{n,k}^{\delta_i}\left(\sigma_{n,k+1}\right) \\ &= \lim_{i\to\infty} T_{n,k}^{\delta_i}\left(\sigma_{n,k+1}\right) = W_{n,k}\left(\sigma_{n,k+1}\right).\end{aligned}$$

Using minimality and uniqueness of $\sigma_{n,k+1}$, we conclude that $\sigma_{n,k+1} = \overline{z}$ and then $z_{n,k+1}^{\delta_i} \rightharpoonup \sigma_{n,k+1}$. Accordingly, $s_{n,k+1}^{\delta_i} \rightharpoonup \sigma_{n,k+1} - \xi$ which implies that $\overline{s} = \sigma_{n,k+1} - \xi$. We prove now that

$$\Delta_p\left(z_{n,k+1}^{\delta_i}, z_{n,k}^{\delta_i}\right) \to \Delta_p\left(\sigma_{n,k+1},\sigma_{n,k}\right) \text{ as } i\to\infty. \tag{A.6}$$

Define

$$\begin{aligned}&a_i := \Delta_p\left(z_{n,k+1}^{\delta_i}, z_{n,k}^{\delta_i}\right), \quad a := \limsup a_i, \quad c := \Delta_p\left(\sigma_{n,k+1},\sigma_{n,k}\right), \\ &re_i := \frac{1}{r}\left\|b_n^{\delta_i} - F_{[n]}'\left(x_n^{\delta_i}\right)s_{n,k+1}^{\delta_i}\right\|^r, \text{ and } re := \liminf re_i.\end{aligned}$$

In view of $(A.5)$, it is enough to prove that $a \leq c$. Suppose that $a > c$. From the definition of lim sup there exists, for all $M \in \mathbb{N}$, some index $i > M$ such that

$$a_i > a - \frac{a-c}{4}. \tag{A.7}$$

From the definition of lim inf, there exists $N_1 \in \mathbb{N}$ such that

$$re_i \geq re - \frac{\alpha_n\left(a-c\right)}{4}, \tag{A.8}$$

for all $i \geq N_1$. As above, $\lim_{i\to\infty} T_{n,k}^{\delta_i}\left(\sigma_{n,k+1}\right) = W_{n,k}\left(\sigma_{n,k+1}\right)$ and then there is an $N_2 \in \mathbb{N}$ such that

$$T_{n,k}^{\delta_i}\left(\sigma_{n,k+1}\right) < W_{n,k}\left(\sigma_{n,k+1}\right) + \frac{\alpha_n\left(a-c\right)}{2} \tag{A.9}$$

for all $i \geq N_2$. Using $(A.4)$ and setting $M = N_1 \vee N_2$, there exists some index $i > M$ such that

$$\begin{aligned}W_{n,k}\left(\sigma_{n,k+1}\right) &\leq re + \alpha_n c = re + \alpha_n a - \alpha_n\left(a-c\right) \\ &\leq re_i + \frac{\alpha_n\left(a-c\right)}{4} + \alpha_n a_i + \frac{\alpha_n\left(a-c\right)}{4} - \alpha_n\left(a-c\right) \\ &= re_i + \alpha_n a_i - \frac{\alpha_n\left(a-c\right)}{2} = T_{n,k}^{\delta_i}\left(z_{n,k+1}^{\delta_i}\right) - \frac{\alpha_n\left(a-c\right)}{2} \\ &\leq T_{n,k}^{\delta_i}\left(\sigma_{n,k+1}\right) - \frac{\alpha_n\left(a-c\right)}{2}\end{aligned}$$

where the second inequality comes from $(A.8)$ and $(A.7)$ and the last one, from the minimality of $z_{n,k+1}^{\delta_i}$. From $(A.9)$ we obtain the contradiction $W_{n,k}\left(\sigma_{n,k+1}\right) <$

$W_{n,k}(\sigma_{n,k+1})$. Thus, $a \leq c$ and $(A.6)$ holds. From the definition of the Bregman distance Δ_p we have $\left\| z_{n,k+1}^{\delta_i} \right\| \to \|\sigma_{n,k+1}\|$. As $z_{n,k+1}^{\delta_i} \rightharpoonup \sigma_{n,k+1}$ we conclude that $z_{n,k+1}^{\delta_i} \to \sigma_{n,k+1}$ as $i \to \infty$, because X is uniformly convex. So far, we have shown that each positive zero-sequence $(\delta_i)_{i\in\mathbb{N}}$ contains a subsequence $\left(\delta_{i_j}\right)_{j\in\mathbb{N}}$ such that $z_{n,k+1}^{\delta_{i_j}} \to \sigma_{n,k+1}$ as $j \to \infty$ which is enough to prove the statement.

- Bregman variation of Tikhonov-Phillips method (3.58)

  This method uses $\widehat{z}_{n,k}^{\delta_i} = x_n^{\delta_i}$ and $v_{n,k}^{\delta_i} = A_n^{\delta_i} s_{n,k+1}^{\delta_i} - b_n^{\delta_i}$.

  We omit this proof because it is very similar to that one given above to the IT method.

- Mixed gradient-Tikhonov methods presented in Subsection 3.1.4

  For these methods, $\widehat{z}_{n,k}^{\delta_i} = z_{n,k}^{\delta_i}$, $v_{n,k}^{\delta_i} = A_n^{\delta_i} s_{n,k}^{\delta_i} - b_n^{\delta_i}$, $\gamma_n^{\delta_i} = \alpha_n^{\delta_i}$ and $K_2^{\delta_i} = K_0/K_3^{\delta_i}$, where K_0 is defined in Assumption 3, page 58 and $K_3^{\delta_i}$ is an upper bound to the sequence $\left(\frac{1}{p}\left\|e_n^{\delta_i} - \overline{x}_n^{\delta_i}\right\|^p\right)_{n\in\mathbb{N}}$ (see Assumption 4, page 62, and Subsection 3.1.4). In addition to the hypotheses of Theorem 47, we need to assume in this case that Y_j, $j = 0, ..., d-1$ are uniformly smooth Banach spaces, which guarantees that the duality mapping j_r is unique and continuous.

  The proof is very similar to that given above for the dual gradient methods, with the difference that here $K_2 \neq 0$ (see Assumption 4). Since $x_n^{\delta_i} \to \xi$ and $\overline{x}_n^{\delta_i} \to \overline{x}_n^{\xi}$ as $i \to \infty$, the bound $K_3^{\delta_i}$ converges to a bound K_3^{ξ} of the sequence $\left(\frac{1}{p}\left\|e_n - \overline{x}_n^{\xi}\right\|^p\right)_{n\in\mathbb{N}}$ and accordingly, $K_2^{\delta_i}\left\|b_n^{\delta_i}\right\|^r \to K_2^{\xi}\left\|\widetilde{b}_n\right\|^r$ as $i \to \infty$. It follows that $0 \leq \lim_{i\to\infty} \gamma_n^{\delta_i} \leq K_2^{\xi}\left\|\widetilde{b}_n\right\|^r$, which implies, similarly to $(A.1)$ and $(A.2)$ that

  $$J_p\left(z_{n,k+1}^{\delta_i}\right) = J_p\left(z_{n,k}^{\delta_i}\right) - \lambda_{n,k}^{\delta_i}\left[F'_{[n]}\left(x_n^{\delta_i}\right)^* j_r\left(A_n^{\delta_i} s_{n,k}^{\delta_i} - b_n^{\delta_i}\right) + \gamma_n^{\delta_i} J_p\left(s_{n,k}^{\delta_i} - \overline{x}_n^{\delta_i}\right)\right]$$

  converges to

  $$J_p(\sigma_{n,k+1}) = J_p(\sigma_{n,k}) - \lambda_{n,k}^{\xi}\left[F'_{[n]}(\xi)^* j_r\left(v_{n,k}^{\xi}\right) + \gamma_n^{\xi} J_p\left(\sigma_{n,k} - \xi - \overline{x}_n^{\xi}\right)\right],$$

  as $i \to \infty$. Since $J_p^{-1} = J_{p^*}^*$ is a continuous function, the vector $z_{n,k+1}^{\delta_i}$ converges to $\sigma_{n,k+1}$ as $i \to \infty$ and the proof is complete.

Bibliography

[1] Giovanni Alessandrini. Stable determination of conductivity by boundary measurements. *Appl. Anal.*, 27(1-3):153–172, 1988.

[2] Kari Astala and Lassi Päivärinta. Calderón's inverse conductivity problem in the plane. *Ann. of Math. (2)*, 163(1):265–299, 2006.

[3] A. B. Bakushinskiĭ. On a convergence problem of the iterative-regularized Gauss-Newton method. *Zh. Vychisl. Mat. i Mat. Fiz.*, 32(9):1503–1509, 1992.

[4] Johann Baumeister, Barbara Kaltenbacher, and Antonio Leitão. On Levenberg-Marquardt-Kaczmarz iterative methods for solving systems of nonlinear ill-posed equations. *Inverse Probl. Imaging*, 4(3):335–350, 2010.

[5] Liliana Borcea. Electrical impedance tomography. *Inverse Problems*, 18(6):R99–R136, 2002.

[6] Kristian Bredies and Dirk Lorenz. *Matematische Bildverarbeitung. Einführung in Grundlagen und moderne Theorie.* Vieweg+Teubner, Berlin, 2011.

[7] Martin Burger and Barbara Kaltenbacher. Regularizing Newton-Kaczmarz methods for nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 44(1):153–182 (electronic), 2006.

[8] Alberto-P. Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980)*, pages 65–73. Soc. Brasil. Mat., Rio de Janeiro, 1980.

[9] Charles Chidume. *Geometric properties of Banach spaces and nonlinear iterations*, volume 1965 of *Lecture Notes in Mathematics*. Springer-Verlag London, Ltd., London, 2009.

[10] Ioana Cioranescu. *Geometry of Banach spaces, duality mappings and nonlinear problems*, volume 62 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1990.

[11] Christian Clason and Bangti Jin. A semismooth Newton method for nonlinear parameter identification problems with impulsive noise. *SIAM J. Imaging Sci.*, 5(2):505–538, 2012.

[12] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[13] Ron S. Dembo, Stanley C. Eisenstat, and Trond Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19(2):400–408, 1982.

[14] Françoise Demengel and Gilbert Demengel. *Functional spaces for the theory of elliptic partial differential equations.* Universitext. Springer, London; EDP Sciences, Les Ulis, 2012. Translated from the 2007 French original by Reinie Erné.

[15] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1996.

[16] Heinz W. Engl, Karl Kunisch, and Andreas Neubauer. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5(4):523–540, 1989.

[17] V. M. Fridman. On the convergence of methods of steepest descent type. *Uspehi Mat. Nauk*, 17(3 (105)):201–204, 1962.

[18] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics.* Pitman (Advanced Publishing Program), Boston, MA, 1985.

[19] Jacques Hadamard. *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques.* Hermann, Paris, 1932.

[20] Markus Haltmeier, Antonio Leitão, and Otmar Scherzer. Kaczmarz methods for regularizing nonlinear ill-posed equations. I. Convergence analysis. *Inverse Probl. Imaging*, 1(2):289–298, 2007.

[21] Martin Hanke. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, 13(1):79–95, 1997.

[22] Martin Hanke. Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. *Numer. Funct. Anal. Optim.*, 18(9-10):971–993, 1997.

[23] Bangti Jin, Taufiquar Khan, and Peter Maass. A reconstruction algorithm for electrical impedance tomography based on sparsity regularization. *Internat. J. Numer. Methods Engrg.*, 89(3):337–353, 2012.

[24] Qinian Jin. Inexact Newton-Landweber iteration for solving nonlinear inverse problems in Banach spaces. *Inverse Problems*, 28(6):065002, 15, 2012.

[25] Qinian Jin. On the order optimality of the regularization via inexact Newton iterations. *Numer. Math.*, 121(2):237–260, 2012.

[26] Qinian Jin and Linda Stals. Nonstationary iterated Tikhonov regularization for ill-posed problems in Banach spaces. *Inverse Problems*, 28(10):104011, 15, 2012.

[27] S. Kaczmarz. Approximate solution of systems of linear equations. *Internat. J. Control*, 57(6):1269–1271, 1993. Translated from the German.

[28] Jari P. Kaipio, Ville Kolehmainen, Erkki Somersalo, and Marko Vauhkonen. Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems*, 16(5):1487–1522, 2000.

[29] Barbara Kaltenbacher, Andreas Neubauer, and Otmar Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics.* Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

[30] Andreas Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences.* Springer, New York, second edition, 2011.

[31] R. Kowar and O. Scherzer. Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems. In *Ill-posed and inverse problems*, pages 253–270. VSP, Zeist, 2002.

[32] Erwin Kreyszig. *Introductory functional analysis with applications.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1989.

[33] L. Landweber. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.

[34] Armin Lechleiter and Andreas Rieder. Newton regularizations for impedance tomography: a numerical study. *Inverse Problems*, 22(6):1967–1987, 2006.

[35] Armin Lechleiter and Andreas Rieder. Newton regularizations for impedance tomography: convergence by local injectivity. *Inverse Problems*, 24(6):065009, 18, 2008.

[36] Armin Lechleiter and Andreas Rieder. Towards a general convergence theory for inexact Newton regularizations. *Numer. Math.*, 114(3):521–548, 2010.

[37] A. Leitão and M. Marques Alves. On Landweber-Kaczmarz methods for regularizing systems of ill-posed equations in Banach spaces. *Inverse Problems*, 28(10):104008, 15, 2012.

[38] Joram Lindenstrauss and Lior Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas].* Springer-Verlag, Berlin-New York, 1979. Function spaces.

[39] Fábio Margotti and Andreas Rieder. An inexact Newton regularization in Banach spaces based on the nonstationary iterated Tikhonov method. *Journal of inverse and Ill-posed Problems*, Ahead of print, 2014.

[40] Fábio Margotti, Andreas Rieder, and Antonio Leitão. A Kaczmarz version of the `REGINN`-Landweber iteration for ill-posed problems in Banach spaces. *SIAM J. Numer. Anal.*, 52(3):1439–1465, 2014.

[41] A. Neubauer and O. Scherzer. A convergence rate result for a steepest descent method and a minimal error method for the solution of nonlinear ill-posed problems. *Z. Anal. Anwendungen*, 14(2):369–377, 1995.

[42] Nick Polydorides and William R. B. Lionheart. A Matlab toolkit for three-dimensional electrical impedance tomography: a contribution to the Electrical Impedance and Diffuse Optical Reconstruction Software project. *Measurement Science and Technology*, 13(12):1871–1873, 2002.

[43] Andreas Rieder. On the regularization of nonlinear ill-posed problems via inexact Newton iterations. *Inverse Problems*, 15(1):309–327, 1999.

[44] Andreas Rieder. On convergence rates of inexact Newton regularizations. *Numer. Math.*, 88(2):347–365, 2001.

[45] Andreas Rieder. *Keine Probleme mit inversen Problemen.* Friedr. Vieweg & Sohn, Braunschweig, 2003. Eine Einführung in ihre stabile Lösung. [An introduction to their stable solution].

[46] Andreas Rieder. Inexact Newton regularization using conjugate gradients as inner iteration. *SIAM J. Numer. Anal.*, 43(2):604–622 (electronic), 2005.

[47] F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22(1):311–329, 2006.

[48] Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S. Kazimierski. *Regularization methods in Banach spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2012.

[49] R. E. Showalter. *Monotone operators in Banach space and nonlinear partial differential equations*, volume 49 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1997.

[50] Erkki Somersalo, Margaret Cheney, and David Isaacson. Existence and uniqueness for electrode models for electric current computed tomography. *SIAM J. Appl. Math.*, 52(4):1023–1040, 1992.

[51] A. N. Tikhonov. On the solution of incorrectly put problems and the regularisation method. In *Outlines Joint Sympos. Partial Differential Equations (Novosibirsk, 1963)*, pages 261–265. Acad. Sci. USSR Siberian Branch, Moscow, 1963.

[52] Robert Winkler and Andreas Rieder. Model-aware newton-type inversion scheme for electrical impedance tomography. *Inverse Problems*, 31(4):045009, 2015.

[53] Zong Ben Xu and G. F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *J. Math. Anal. Appl.*, 157(1):189–210, 1991.