Westfälische
Wilhelms-Universität
Münster

› Reconstruction Using Local Sparsity

A Novel Regularization Technique and an Asymptotic Analysis of Spatial Sparsity Priors

Pia Heins

– 2014 –

Angewandte
Mathematik
Münster

Westfälische
Wilhelms-Universität
Münster

Fach: Mathematik

Reconstruction Using Local Sparsity

A Novel Regularization Technique and an Asymptotic Analysis of Spatial Sparsity Priors

Inaugural Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

– Dr. rer. nat. –

im Fachbereich Mathematik und Informatik

der Mathematisch-Naturwissenschaftlichen Fakultät

der Westfälischen Wilhelms-Universität Münster

eingereicht von

Pia Heins

aus Oberhausen

– 2014 –

Dekan:	Prof. Dr. Martin Stein
Erster Gutachter:	Prof. Dr. Martin Burger (Universität Münster)
Zweiter Gutachter:	Prof. Dr. Samuli Siltanen (University of Helsinki)
Tag der mündlichen Prüfung:	21.01.2015
Tag der Promotion:	21.01.2015

Abstract

The specific field of inverse problems, where the unknown obtains apart from its spatial dimensions at least one additional dimension, is of major interest for many applications in imaging, natural sciences and medicine. Enforcing certain sparsity priors on such unknowns, which can be written as a matrix, has thus become current state of research. This thesis deals with a special type of sparsity prior, which enforces a certain structure on the unknown matrix. Furthermore, we consider the asymptotics of spatial sparsity priors. Therefore, this thesis can be roughly divided into two parts accordingly.

In the first part we present and analyze a novel regularization technique promoting so-called *local sparsity* by minimizing the $\ell^{1,\infty}$-norm as a regularization functional in a variational approach. This regularization can be used for dictionary based matrix completion problems, where the unknown dynamic image can be written as a product of a certain known dictionary and a coefficient matrix, which we would like to reconstruct. Typical areas of application are for instance in dynamic positron emission tomography or spectral imaging. In this context we have the a-priori knowledge that every pixel of the dynamic image should consist of a linear combination of the basis vectors with as few nonzero components as possible. This can be realized by using the above-mentioned regularization functional. Due to the rather involved mathematical structure of this regularization functional, we provide an equivalent formulation, which facilitates its analysis and simplifies its computation. We not only analyze these different formulations, but also present algorithms for the solution of this variational model on the basis of the alternating direction method of multipliers (cf. Rockafellar 1976). Our numerical results, gained by applying these algorithms to synthetic data, validate our method and illustrate its potential.

The second part of this thesis is concerned with the theoretical analysis of the asymptotics of certain sparsity priors. We consider discrete sparsity promoting functionals and analyze their behavior as the discretization becomes finer. In so doing, we are able to compute some Γ-limits. We not only consider usual ℓ^p-norms for $p \geq 1$, but also analyze the asymptotics of the ℓ^0-“norm”. On the basis of these insights, we moreover deduce some Γ-limits for certain types of mixed norms. In order to verify our results numerically, we furthermore consider the deconvolution of a sparse spike pattern for different discretizations as the step size of the grid becomes smaller.

Keywords: Local Sparsity, $\ell^{1,\infty}$-Regularization, Sparsity, ℓ^1-Regularization, Compressed Sensing, Mixed Norms, Augmented Lagrangian, Inverse Problems, Regularization Theory, Imaging, Image Processing, Image Reconstruction, Dynamic Positron Emission Tomography, Asymptotic Sparsity, Super-Resolution, Radon Measures

Acknowledgements

I would like to thank

Martin Burger first of all for offering me an interesting job as a student assistant and consequently asking me to continue my research as a PhD student and for supervising this thesis. Moreover, for giving me the opportunity to sometimes but regularly work from home (which is about 200km from Münster), for taking the time to patiently answer my questions, for many interesting discussions and furthermore for answering e-mails quite quickly, even if the answer contains only a word.

Samuli Siltanen for co-reviewing my thesis. For making many conferences greatly enjoyable and for introducing me to the world of Finnish walrus-wrestling, a competition even took place at our skiseminar in Kleinwalsertal, Austria.

Carola-Bibiane Schönlieb for kindly inviting me to Cambridge, for being an admirable role model and for bringing the mentoring program to Münster.

Xiaoqun Zhang for interesting discussions and for inviting me to Shanghai.

Caterina Ida Zeppieri for her kind help on Γ-convergence.

my colleagues and ex-colleagues from the imaging work group and the whole Institute for Computational and Applied Mathematics, in particular

- *Hendrik Dirks* for being a really good friend for many years, for sharing the office with me, for many conversations (interesting and silly), for sharing my dedication to vegetable gardening and for sometimes taking care of my apartment and my plants when I am not at home.
- *Felix Lucka* for accompanying me to many conferences, for being quite funny

while trying to eat with chopsticks, for proofreading and especially for introducing me to my boyfriend Florian.

- *Rachel Hegemann* for being a good friend, for helping me out with some language problems and for joint painting (almost without painting at all).
- *Jan Hegemann* for his advice on nerdy TV shows and for sometimes startling me with his sneezing.
- *Michael Möller* for many fruitful discussions, for proofreading and, of course, for his great Lego robots, which make mathematics much more applicable and enjoyable.
- *Martin Benning* for his bad jokes, for being the cause of the expression "That was a classical Benning!" and for being a good mentor.
- *Jahn Müller* for organizing those awesome skiseminars, for helping me with any Apple-related question and for making a funny face when something does not work at all.
- *Frank Wübbeling* for always having a good advice, for forthrightly speaking his mind, for usually being in a good mood and, of course, for his laugh.
- *Alex Sawatzky* for his high spirits and his costume during snowboarding at the skiseminar.
- *Christoph Brune* for useful advices, for many answers and for good discussions.
- *Ralf Engbers* for being a quite talented snowboard student and for tumbling on purpose after seeing me jumping to the snow.
- *Lena Frerking* for proofreading, for being a MINT-mentor as well and for co-organizing our large mentoring get-togethers.
- *Eva-Maria Brinkmann* and *Joana Grah* for proofreading and for reminding me of my beginning as a PhD-student.
- *Ole Løseth Elvetun* for improving my German, since I had to get used to speak clearly and grammatically correct, and for sharing the office with me.
- *René Milk* for almost always having a solution for any kind of technical problem and for patiently explaining it in an understandable way.
- *Claudia Giesbert* and *Carolin Gietz* for magically always having a solution for any kind of organizational problem.

Hanne Kekkonen for meeting with me on so many conferences, for making a good time out of them and for introducing me to Finnish culture.

Sarah Jane Hamilton for making conferences much more enjoyable and for demonstrating that a mathematical talk does not have to send the audience to sleep.

my parents and their partners *Ulrike Grabe*, *Dirk Heins*, *Hartmut Grabe*, *Sabine Herzog* for always believing in me, for building me up in times of doubt and for supporting me unconditionally.

Jannis Heins for being the best brother I could ever imagine, for sometimes making me miss my childhood, for turning me into a proud big sister and for being better in catching the quoit than I am.

Meinen Großeltern *Ursula* und *Friedel Lechtleitner* für ihre Liebe und Unterstützung, für alles was ich von ihnen lernen durfte und für jegliche gutgemeinten Ratschläge. Ihr seid großartig!

Meiner Großmutter *Sofia Heins* für ihre Liebe und Unterstützung und weil sie der großherzigste und gutmütigste Mensch ist, den ich jemals kennen lernen durfte. Ich bewundere dich zutiefst!

Florian Knigge for believing in me, supporting me and always being there for me, for building me up, when I am sometimes dispirited. For being patient with me, when my memory is playing tricks on me. For always being able to make me laugh and, of course, for being the best and irreplaceable travel mate ever!

All my friends for supporting me and being there for me when I need them, especially *Michelle Bouanane*, *Nora Gärtner*, *Carolin Wirtz*, *Josephine Soei*, *Claudia Denecke*, *Antje Friedrich*, *Jennifer Schulz* and many many more.

and of course everybody, who deserves to be mentioned here, but has been forgotten. Sorry for that!

Contents

List of Figures

List of Algorithms

List of Tables

1
Introduction

In the course of this thesis, we will propose a new model for dictionary-based matrix completion problems, which promotes local sparsity. Furthermore, we will discuss the asymptotics of spatial sparsity priors.
Before going into the details, we firstly want to motivate our principal topic. Afterwards, we give an introduction to inverse problems and their method of resolution, namely variational models. Finally, we state our contributions and summarize the organization of this work.

1.1. Motivation

First of all, let us motivate our work by introducing applications we have in mind and explaining the general problem, which emerges from these applications.
For a moment let us put ourselves into the position of one of the poor human beings trapped in the cave of Plato's *Allegory of the Cave*, (cf. for instance Plato 2010), which the Greek philosopher presented in his work "The Republic (514a-520a)" around 380 BC. We have lived our whole lives chained to the wall of a cave facing a blank wall. We can only watch shadows projected onto the wall by things passing in between a fire and ourselves behind us, cf. Figure 1.1.
Now let us assume there is an elephant standing between the fire and our backs. We have never seen an elephant before and have no idea what this object could be, which is projected onto the wall in front of us. We are not able to identify the object by only having the information of its shadow. However, there is one fellow prisoner, who has been outside the cave before and has seen an elephant in real life. Due to his prior knowledge about the shape of an elephant, he is able to directly identify the object as such.

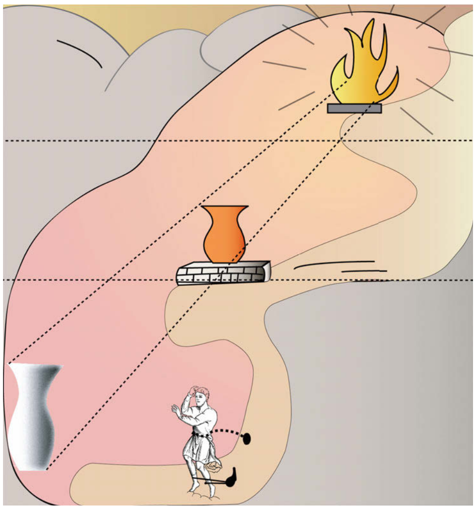


Figure 1.1.: Detail of the allegory of the cave by Gothika (2008)

Luckily, we are not captive in a cave facing the shadows on the wall. However, many real life problems are not solvable without prior knowledge of the subject. Humans usually intuitively include their prior knowledge into their decisions and problem solving. In order to teach a computer reasonable problem solving, we thus have to offer a way of including a-priori knowledge to the model.

As a real life application, let us consider medical imaging and as an example thereof computerized tomography (CT). In order to obtain an image of the inside of a human body, the patient is x-ray scanned from different angles. Then a 3-dimensional image of the inside of the body is reconstructed by using these 2-dimensional x-ray images. The more images from different angles we have, the better is the quality of the reconstructed image. Unfortunately, in this case the patient would be exposed to a lot of x-ray radiation, which is a danger to his health. Thus it is desirable to use fewer angles to preserve the health of the patient. These types of problems are called *inverse problems* and are usually hard to solve.

In order to still be able to solve certain problems, we include a-priori knowledge about the application. For instance we could incorporate that the edges in the images should be sharp and the area in between should be homogeneous. Another type of including a-priori knowledge is to provide a so called *dictionary* for the reconstruction process. The reconstructed image is supposed to be a linear combination of these dictionary elements using as few elements as possible. This signal or image processing technique is called *compressed sensing* and promotes *sparse* solutions, i.e. solutions with few nonzero entries.

Let us now mathematically introduce inverse problems and variational models, i.e. methods for the solution of these problems.

1.2. Inverse Problems

Applications in imaging, natural sciences and medicine frequently tend to result in so called *inverse problems*, that is determining an unknown cause $\hat{z} \in \mathcal{U}(\Omega)$ only by measuring its effect $\hat{w} \in \mathcal{V}(\Sigma)$. In the mathematical modelling behind these problems $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$ denote Banach spaces of functions on bounded and compact sets Ω, respectively Σ. In the attempt of solving such an inverse problem, the cause $\hat{z}$ and its effect $\hat{w}$, i.e. the ideally measured data, have to be connected by a mathematical model $\mathcal{A}$. Thus solving an inverse problem yields solving the operator equation

$$\mathcal{A}\hat{z} = \hat{w} \, , \tag{1.1}$$

where $\mathcal{A} : \mathcal{U}(\Omega) \longrightarrow \mathcal{V}(\Sigma)$ denotes a linear and compact operator.
Inverse problems usually do not fulfill HADAMARD's postulates of *well-posedness* and thus are called *ill-posed*. This characterization of a problem, which we present subsequently, traces back to the French mathematician JACQUES HADAMARD (1902).

Definition 1.1.
A problem is called *well-posed*, if

- there exists a solution of the problem,
- the solution is unique,
- the solution depends continuously on the input data.

If a problem is not well-posed, it is called *ill-posed*.

Since the operator $\mathcal{A}$ is compact, it usually cannot be inverted continuously (cf. ENGL et al. (1996)). Thus in the majority of the cases the third point is violated, i.e. the solution does not depend continuously on the input data. In addition, we usually encounter the difficulty that the exact data $\hat{w}$ are not available in real life. This condition complicates the problem even further. Thus considering (1.1) might not even be possible, since we face measurement errors. In practice we more often encounter the inverse problem

$$\mathcal{A}z = w \, , \tag{1.2}$$

with $z \in \mathcal{U}(\Omega)$ and $w \in \mathcal{V}(\Sigma)$. Here w are the actually measured data, which differ to some extent from $\hat{w} = \mathcal{A}\hat{z}$, which could be derived if the exact cause $\hat{z}$ was known. This difference between the ideal data $\hat{w}$ and the real data w is termed *noise*. Thus (1.2) is an approximation of the exact inverse problem (1.1). In the discrete and linear case the inverse problem (1.2) reduces to a linear system

$$Az = w \ ,$$

where $z \in \mathbb{R}^M$ and $w \in \mathbb{R}^L$ are vectors and $A \in \mathbb{R}^{L \times M}$ is a matrix, often referred to as the *observation matrix*. Including an additional dimension such as time or spectral information, we can even consider the matrix equation

$$AZ = W \ , \tag{1.3}$$

with $A \in \mathbb{R}^{L \times M}$, $Z \in \mathbb{R}^{M \times T}$ and $W \in \mathbb{R}^{L \times T}$, where Z again is unknown. This is the case in many applications, for instance dynamic positron emission tomography or spectral imaging.

Since in (1.2) the operator $\mathcal{A}$ is often not invertible, one seeks for an approximate solution $\tilde{z}$ close to the true solution $\hat{z}$. In order to work with problems like (1.2), the variational approach

$$\tilde{z} = \operatorname*{argmin}_{z \in \operatorname{dom} \mathcal{R}} \ \{\mathcal{D}_w(\mathcal{A}z) + \alpha \mathcal{R}(z)\} \tag{1.4}$$

is quite convenient. Since it is easier to claim the distance of $\mathcal{A}z$ and w being minimal, instead of asking for equality in (1.2), $\mathcal{D}_w$ should be a suitable distance measure. In order to include problem dependent a-priori information and to promote solutions with certain properties, the regularization functional $\mathcal{R}$ should be small, in case the solution fits best to the a-priori knowledge and it should be highly increasing, if the solution deviates from the prior. In (1.4) α is called *regularization parameter* and acts as weighting between the data fidelity term $\mathcal{D}_w$ and the prior $\mathcal{R}$. It should be chosen depending on the noise level.

When it comes to choosing a suitable data term $\mathcal{D}_w$ the quadratic L^2 data fidelity term

$$\mathcal{D}_w(\mathcal{A}z) = \frac{1}{2} \left\| \mathcal{A}z - w \right\|_{L^2(\Omega)}^2 \tag{1.5}$$

is quite common, especially in the presence of additive Gaussian noise. Depending on the type of noise, different data terms can be derived by using a statistical approach often referred to as *Bayesian modeling*. Apart from (1.5) the following exemplary data fidelity terms can be derived:

$$\text{additive Laplacian noise} \quad \rightsquigarrow \quad \mathcal{D}_w(\mathcal{A}z) = \|\mathcal{A}z - w\|_{L^1(\Omega)}$$
$$\text{Poisson noise} \quad \rightsquigarrow \quad \mathcal{D}_w(\mathcal{A}z) = \int_\Omega \left(w \log \frac{w}{\mathcal{A}z} + \mathcal{A}z - w \right) \mathrm{d}x$$

The latter is also known as *Kullback-Leibler divergence*. The interested reader may find further information on the derivation of these data fidelity terms in Heins (2011, Section 2.2), Brune (2010, Section 2.2) or in Benning (2011, Section 2.3).
In (1.4) convex regularization terms are typically quadratic functionals of the form $\mathcal{R}(z) = \|\Gamma z\|^2_{L^2(\Omega)}$ with a linear operator Γ. This regularization term in combination with (1.5) yields the classical Tikhonov regularization, which traces back to the Russian mathematician Andrey Tikhonov (1943). In case the solution is expected to be smooth, using $\|\Gamma z\|^2_{L^2(\Omega)}$ with $\Gamma = \nabla$ seems to be a reasonable choice.
However, this is certainly not always the case. Especially in imaging or image processing the solutions are rather expected to have sharp edges. Thus $\mathcal{R}(z) = \|\nabla z\|^2_{L^2(\Omega)}$ would be a disadvantageous choice.
In contrast to $\mathcal{R}(z) = \|\nabla z\|^2_{L^2(\Omega)}$, the *total variation* semi-norm

$$\mathrm{TV}(z) := \sup_{\substack{p \in C_0^\infty(\Omega) \\ \|p\|_{L^\infty(\Omega)} \leq 1}} \int_\Omega z \, \nabla \cdot p \, \mathrm{d}x \;\approx\; \|\nabla z\|_{L^1(\Omega)} \tag{1.6}$$

promotes the recovery of functions with discontinuities, i.e. sharp edges. Nowadays it is very common to use the so-called *Rudin-Osher-Fatemi (ROF) model* for denoising, i.e.

$$\mathrm{ROF}(z) := \frac{1}{2} \|z - w\|^2_{L^2(\Omega)} + \mathrm{TV}(z) \ ,$$

cf. Rudin, Osher and Fatemi (1992). The idea of total variation minimization makes use of the fact that the L^1-norm itself promotes *sparse* solutions, i.e. solutions with only a few nonzero entries. Taking the L^1-norm of the gradient of a function promotes solutions with derivatives, which have only a few nonzero entries. Hence regularizing with (1.6) favors solutions with sharp edges and constant parts between the edges.

1.3. Contributions

In signal and image processing sparse reconstruction using standard ℓ^1-minimization is a very useful and versatile tool and by now it is quite well understood. The basic idea behind this is the knowledge of the ℓ^1-norm being the convex relaxation of the ℓ^0-“norm” (cf. Donoho and Elad (2003)). Recent research went further in this direction by adapting this concept to problems with unknowns being matrices. The outcome of this were even more advanced sparsity promoting methods.

Instead of considering usual sparsity regularizations componentwise on the unknown matrix such as minimizing the $\ell^{1,1}$-norm, Yuan and Lin (2006) considered for instance a generalization of the lasso method (cf. Tibshirani (1996)), which they call *group lasso*. The original lasso method (cf. also the review Tibshirani (2011)) is a shrinkage and selection method for linear regression, i.e. it basically minimizes the sum of a squared ℓ^2 data term and an ℓ^1-regularization term. The group lasso, however, generalizes this method by using a slightly different regularization term, i.e. it minimizes the ℓ^1-norm of a weighted ℓ^2-norm. Later this $\ell^{2,1}$-regularization was further generalized by Fornasier and Rauhut (2008) and Teschke and Ramlau (2007), which then became known under the term *joint sparsity*. This method mainly consists of minimizing $\ell^{p,1}$-norms, which are used to include even more prior knowledge about the unknown such as additional structures like block sparsity (or collaborative sparsity).

However, for many applications, such as dynamic positron emission tomography or unmixing problems, it turns out to be useful to incorporate another type of sparsity. Enhancing the idea of usual ℓ^1-sparsity to what we call *local sparsity* is one of the main contributions of this thesis. Local sparsity turns out to be beneficial when working on problems including inversion with some spatial dimensions and at least one additional dimension such as time or spectral information.
In order to incorporate the idea of local sparsity, we motivate the use of the $\ell^{1,\infty}$-norm as regularization functional in a variational framework for dictionary based reconstruction of matrix completion problems. Working with the $\ell^{1,\infty}$-norm turns out to be rather difficult, which is why we additionally propose alternative formulations of the problem. By accompanying the $\ell^{1,\infty}$-regularization with total variation minimization, we include even more prior knowledge and thus improve the results even further.
Besides this, we discuss basic properties of the $\ell^{1,\infty}$-functional and potential exact recovery. In addition, we propose different algorithms for the solution of $\ell^{1,\infty}$-regularized problems and show computational results for synthetic examples.

Sparsity promoting regularizations are usually finite-dimensional and therefore highly dependent on their discretization. Due to this reason, the solutions of corresponding variational problems rely on the same discretization and may therefore not even be stable for different grid sizes. Moreover, in many applications an infinite-dimensional modeling seems to be more reasonable and thus it would be disadvantageous to consider its discrete counterpart instead.

The establishment of an appropriate asymptotic theory behind these discrete approaches is thus of particular importance. Having an underlying asymptotic theory permits the analysis of variational problems independent of their discretization and thus yields robustness. In order to analyze sparsity regularizations in an infinite dimensional setting, recent research investigates approaches in the space of finite Radon measures, cf. for instance Bredies and Pikkarainen (2013), Duval and Peyré (2013) and Scherzer and Walch (2009).

This thesis contributes to this topic by analyzing the asymptotics of regularization methods applicable for instance to spatial sparsity. By working in the space of finite Radon measures and utilizing Γ-convergence, we aim to make variational problems independent of their discretization. This analysis can be performed by considering the priors only. We not only investigate ℓ^p-norms for $p \geq 1$ and $p = 0$, but also examine mixed $\ell^{p,q}$-norms as the step size of the grid vanishes.
Furthermore, we consider the deconvolution of a sparse spike pattern and investigate the structure of a solution in case the exact spikes are located at the nodes as well as in between the grid points. We verify our results by numerically computing a deconvolution of a few sparse spikes for different discretizations as the grid becomes finer.

1.4. Organization of this Work

In the last sections we motivated this work, briefly introduced the field of inverse problems and emphasized our contributions to this topic. In order to complete this introductory chapter, we now give an overview on how this work is organized.

Chapter 2 will provide the mathematical background needed for the sequel of this thesis. We will start by recalling some basic calculus of variations and some fundamentals about Γ-convergence, before we will introduce important function and sequence spaces such as the Lebesgue spaces $L^p(\Omega)$, the sequence spaces $\ell^p(\Omega)$ and the Sobolev spaces $W^{k,p}(\Omega)$. Furthermore, we will introduce the space of functions with bounded variation $\mathrm{BV}(\Omega)$, which is fundamental in imaging. The space $\mathrm{BV}(\Omega)$ will be relevant for Chapter 5 and 6, since total variation regularization will be used as an additional regularization on the main problem, in order to promote sharp edges in the reconstructed images. Moreover, we will introduce the space of finite Radon measures, which will be of particular importance for a theoretical investigation about the asymptotics of spatial sparsity regularizations, which can be found in Chapter 7. Afterwards, we will review some basic convex analysis and recall the concept of subdifferential calculus and the Legendre-Fenchel duality. We will end this chapter of mathematical preliminaries by giving an introduction to sparsity regularization, which builds the basis for other sparsity promoting regularization techniques, c.f. for instance Chapters 3 and 4. Starting with the ℓ^0-"norm", which is not a norm in the classical sense, we will afterwards continue with its convex relaxation, which will lead us to the well known sparsity-promoting ℓ^1-regularization. Furthermore, we will shortly discuss regularization with Radon measures, which will become relevant for Chapter 7.

After this fundamental chapter, we will introduce mixed $\ell^{p,q}$-norms in Chapter 3. We will motivate the use of mixed norms to promote sparsity for matrices, which builds the basis for our analysis in Chapter 4. We will see that not only classical sparsity, such as using the $\ell^{1,1}$-regularization, is possible. Mixed norms can moreover be used to promote some kinds of structured sparsity. In particular, we will consider joint sparsity via $\ell^{p,1}$-norms and local sparsity via $\ell^{1,q}$-norms.

The latter one yields the local sparsity promoting $\ell^{1,\infty}$-regularization, which will be the main subject of Chapter 4. We will discuss basic properties of this regularization such as different problem formulations, existence and uniqueness. Moreover, we will give attention to subdifferentials and source conditions for these different formulations. After showing the equivalence of those formulations, we will analyze the asymptotics in case that the regularization parameter tends to infinity. Another section shall be devoted to the analysis of exact recovery of locally 1-sparse solutions, where we will introduce certain conditions for exact recovery. We will end this chapter by including an additional regularization, which shall further improve the numerical results presented in Chapter 6, namely total variation regularization.

In Chapter 5 we will propose algorithms for the reconstruction with local sparsity, which are based upon the alternating direction method of multipliers (ADMM). Computational experiments using these algorithms can be found later in Chapter 6. First we will propose an algorithm for the reconstruction including $\ell^{1,\infty}$- and $\ell^{1,1}$-regularization. Then we will state another two algorithms for the reconstruction including $\ell^{1,\infty}$- and $\ell^{1,1}$-regularization and an additional total variation regularization, both on the images in every time step as well as on the coefficient matrices for every basis function.

On the basis of the previous chapters, we will apply our model proposed in Chapter 4 to dynamic positron emission tomography, which will be used to visualize myocardial perfusion, in Chapter 6. We will firstly motivate, why the knowledge about the perfusion of the heart muscle is of particular importance and give a short introduction to the medical and technical background of dynamic positron emission tomography. We will briefly discuss a model for blood flow and tracer exchange, i.e. kinetic modeling, which will then yield the same inverse problem, which we will aspire to solve in Chapter 4. Afterwards, we will apply the algorithms, which will be deduced in Chapter 5, to artificial data in order to verify our model and illustrate its potential. Moreover, we will discuss the quality of the proposed algorithms and show some examples.

Chapter 7 is devoted to the analysis of the asymptotics of sparsity promoting regularization functionals. We will motivate the establishment of an appropriate asymptotic theory behind commonly used sparsity priors. This theory could improve the solution of inverse problems in applications, where an infinite-dimensional modeling is more reasonable. Moreover, we aim to make regularization functionals independent of their discretization. We will discuss spatial sparsity under Γ-convergence and consider the cases of different ℓ^p-norms for $p \geq 1$ as well as for $p = 0$. Furthermore, we will discuss the asymptotics for certain types of mixed $\ell^{p,q}$-norms, which will thus establish a connection to the previously examined functionals.

As an application for the theory of Chapter 7, we will examine the deconvolution of a sparse spike pattern in Chapter 8. We will theoretically investigate the structure of the discrete deconvolved solutions as the positions of the exact spikes are located firstly on the nodes and secondly in between the grid points. In order to support our theory numerically, we will consider a standard variational problem including the sparsity promoting ℓ^1-regularization, which we will implement via the alternating direction method of multipliers (ADMM). Previously to the reconstruction, we will convolve the exact δ-spikes analytically with a Gaussian kernel and then discretize the convolved data for different grid sizes. On the basis of this, we will test our algorithm for different discretizations as the step size of the grid becomes smaller. We will indeed obtain a certain limit, which will be consistent with our theoretical considerations as well as with the literature.

Finally we will summarize the content of this thesis in Chapter 9, which will be divided into two sections, accordingly to the main topics of this thesis. First we will summarize our work on local sparsity and draw conclusions. We will moreover present questions and ideas, which still remain to be examined. Afterwards, we will discuss conclusions for our analysis of the asymptotics of sparsity priors and give an outlook on possible future work.

In addition to this summary, Figure 1.2 illustrates, how this thesis is organized.

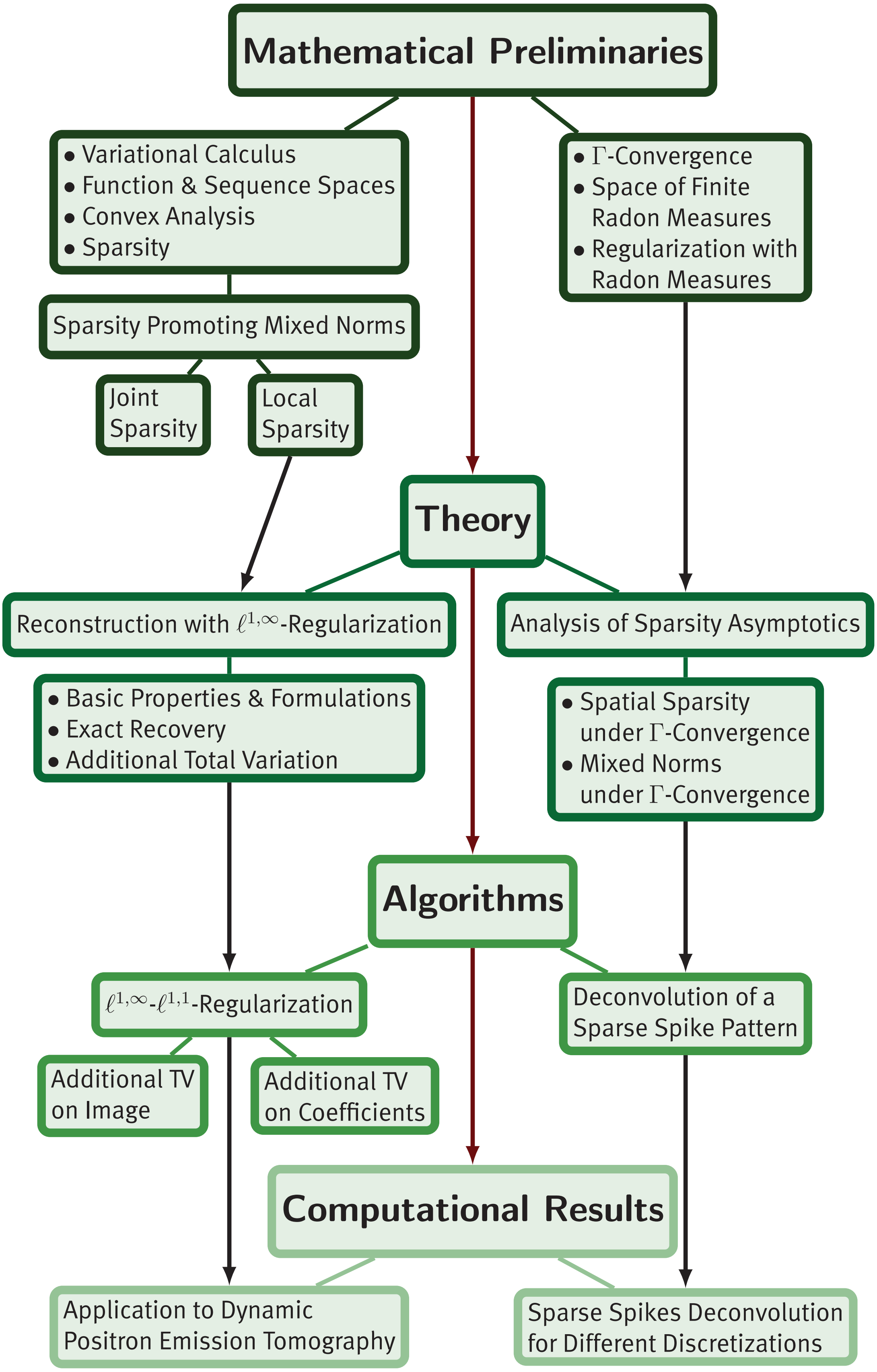


Figure 1.2.: Overview of the organization of this thesis

2

Mathematical Preliminaries

In this chapter we provide the mathematical background, which is needed in the course of this thesis.
Starting by summarizing some basic principles of calculus of variations, we proceed with a section about essential function and sequence spaces. This section not only contains the definitions of Lebesgue spaces and the sequence spaces ℓ^p, but also the definition of Sobolev spaces. The latter ones are then utilized to motivate the usage of the space of functions with bounded variation. We end this section by introducing the space of finite Radon measures, which will be important for the analysis of asymptotic sparsity. The next section gives a short summary of basic convex analysis including subdifferential calculus and the concept of Legendre-Fenchel duality. Finally we end this chapter with a short introduction to sparsity regularization.

2.1. Variational Calculus

In this section we give an introduction to the calculus of variations. We start by recalling basic definitions and state important theorems. Afterwards, we introduce convergence for functionals, namely Γ-convergence, which will be of importance during our asymptotic analysis of sparsity regularizations in Chapter 7.

2.1.1. Introduction

Since the minimization of functionals like (1.4) is related to the calculus of variations, we are going to recall some definitions and results of this subject, which are needed in this thesis. For the sake of compactness we refer to Dacorogna (2004), Ekeland and Témam (1999) or Zeidler (1985) for a more detailed introduction on this subject.

Before getting more into the details, we would like to remind the reader of the difference between operators and functionals.

Definition 2.1 (Operators & Functionals).
Let $\mathcal{U}$ and $\mathcal{V}$ denote two Banach spaces with topology τ_1, τ_2 respectively. A mapping $J\colon(\mathcal{U},\tau_1)\longrightarrow(\mathcal{V},\tau_2)$ between those Banach spaces is called *operator*.
In the special case that $\mathcal{V}$ is a field, J is called *functional*.

We consider functionals $J\colon\mathcal{U}\longrightarrow\overline{\mathbb{R}}$ with a nonempty domain. Therefore, we recall the definition of a proper functional.

Definition 2.2 (Proper Functional).
Let $\mathcal{U}$ be a Banach space. A functional $J\colon\mathcal{U}\longrightarrow\overline{\mathbb{R}}$ is called *proper* if $J(u)\neq-\infty$ for all $u\in\mathcal{U}$ and if there exists at least one $u\in\mathcal{U}$ with $J(u)\neq+\infty$. In this setting

$$\mathrm{dom}J := \{u\in\mathcal{U}|J(u)<\infty\}$$

is called *effective domain* of the functional J.

After recalling these basic definitions we now introduce generalized derivatives in Banach spaces.

Definition 2.3 (Directional Derivative).
Let $J\colon U\subset\mathcal{U}\longrightarrow\mathcal{V}$ denote an operator between Banach spaces and let U be a non empty subset. Then the *directional derivative* at the point $u\in U$ in direction $v\in U$ is defined as

$$\mathrm{d}J(u,v)=\lim_{t\searrow 0}\frac{J(u+tv)-J(u)}{t}\ ,$$

if the limit exists. In case the directional derivative exists for all $v\in\mathcal{U}$, J is called *directionally differentiable.*

Definition 2.4 (Gâteaux Differentiability).
An operator J with preconditions as in Definition 2.3 is called *Gâteaux differentiable*, in case the directional derivative $J'(u)\colon\mathcal{U}\longrightarrow\mathcal{V}$ with $v\mapsto\mathrm{d}J(u,v)$ is linear and bounded.

Definition 2.5 (Fréchet Differentiability).
Let for a Gâteaux differentiable operator J hold

$$\|J(u+v)-J(u)-J'(u)v\|_{\mathcal{V}}=o(\|v\|_{\mathcal{U}})\quad\text{for}\quad\|v\|_{\mathcal{U}}\to 0\ .$$

Then J is called *Fréchet differentiable.*

Remark 2.1.
In the case that J is Gâteaux differentiable in a neighborhood of u and $J'(u)$ is continuous at u, J is Fréchet differentiable at u as well.

In order to state some results about the existence of a minimizer of variational problems like (1.4), we will first recall some definitions, which will then lead us to the fundamental theorem of optimization.

Definition 2.6 (Lower semi-continuity)**.**
Let $J : (\mathcal{U}, \tau) \longrightarrow \mathbb{R} \cup \{\infty\}$ be a functional on a Banach space $\mathcal{U}$ with metric topology τ. J is called *lower semi-continuous* at $u \in \mathcal{U}$ if

$$J(u) \leq \liminf_{k \to \infty} J(u_k) \tag{2.1}$$

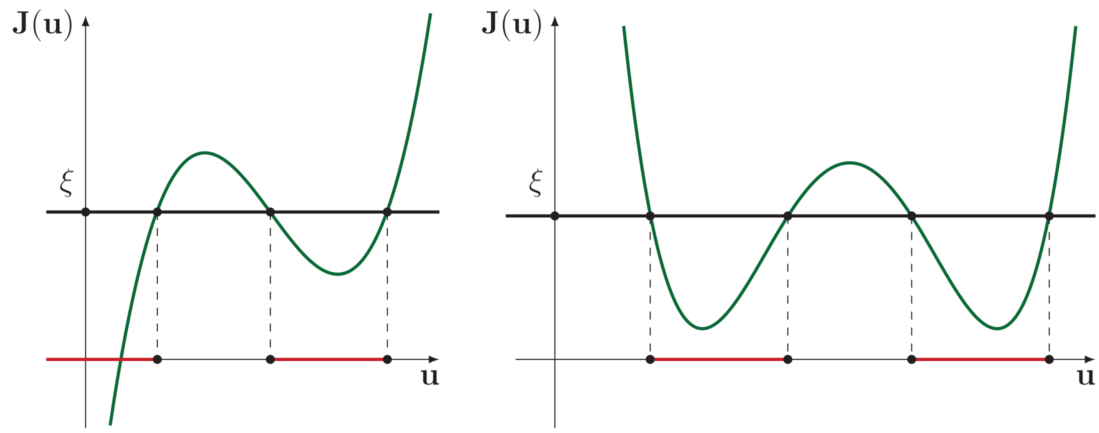
holds for all sequences $(u_k)_{k \in \mathbb{N}}$ with $u_k \to u$ in the topology τ.

Definition 2.7 (Compactness of Sublevel Sets)**.**
Let $J : (\mathcal{U}, \tau) \longrightarrow \mathbb{R} \cup \{\infty\}$ be a functional on a Banach space $\mathcal{U}$ with topology τ.
For a $\xi \in \mathbb{R}$ the corresponding *sublevel set* of J is defined as

$$\mathcal{S}_\xi := \{u \in \mathcal{U} \mid J(u) \leq \xi\} \ . \tag{2.2}$$

Moreover, we call $\mathcal{S}_\xi$ *compact* if it is not empty and compact in the topology τ.

An illustration of different sublevel sets can be found in Figure 2.1.



(a) This function does *not* have compact sublevel sets

(b) This function has compact sublevel sets

Figure 2.1.: Illustration of sublevel sets (red) of two different functions

By using these two definitions, we can now state the fundamental theorem of optimization (cf. Zeidler (1985) or Aubert and Kornprobst (2002)).

Theorem 2.1 (Fundamental Theorem of Optimization).
Let $\mathcal{U}$ be a Banach space with topology τ. Furthermore, let $J\colon(\mathcal{U},\tau)\longrightarrow\mathbb{R}\cup\{\infty\}$ be a lower semi-continuous functional. In addition, let there exist a $\xi\in\mathbb{R}$ such that the corresponding sublevel set $\mathcal{S}_\xi$ of J is compact.
Then there exists a global minimum $\hat{u}\in\mathcal{U}$, i.e.

$$J(\hat{u})=\inf_{u\in\mathcal{U}} J(u)\ .$$

Proof.
Let $(u_k)_{k\in\mathbb{N}}$ be a minimizing sequence, i.e. $J(u_k)\to\inf\limits_{u\in\mathcal{U}} J(u)$. We have $u_k\in\mathcal{S}_\xi$ for $k\geq k_0$ sufficiently large and hence $(u_k)_{k\geq k_0}$ is contained in a compact set. Thus it has a convergent subsequence, which we again denote with (u_k). Furthermore, let its limit be $\hat{u}$. Due to the lower semi-continuity (2.1) of J, it follows that

$$\inf_{u\in\mathcal{U}} J(u)\leq J(\hat{u})\leq\liminf_{k\to\infty} J(u_k)=\inf_{u\in\mathcal{U}} J(u)$$

holds. Therefore, $\hat{u}$ is a global minimum of J. □

Since boundedness causes compactness in finite dimensions, this theorem easily leads to a proof of existence in finite-dimensional optimization. However, this does not hold in function spaces due to their infinite dimension. In order to still be able to conclude compactness from boundedness, we have to define *weak* and *weak-*topologies*.

Definition 2.8 (Weak Topology and Weak-*Topology).
Let $\mathcal{U}$ be a Banach space with dual space $\mathcal{U}^*$. Let $(u_k)_{k\in\mathbb{N}}$ and $(v_k)_{k\in\mathbb{N}}$ be two sequences in $\mathcal{U}$ and $\mathcal{U}^*$ respectively. Let be $u\in\mathcal{U}$ and $v\in\mathcal{U}^*$. Then the *weak topology* on $\mathcal{U}$ is defined by

$$u_k\rightharpoonup u\ :\Longleftrightarrow\ \langle v,u_k\rangle\to\langle v,u\rangle\qquad\forall\, v\in\mathcal{U}^*\ .$$

Furthermore, the *weak-*topology* on $\mathcal{U}^*$ is defined by

$$v_k\rightharpoonup^* v\ :\Longleftrightarrow\ \langle v_k,u\rangle\to\langle v,u\rangle\qquad\forall\, u\in\mathcal{U}\ .$$

Because we have $\mathcal{U}\subset\mathcal{U}^{**}$ the weak-*topology on $\mathcal{U}^*$ is even weaker than the weak topology on $\mathcal{U}$. Note that the weak and weak-*topology coincide in reflexive Banach spaces (i.e. $\mathcal{U}=\mathcal{U}^{**}$).
We now want to propose the theorem of *Banach-Alaoglu*. This theorem allows us to deduce compactness from boundedness at least in the weak-*topology and is therefore a central result concerning compactness.

Theorem 2.2 (Theorem of Banach-Alaoglu).
Let $\mathcal{U}$ be a Banach space and $\mathcal{U}^*$ its dual space.
Then the set

$$\{v \in \mathcal{U}^* \mid \|v\|_{\mathcal{U}^*} \leq C\}$$

for $C > 0$ is compact in the weak-*topology.

A proof can for instance be found in Rudin (1973, p. 66-68, Chapter 3, Theorem 3.15). In case we can prove lower semi-continuity in the weak-*topology, we are able to prove existence of a global minimum for a given infinite dimensional optimization problem by using Theorem 2.2 of Banach-Alaoglu. Then we could obtain the minimum by computing the Fréchet-derivative, however, normally proving lower semi-continuity in the weak-*topology is not that easy.

2.1.2. Γ-Convergence

In order to facilitate the transition from discrete to continuum models, we introduce a special convergence for functionals, namely Γ-*convergence*. The following definitions and properties can be found amongst others in Braides (2002) and Maso (1993).

Definition 2.9 (Γ-Convergence).
Let $\mathcal{U}$ be a topological space. Furthermore, let $J_k : \mathcal{U} \longrightarrow [0, \infty]$ be a sequence of functionals on $\mathcal{U}$. The sequence $(J_k)_{k\in\mathbb{N}}$ converges in the sense of Γ-convergence to the Γ-limit $J : \mathcal{U} \longrightarrow [0, \infty]$ in the case that the following conditions hold:

- *Lower bound inequality:* For every sequence $(u_k)_{k\in\mathbb{N}}$ with $u_k \in \mathcal{U}$ such that $u_k \to u$ for $k \to \infty$ holds
$$J(u) \leq \liminf_{k\to\infty} J_k(u_k) \ .$$
- *Upper bound inequality:* For all $u \in \mathcal{U}$ exists a sequence $(u_k)_{k\in\mathbb{N}}$ with $u_k \to u$ for $k \to \infty$ such that
$$J(u) \geq \limsup_{k\to\infty} J_k(u_k) \ .$$

Note that the meaning of the first condition is that J provides an asymptotic common lower bound for $(J_k)_{k\in\mathbb{N}}$. The second condition, however, yields the fact that this lower bound is optimal.
In case that $(J_k)_{k\in\mathbb{N}}$ Γ-converges to J we write

$$J_k \xrightarrow{\Gamma} J \ .$$

In order to facilitate the proof of the upper bound inequality, one may use the following density argument, which we cite from Braides (2002, Remark 1.29, p. 32).

Theorem 2.3 (Upper Bound Inequality by Density)**.**
Let d' be a distance on the metric space $(\mathcal{U},\ d)$ with stronger topology than the one induced by d, i.e. for every sequence $u_k \to u$ in $\mathcal{U}$ holds that

$$d'(u_k, u) \to 0 \qquad \Rightarrow \qquad d(u_k, u) \to 0 \ .$$

Furthermore, let $J_k : \mathcal{U} \longrightarrow [0, \infty]$ be a sequence of functionals on $\mathcal{U}$ and let the functional $J : \mathcal{U} \longrightarrow [0, \infty]$ be continuous with respect to d. Moreover, let $\mathcal{D}$ be a dense subset of $\mathcal{U}$ for d' and let for every $u \in \mathcal{U}$ exist a sequence $u_k \to u$ for $k \to \infty$ such that

$$\limsup_{k\to\infty} J_k(u_k) \leq J(u) \tag{2.3}$$

holds on $\mathcal{D}$. Then (2.3) holds on $\mathcal{U}$ as well.

In order to obtain convergence of minimizers, we need equi-coercivity of a sequence of functionals.

Definition 2.10 (Coerciveness Conditions)**.**
Let $\mathcal{U}$ be a topological space. A functional $J : \mathcal{U} \longrightarrow \overline{\mathbb{R}}$ is called

- *coercive* if for every $c \in \mathbb{R}$ holds that the set $\{J(u) \leq c\}$ is pre-compact,
- *mildly coercive* if there exists a set $K \subset \mathcal{U}$, which is non-empty and compact and it holds
$$\inf_{u\in\mathcal{U}} J(u) = \inf_{u\in K} J(u) \ .$$

Furthermore, a sequence $(J_k)_{k\in\mathbb{N}}$ of functionals $J_k : \mathcal{U} \longrightarrow \overline{\mathbb{R}}$ is called *equi-mildly coercive* or *equi-coercive* if there exists a set $K \subset \mathcal{U}$, which is non-empty and compact and for all $k \in \mathbb{N}$ holds
$$\inf_{u\in\mathcal{U}} J_k(u) = \inf_{u\in K} J_k(u) \ .$$
Cicalese et al. (2009) state an equivalent definition of equi-coercivity for the L^2-setting, which we cite for the general case.

Definition 2.11 (Equi-Coercivity)**.**
Let $\mathcal{U}$ be a Banach space. A sequence $(J_k)_{k\in\mathbb{N}}$ of functionals $J_k : \mathcal{U} \longrightarrow \overline{\mathbb{R}}$ is called *equi-coercive* if for all sequences $(u_k)_{k\in\mathbb{N}} \subset \mathcal{U}$ with

$$\sup_{k\in\mathbb{N}} J_k(u_k) < \infty \ ,$$

up to subsequences, holds that $u_k \to u$ in $\mathcal{U}$ for $k \to \infty$ and $u \in \mathcal{U}$.

Let us now state a fundamental result about the convergence of minimizers under Γ-convergence.

Theorem 2.4 (Convergence of Minimizers)**.**
Let $(\mathcal{U}, d)$ be a metric space. Moreover, let $(J_k)_{k\in\mathbb{N}}$ be a sequence of equi-coercive functionals and let $J_k \xrightarrow{\Gamma} J$ for $k \to \infty$.
Then there exists a minimizer of J with

$$\min_{u\in\mathcal{U}} J(u) = \lim_{k\to\infty} \inf_{u\in\mathcal{U}} J_k(u)\ .$$

Furthermore, in case that $(u_k)_{k\in\mathbb{N}}$ is a pre-compact sequence with

$$\lim_{k\to\infty} J_k(u_k) = \lim_{k\to\infty} \inf_{u\in\mathcal{U}} J_k(u)\ ,$$

every limit of a subsequence of $(u_k)_{k\in\mathbb{N}}$ is a minimizer of J.

A proof can be found in Braides (2002, Section 1.5).
Finally, we give a summary of some properties of Γ-convergence.

Remark 2.2 (Properties of Γ-Convergence)**.**

- Minimizers converge to minimizers: In case we have $J_k \xrightarrow{\Gamma} J$ and u_k minimizes J_k for every $k \in \mathbb{N}$, every cluster point of $(u_k)_{k\in\mathbb{N}}$ is a minimizer of J.
- Γ-limits are always lower semi-continuous.
- Γ-convergence is stable under continuous perturbations: Let $J_k \xrightarrow{\Gamma} J$ and let $H : \mathcal{U} \longrightarrow [0, \infty]$ be a continuous functional. Then $J_k + H$ Γ-converges to $J + H$.
- Let us consider a constant sequence of functionals, i.e. $J_k = J$ for all $k \in \mathbb{N}$. Then J_k does not necessarily Γ-converge to J. However, it Γ-converges to the relaxation of J, i.e. the largest lower semi-continuous functional below J.

2.2. Function Spaces and Sequence Spaces

This section summarizes basic definitions concerning function and sequence spaces, which are needed for the course of this thesis. For the definition of most of these spaces, basic measure theory is necessary, which we recall in the first subsection. Subsequently we shortly give a reminder of Lebesgue spaces and the sequence spaces ℓ^p, afterwards we recall Sobolev spaces. Since neither Lebesgue nor Sobolev spaces are appropriate for images, we introduce the space of functions with bounded variation. Furthermore, the space of finite Radon measures will be of particular importance for the analysis of asymptotic sparsity, whose introduction will end this section.

2.2.1. Basic Measure Theory

Let us now recall some basic definitions from measure theory. The content of this subsection is mainly based upon Elstrodt (2004).

Definition 2.12 (σ-Algebra)**.**
Let Ω be a set and $\mathcal{P}(\Omega)$ its power set. A collection $\mathcal{A}$ of subsets of $\mathcal{P}(\Omega)$ is called *σ-algebra* if the following conditions hold:

1. *Non-emptiness*: $\Omega \in \mathcal{A}$.

2. *Closure under complementation*: If $A \in \mathcal{A}$ holds, then we have $A^c \in \mathcal{A}$, where $A^c := \Omega \backslash A$.

3. *Closure under countable unions*: If $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}$ holds, then we have $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Example 2.1 (Borel σ-Algebra)**.**
An important example is the *Borel σ-algebra* $\mathcal{B}(X)$, which builds a bridge between measure theory and topology. Every topological space X can be endowed with the unique Borel σ-algebra, which is defined as the smallest σ-algebra containing all open sets of X. The elements of the Borel σ-algebra are called *Borel sets*.
This σ-algebra is named after Émile Borel (1950, Chapter 3), who implicitly introduced Borel subsets of the unit interval. For these sets he suggested a concept of length, which complies with the property of σ-additivity.

Definition 2.13 (Measure)**.**
Let $\mathcal{A}$ be a σ-algebra over a non-empty set Ω. A function $\mu\colon \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ is called a *measure* if it satisfies the following conditions:

1. *Non-negativity:* For all sets $A \in \mathcal{A}$ holds $\mu(A) \geq 0$.

2. *Null empty set:* $\mu(\emptyset) = 0$.

3. *σ-additivity:* For all countable collections $(A_n)_{n\in\mathbb{N}}$ of pairwise disjoint sets in $\mathcal{A}$ holds that
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right) \ .$$

Note that a measure is always monotonic, i.e. in case that $A \subset B$, we have $\mu(A) \leq \mu(B)$, cf. Figure 2.2.
The function μ is called a *signed measure* in the case that only the second and third conditions hold and μ takes on at most one of the values $\pm\infty$.

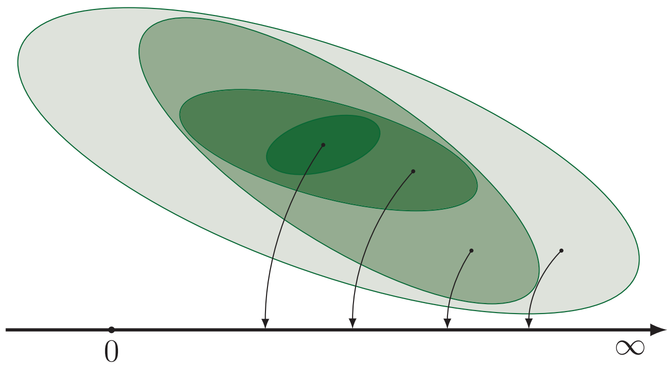


Figure 2.2.: Monotonicity of a measure

Definition 2.14 (Measurable Space & Measure Space)**.**
Let Ω be a non-empty set and $\mathcal{A}$ be a σ-algebra on Ω. We call the pair $(\Omega, \mathcal{A})$ a *measurable space*. The members of $\mathcal{A}$ are called *measurable sets*.
Let in addition $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ be a measure. Then the triple $(\Omega, \mathcal{A}, \mu)$ is called *measure space*.

Example 2.2.

1. Let $(\Omega, \mathcal{A})$ be a measurable space, $x \in \Omega$ and $A \in \mathcal{A}$. Then the *Dirac measure*, named after the English theoretical physicist Paul Dirac (1958), is defined as

$$\delta_x(A) := \begin{cases} 1, & \text{if } x \in A\ , \\ 0, & \text{else.} \end{cases}$$

2. Consider the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The *Lebesgue measure* on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ relates an geometric object to its content (length, area, volume, ...). It is the unique measure λ, which meets the following property:

$$\lambda\left([a_1, b_1] \times \ldots \times [a_n, b_n]\right) = (b_1 - a_1) \cdots (b_n - a_n)\ ,$$

i.e. it relates an n-dimensional hyperrectangle to its volume. The French mathematician Henri Lebesgue (1902) published this measure as part of his dissertation.

Definition 2.15 (Measurable Function)**.**
Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces. A function $u : \Omega_1 \longrightarrow \Omega_2$ is called *measurable* if the pre-image of every set $A \in \mathcal{A}_2$ under the function u is in $\mathcal{A}_1$, i.e.

$$u^{-1}(A) = \{x \in \Omega_1 \mid u(x) \in A\} \in \mathcal{A}_1 \qquad \forall\, A \in \mathcal{A}_2\ .$$

2.2.2. Lebesgue Spaces

We now shortly review the definition of Lebesgue spaces and their norms, namely L^p-norms. For the sake of compactness, we do not recall most of the measure theory and take the knowledge about Lebesgue integration for granted. We refer the reader to standard books on measure theory and analysis such as RUDIN (1987) in case further information should be necessary.

Definition 2.16 (The Lebesgue Spaces $L^p(\Omega)$ and $L^\infty(\Omega)$)**.**
Let be $p \in \mathbb{R}^+$. The space $L^p(\Omega)$ is defined as the set of all measurable functions $u{:}\Omega \longrightarrow \mathbb{C}$ on the domain Ω, for which holds

$$\int_\Omega |u(x)|^p \mathrm{d}x < \infty \, .$$

In case that we have $p = \infty$, $L^\infty(\Omega)$ is considered to be the space of all measurable functions $u{:}\Omega \longrightarrow \mathbb{C}$, for which there exists a constant $\alpha > 0$ with

$$\lambda\left(\{x \in \Omega \mid |u(x)| \geq \alpha\}\right) = 0 \, ,$$

where λ denotes the Lebesgue measure.

For the case that $p \geq 1$ holds, those spaces are naturally Banach spaces with the following norms:

Theorem 2.5 ($L^p(\Omega)$ and $L^\infty(\Omega)$ are Banach Spaces)**.**
Let be $1 \leq p < +\infty$. The space $L^p(\Omega)$ is a Banach space with norm

$$\|u\|_p = \left(\int_\Omega |u(x)|^p \, \mathrm{d}x\right)^{\frac{1}{p}} \, .$$

Furthermore, the space $L^\infty(\Omega)$ is a Banach space with norm

$$\|u\|_\infty = \operatorname*{ess\,sup}_{x\in\Omega} |u(x)| := \inf\left\{\alpha > 0 \mid \lambda\left(\{x \in \Omega \mid |u(x)| \geq \alpha\}\right) = 0\right\} \, .$$

Remark 2.3.
In the special case of $p = 2$ the Lebesgue space $L^2(\Omega)$ is even a Hilbert space with scalar product

$$\langle u, v\rangle_{L^2(\Omega)} := \int_\Omega \langle u(x), v(x)\rangle \, \mathrm{d}x \, . \tag{2.4}$$

2.2.3. Sequence Spaces ℓ^p

In the course of this thesis sequence spaces are elementary, thus we recall their definitions in this subsection. We concentrate on ℓ^p-spaces for $p \in \mathbb{R}^+ \cup \{+\infty\}$, since those are essential for our further results.

Definition 2.17 (Sequence Spaces $\ell^p(\Omega)$ and $\ell^\infty(\Omega)$)**.**
Let Ω be a field and let $u = (u_n)_{n\in\mathbb{N}}$ be a sequence with $u_n \in \Omega$ for all $n \in \mathbb{N}$. Let be $0 < p < \infty$. Then the subspace $\ell^p(\Omega)$ of the sequence space $\Omega^{\mathbb{N}}$ is defined as

$$\ell^p(\Omega) := \left\{ u \in \Omega^{\mathbb{N}} \,\middle|\, \sum_{n=1}^{\infty} |u_n|^p < \infty \right\} .$$

Furthermore, $\ell^\infty(\Omega)$ denotes the space of bounded sequences, i.e.

$$\ell^\infty(\Omega) := \left\{ u \in \Omega^{\mathbb{N}} \,\middle|\, \sup_{n\in\mathbb{N}} |u_n| < \infty \right\} .$$

Note that we obviously obtain the following relation:

$$\ell^p(\Omega) \subseteq \ell^\infty(\Omega) .$$

Similar to the case of L^p-spaces, the sequence spaces $\ell^p(\Omega)$ for $p \geq 1$ and $\ell^\infty(\Omega)$ are Banach spaces.

Theorem 2.6 ($\ell^p(\Omega)$ and $\ell^\infty(\Omega)$ are Banach Spaces)**.**
Let be $1 \leq p < +\infty$. The sequence space $\ell^p(\Omega)$ is a Banach space with norm

$$\|u\|_p = \left(\sum_{n=1}^{\infty} |u_n|^p \right)^{\frac{1}{p}} .$$

Note that $\ell^2(\Omega)$ is even a Hilbert space. Furthermore, $\ell^\infty(\Omega)$ is a Banach space with norm

$$\|u\|_\infty = \sup_{n\in\mathbb{N}} |u_n| .$$

In Figure 2.3 we see unit circles for different ℓ^p-norms. Since the unit circles for ℓ^p-"norms" with $0 < p < 1$ are not convex, we especially learn from this illustration that ℓ^p-"norms" with $0 < p < 1$ satisfy all of the norm axioms except for the triangle inequality. However, they fulfill the weaker inequality

$$\|x + y\|_p \leq k \left(\|x\|_p + \|y\|_p \right) \qquad \text{for} \qquad k > 1$$

instead and thus are at least *quasi-norms*.

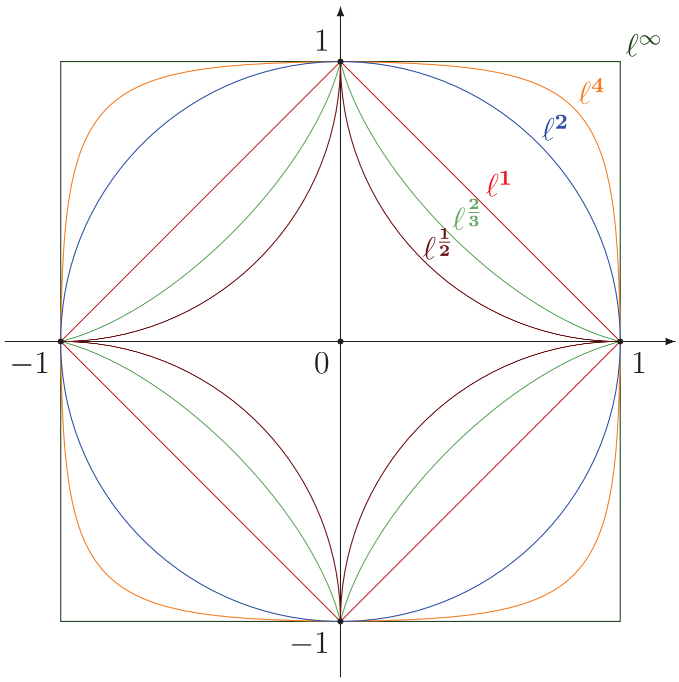


Figure 2.3.: Illustration of unit circles in different ℓ^p-norms and quasi-norms

2.2.4. Sobolev Spaces

The Lebesgue spaces $L^p(\Omega)$ contain many different functions including highly oscillating functions. In application to inverse problems (cp. Section 1.2) functions with high oscillation usually mean noise and thus are not desirable. As an alternative one might search for functions, whose derivatives are additionally Lebesgue integrable. This idea leads us to so-called *Sobolev spaces*, which we recall in this chapter. In order to do so, we first introduce the concept of weak derivatives.

Definition 2.18 (Weak Derivative)**.**
Let $\Omega \subset \mathbb{R}^N$ be an open set and let $\alpha \in \mathbb{N}_0^N$ be a multiindex. Let u be locally L^1-integrable, i.e. $u \in L^1_{\text{loc}}(\Omega)$. A function $w \in L^1_{\text{loc}}(\Omega)$ with

$$\int_\Omega w\varphi \,\mathrm{d}x = (-1)^{|\alpha|} \int_\Omega u \,\mathrm{D}^\alpha \varphi \,\mathrm{d}x \qquad \forall\, \varphi \in C_0^\infty(\Omega)$$

is called *weak partial derivative* of order $|\alpha|$ of u. In this context we have $|\alpha| = \sum_{i=1}^N \alpha_i$ and $\mathrm{D}^\alpha = \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} x_1 \ldots \partial^{\alpha_N} x_N}$. The weak derivative is often denoted by $w = \mathrm{D}^\alpha u$.

Note that a weak derivative does not have to exist.
Since the space $L^1_{\text{loc}}(\Omega)$ is not quite convenient, we rather consider functions with weak derivatives in Lebesgue spaces. This brings us to the definition of so-called Sobolev spaces.

Definition 2.19 (Sobolev Spaces).
Let be $1 \leq p \leq \infty$, $k \in \mathbb{N}$ and Ω be an open set. The *Sobolev space* $W^{k,p}(\Omega)$ is defined as

$$W^{k,p}(\Omega) := \{u \in L^p(\Omega) \mid \mathrm{D}^\alpha u \in L^p(\Omega) \text{ for all } |\alpha| \leq k\}$$

and is naturally a Banach spaces with *Sobolev norm*

$$\|u\|_{W^{k,p}(\Omega)} := \left(\sum_{|\alpha| \leq k} \|\mathrm{D}^\alpha u\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}}$$

for $1 \leq p < \infty$ and

$$\|u\|_{W^{k,\infty}(\Omega)} := \max_{|\alpha| \leq k} \|\mathrm{D}^\alpha u\|_{L^\infty(\Omega)} .$$

Remark 2.4.
By using the scalar product (2.4) of $L^2(\Omega)$, we obtain a scalar product in $W^{k,2}(\Omega)$ by

$$\langle u, v \rangle_{W^{k,2}(\Omega)} := \sum_{|\alpha| \leq k} \langle D^\alpha u, D^\alpha v \rangle_{L^2(\Omega)} .$$

Thus we see that $W^{k,2}(\Omega)$ is a Hilbert space, which is often denoted by

$$H^k(\Omega) := W^{k,2}(\Omega) \qquad \forall\, k \in \mathbb{N} .$$

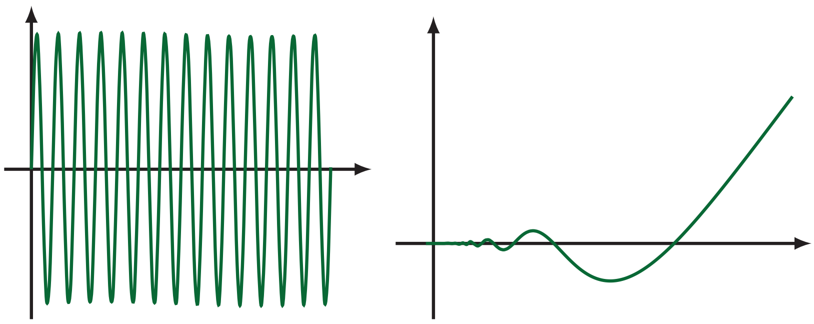
2.2.5. Space of Functions with Bounded Variation

Let us now introduce the space of functions with bounded variation, which will be used in the course of this thesis. This function space is used for images with certain properties. One could ask, why either the Lebesgue spaces $L^1(\Omega)$ and $L^2(\Omega)$ or the Sobolev space $W^{1,1}(\Omega)$ are not appropriate for this assignment. The problem is, however, that $L^1(\Omega)$ and $L^2(\Omega)$ contain not only the images we have in mind, but also the noise, which is not desirable. Considering $W^{1,1}(\Omega)$ instead is neither advantageous, since $W^{1,1}(\Omega)$ does not contain functions with discontinuities. This property, however, would be recommendable in order to examine reasonable images, which naturally come with discontinuities, i.e. sharp edges.
In this subsection we only give a short summary of functions with bounded variation. However, we refer the interested reader to Burger et al. (2013) and Ambrosio et al. (2000) for further details on this subject.
In this subsection we assume $\Omega \subset \mathbb{R}^N$ to be sufficiently regular and open. We begin by recalling the definition of the *total variation* of a function u.

Definition 2.20 (Total Variation).
The *total variation* $\mathrm{TV}(u)$ of a function $u \in L^1(\Omega)$ is defined as

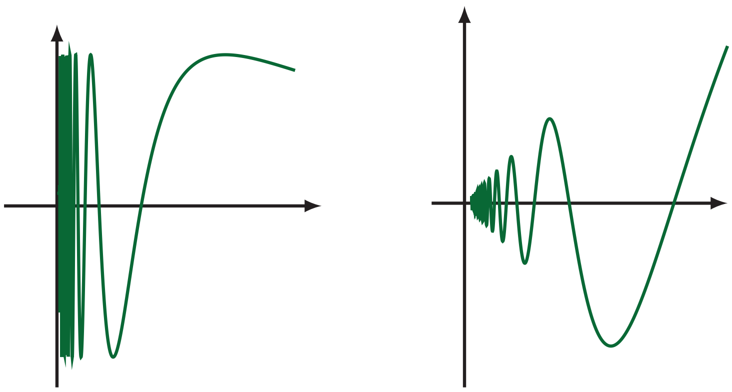
$$\mathrm{TV}(u) := \sup_{\substack{p \in C_0^\infty(\Omega) \\ \|p\|_{L^\infty(\Omega)} \leq 1}} \int_\Omega u \, \nabla \cdot p \, \mathrm{d}x \; . \tag{2.5}$$

In Figures 2.4 and 2.5 we see examples for functions with bounded and unbounded total variation.



(a) $f(x) = \sin(\alpha x)$ (b) $g(x) = x^2 \sin\left(\frac{1}{x}\right)$

Figure 2.4.: Functions with *bounded* total variation



(a) $h(x) = \sin(\frac{1}{x})$ (b) $k(x) = x \sin(\frac{1}{x})$

Figure 2.5.: Functions with *unbounded* total variation

By using this definition, we are able to define the space of functions with bounded variation.

Definition 2.21 (The Space $\mathrm{BV}(\Omega)$).
The *space of functions with bounded variation* $\mathrm{BV}(\Omega)$ is defined as the space of all functions $u \in L^1(\Omega)$ with finite total variation $\mathrm{TV}(u)$, i.e.

$$\mathrm{BV}(\Omega) := \left\{ u \in L^1(\Omega) \mid \mathrm{TV}(u) < \infty \right\} \; .$$

BV(Ω) is a Banach space equipped with the norm

$$\|u\|_{\mathrm{BV}(\Omega)} := \mathrm{TV}(u) + \|u\|_{L^1(\Omega)} \ ,$$

as we can see in Giusti (1984).

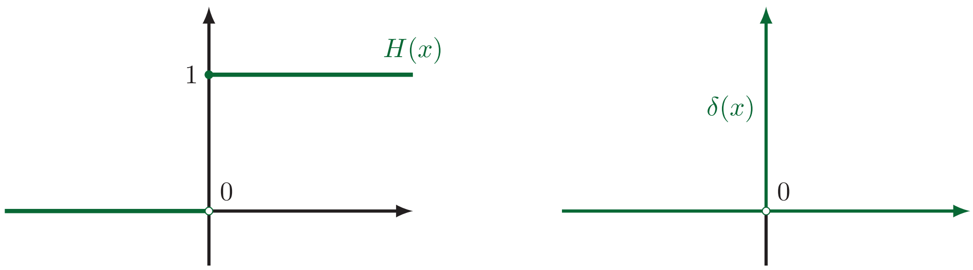
Remark 2.5.
In case the function of interest is sufficiently smooth, i.e. $u \in W^{1,1}(\Omega)$, the definition of $\mathrm{TV}(u)$ can be written in a primal setting as

$$\mathrm{TV}(u) = \int_\Omega |\nabla u| \,\mathrm{d}x \ . \tag{2.6}$$

In an informal setting this definition is often used for general $u \in \mathrm{BV}(\Omega)$.
Furthermore, we see that the Sobolev space $W^{1,1}(\Omega)$ is contained in BV(Ω). However, the *Heaviside function*

$$\mathrm{H}(x) := \begin{cases} 1, & \text{if } x \geq 0\,, \\ 0, & \text{else,} \end{cases}$$

which is contained in BV(Ω), can be indicative for the fact that this is a strict inclusion, since its distributional derivative, i.e. the *Dirac delta distribution* $\delta(x)$, is singular with respect to the Lebesgue measure, cf. Figure 2.6.



(a) Heaviside function (b) Dirac delta distribution

Figure 2.6.: The Heaviside function $\mathrm{H}(x)$ and its weak derivative $\delta(x)$

In addition BV(Ω) can be nicely embedded in $L^p(\Omega)$, which we see in the following Lemma. This property is especially useful for $N \leq 2$.

Lemma 2.1 (Embedding of BV(Ω))**.**
Let be $\Omega \subset \mathbb{R}^N$. If $\frac{p}{p-1} \geq N$, then we have

$$\mathrm{BV}(\Omega) \hookrightarrow L^p(\Omega) \ .$$

In case that $\frac{p}{p-1} > N$, the embedding is compact.
A proof can be found in Giusti (1984).

In order to make working with the total variation easier, it is important to note that $\mathrm{TV}(u)$ is lower semi-continuous in $\mathrm{BV}(\Omega)$ with respect to the strong topology in $L^1_{\mathrm{loc}}(\Omega)$ (c.f. AMBROSIO et al. (2000, Proposition 3.6)). If we want to give the optimality condition in terms of subgradients, we additionally need convexity of the total variation. However, this can be seen rather easily.

Lemma 2.2 (Convexity of the Total Variation)**.**
The total variation $\mathrm{TV}(u)$ with its definition in (2.5) is convex.

Proof.
Let be $u,\ v \in L^1(\Omega)$ and let $\alpha \in [0,1]$ hold. Then we have

$$\begin{aligned}\mathrm{TV}(\alpha u + (1-\alpha)v) &= \sup_{\|p\|_{L^\infty(\Omega)}\leq 1} \int_\Omega (\alpha u + (1-\alpha)v)\ \nabla\cdot p \,\mathrm{d}x \\ &\leq \alpha \sup_{\|p\|_{L^\infty(\Omega)}\leq 1} \int_\Omega u\ \nabla\cdot p \,\mathrm{d}x + (1-\alpha) \sup_{\|p\|_{L^\infty(\Omega)}\leq 1} \int_\Omega v\ \nabla\cdot p \,\mathrm{d}x \\ &= \alpha \mathrm{TV}(u) + (1-\alpha)\mathrm{TV}(v)\ .\end{aligned}$$

□

2.2.6. Space of Finite Radon Measures

In the course of this thesis we will amongst others work with linear functionals with compact support. In order to examine these functionals in a more theoretical way, we introduce the space of finite Radon measures. For its definition we previously recall Radon measures and important representation theorems.
The content of this subsection is mainly based upon ELSTRODT (2004). However, many of the results can be found in SCHWARTZ (1973) and RUDIN (1987) as well.
In order to introduce Radon measures, we first need some further definitions.

Definition 2.22 ((Locally) Finite Measure)**.**
Let (X,τ) be a topological Hausdorff space and let $\mathcal{A} \supset \mathcal{B}(X)$ be a σ-algebra on X, which contains the topology τ. The (signed) measure $\mu\colon \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ on the measurable space $(X,\mathcal{A})$ is called

- *finite* if $|\mu(X)| < \infty$,
- *σ-finite* if X is the countable union of measurable sets with finite measure,
- *locally finite* if for every point $x \in X$ exists an open neighbourhood N of x such that $|\mu(N)| < \infty$.

Definition 2.23 (Regular Measure)**.**
Let (X, τ) be a topological Hausdorff space and let $\mathcal{A} \supset \mathcal{B}(X)$ be a σ-algebra on X, which contains the topology τ. The measure $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ on the measurable space $(X, \mathcal{A})$ is called

1. *inner regular* if for every set $A \in \mathcal{A}$ holds that

$$\mu(A) = \sup\{\mu(K) \mid K \subset A,\, K \text{ compact}\}\ ,$$

2. *outer regular* if for every set $A \in \mathcal{A}$ holds that

$$\mu(A) = \inf\{\mu(U) \mid U \supset A,\, U \text{ open}\}\ ,$$

3. *regular* if μ is inner and outer regular.

Now we have everything that is required for the definition of Radon measures.

Definition 2.24 (Radon Measure)**.**
Let (X, τ) be a topological Hausdorff space and let $\mathcal{B}(X)$ be the Borel σ-algebra on X. A *Radon measure* is a (signed) measure $\mu : \mathcal{B}(X) \longrightarrow \overline{\mathbb{R}}$ on the measurable space $(X, \mathcal{B}(X))$, which is locally finite and inner regular.

Let us now introduce total variation for signed measures. For this purpose we first need a decomposition theorem, which traces back the to Austrian mathematician Hans Hahn (1921).

Theorem 2.7 (Hahn Decomposition Theorem)**.**
Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ be a signed measure. Then there exists a positive measurable set P and a negative measurable set N in $\mathcal{A}$ such that $P \cup N = \Omega$ and $P \cap N = \emptyset$. P and N are unique except for μ-empty sets.
Furthermore, $\Omega = P \cup N$ is called *Hahn decomposition*.

Definition 2.25 (Variation of Measures)**.**
Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ be a signed measure with Hahn decomposition $\Omega = P \cup N$. For $A \in \mathcal{A}$ we call

1. $\mu^+ : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ with

$$\mu^+(A) := \mu(A \cap P)$$

   positive variation,

2. $\mu^- : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ with

$$\mu^-(A) := -\mu(A \cap N)$$

   negative variation and

3. $\|\mu(A)\|_{\text{TV}} : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ with

$$\|\mu(A)\|_{\text{TV}} := \mu^+(A) + \mu^-(A)$$

total variation of μ.

Definition 2.26 (Regular Signed Measure)**.**
Let X be a locally compact Hausdorff space and $\mathcal{B}(X)$ the Borel σ-algebra on X. A signed measure $\mu : \mathcal{B}(X) \longrightarrow \mathbb{R}$ is called *regular* if for all $A \in \mathcal{B}(X)$ and $\varepsilon > 0$ exists a compact set K and an open set U such that $K \subset A \subset U$ and $\|\mu(U \backslash K)\|_{\text{TV}} < \varepsilon$.
Furthermore, we define $\mathcal{M}(X)$ as the set of finite regular signed measures $\mu : \mathcal{B}(X) \longrightarrow \mathbb{R}$.

Remark 2.6.
$\mathcal{M}(X)$ is even a Banach space with norm $\|\mu\|_{\text{TV}(X)} := \|\mu(X)\|_{\text{TV}}$.
Note that a finite regular signed measure is the same as a finite signed Radon measure, since a finite Radon measure is always regular in case that X is locally compact. In the literature these terms are used interchangeably. However, we act in accordance with the literature used in the ongoing chapters and call $\mathcal{M}(X)$ the *space of finite Radon measures*.
In a more general setting we could consider the space $\mathcal{M}(X, \mathbb{R}^m)$ of finite vector-valued Radon measures $\mu : \mathcal{B}(X) \longrightarrow \mathbb{R}^m$, however, in this chapter it is sufficient to use $\mathcal{M}(X)$.

Definition 2.27 (Absolutely Continuous Measure)**.**
Let μ and ν be two measures on the measurable space $(\Omega, \mathcal{A})$. Then we call μ *absolutely continuous with respect to* ν if for every set $A \in \mathcal{A}$ holds

$$\nu(A) = 0 \quad \Rightarrow \quad \mu(A) = 0 \; .$$

Absolute continuity of measures is a preorder and we write $\mu \ll \nu$. The measures μ and ν are said to be *equivalent* if it holds $\mu \ll \nu$ and $\nu \ll \mu$.

Note that if we consider the case of signed measures, μ is absolutely continuous with respect to ν if its total variation $\|\mu\|_{\text{TV}(\Omega)}$ satisfies $\|\mu\|_{\text{TV}(\Omega)} \ll \nu$.

Another important term in measure theory is the concept of density functions of measures. Thus in order to shortly introduce density functions for measures, we recall the *Radon-Nikodym theorem*, which the Polish mathematician Otto Nikodym (1930) proofed in its general version.

Theorem 2.8 (Radon-Nikodym Theorem).
Let μ and ν be two σ-finite measures on the measurable space $(\Omega, \mathcal{A})$. Let μ be absolutely continuous with respect to ν. Then there exists a measurable function $\rho\colon\Omega \longrightarrow [0,\infty)$ such that for any measurable subset $A \in \mathcal{A}$ holds

$$\mu(A) = \int_A \rho \, \mathrm{d}\nu \; .$$

The function ρ is called *Radon-Nikodym derivative* or *density function* and is also denoted by $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.
Note that ρ is uniquely defined up to a ν-null set.

Let us now consider Radon measures on locally compact spaces. The idea is to consider a locally compact topological space X as underlying measure space. By doing so Radon measures can be expressed in terms of continuous linear functionals on the space of continuous functions with compact support, i.e $C_c(X)$. This approach, which has amongst others been taken by Bourbaki (2004), allows the development of measure and integration theory in terms of functional analysis. However, in order to identify continuous linear functionals on $C_c(X)$ with Radon measures, we recall an essential result. The following theorem traces back to the Hungarian mathematician Frigyes Riesz (1909), who stated the theorem for continuous functions on the unit interval. Then Andrey Markov (1938) extended the result to some non-compact spaces. Finally, Shizuo Kakutani (1941) further extended the assertion to compact Hausdorff spaces.

Theorem 2.9 (Riesz-Markov-Kakutani Representation Theorem).
Let X be a locally compact Hausdorff space and $\mathcal{B}(X)$ the Borel σ-algebra on X. Let furthermore $I\colon C_c(X) \longrightarrow \mathbb{R}$ be a non-negative linear functional. Then there exists a unique Radon measure $\mu\colon\mathcal{B}(X) \longrightarrow [0,\infty]$ such that

$$I(\psi) = \int_X \psi \, \mathrm{d}\mu$$

for all $\psi \in C_c(X)$ with

$$\mu(K) = \inf\{I(\psi) \mid \psi \in C_c(X),\ \psi \geq \chi_K,\ K \text{ compact}\} \; ,$$
$$\mu(A) = \sup\{\mu(K) \mid K \subset A,\ K \text{ compact},\ A \in \mathcal{B}(X)\} \; .$$

This theorem can be extended in several ways. We just want to give a brief summary about representations of non-negative linear functionals.

Lemma 2.3 (Summary Representation Theorems).
Let X be a locally compact Hausdorff space and $\mathcal{B}(X)$ the Borel σ-algebra on X. Then we have equivalence between the non-negative linear functionals on

1. $C_c(X)$ and Radon measures on $\mathcal{B}(X)$,

2. $C_0(X)$ and finite Radon measures on $\mathcal{B}(X)$,

3. $C(X)$ and Radon measures with compact support in case that X is σ-compact.

By using Banach spaces of finite signed Radon measures, these representation theorems allow the description of certain Banach spaces of continuous functions. However, the general introduction to the theory using signed or complex Radon measures is quite challenging. For lack of space further details shall be omitted and we concentrate on $(C_0(X), \|\cdot\|_\infty)$. The interested reader may find further information in Schwartz (1973). For our purposes it shall be sufficient to consider Theorem 2.9 for finite signed Radon measures. In so doing, we will determine the dual space of $C_0(X)$, where X is a locally compact Hausdorff space.

Theorem 2.10 (Representation Theorem for the Dual of $C_0(X)$).
Let X be a locally compact Hausdorff space. Then

$$\begin{aligned} &\Phi : \mathcal{M}(X) \longrightarrow C_0^*(X) \qquad \text{with} \\ &\Phi(\mu) := \int_X \psi \, \mathrm{d}\mu \qquad \forall\, \psi \in C_0(X) \, , \end{aligned}$$

is a monotonic isometric isomorphism, for which holds

$$\|\Phi(\mu)\| = \|\mu\|_{\mathrm{TV}(X)} \ .$$

For lack of space, we omit further deductions and proofs of this subsection and refer again to Elstrodt (2004) and Schwartz (1973).

Remark 2.7.
Concluding, we note that the total variation of a measure may equivalently be defined as

$$\|\mu\|_{\mathrm{TV}(X)} := \sup \left\{ \int_X \psi \, \mathrm{d}\mu \mid \psi \in C_0(X), \ \|\psi\|_\infty \leq 1 \right\} \ .$$

2.3. Convex Analysis

In this section we recall some results from convex analysis needed in this thesis. For the sake of brevity, we will only discuss some subdifferential calculus and end with the concept of Legendre-Fenchel duality. The reader may find further and more detailed information on convex analysis in Boyd and Vandenberghe (2004), Ekeland and Témam (1999), Rockafellar (1970) or Hiriart-Urruty and Lemaréchal (1993, Chapter 4).

We start by recalling the definitions of a convex set and subsequently a convex functional.

Definition 2.28 (Convex Set)**.**
Let $\mathcal{U}$ be a Banach space. A subset $U \subset \mathcal{U}$ is called *convex* if for all $u,\ v \in U$ holds

$$\lambda u + (1-\lambda)v \ \in \ U \qquad \forall\, \lambda \in [0,1]\ .$$

Definition 2.29 (Convex Functional)**.**
Let U be a convex set. A functional $J : U \longrightarrow \mathbb{R} \cup \{\infty\}$ is called *convex* if for all $u,\ v \in U$ holds

$$J(\lambda u + (1-\lambda)v) \leq \lambda J(u) + (1-\lambda)J(v) \qquad \forall\, \lambda \in [0,1]\ . \tag{2.7}$$

J is called *strictly convex* if (2.7) is a strict inequality for all $\lambda \in (0,1)$ and $u \neq v$. In the case that J and U are convex, the optimization problem (1.4) is also called *convex*.

In Figure 2.7 we see the illustration of Definition 2.29 for J being a strictly convex function.

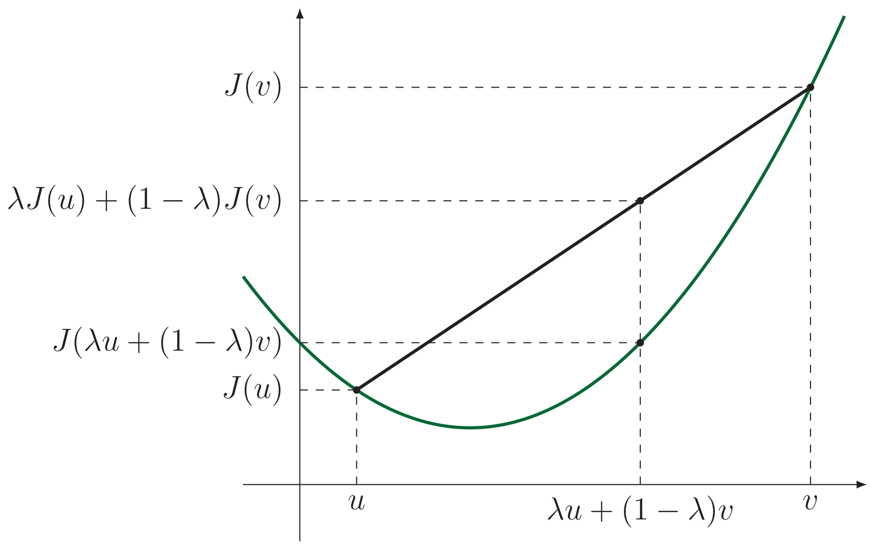


Figure 2.7.: Strictly convex function

2.3.1. Subdifferential Calculus

In order to include functionals, which are not Fréchet-differentiable, into our considerations, a generalized concept of differentiability is necessary. Thus we introduce the *subdifferential* of a convex functional to gain some insight about the solution of a variational problem.

Definition 2.30 (Subdifferential)**.**
Let $\mathcal{U}$ be a Banach space with dual space $\mathcal{U}^*$. Furthermore, let $J : \mathcal{U} \longrightarrow \mathbb{R} \cup \{\infty\}$ be a proper convex functional.
Then J is called *subdifferentiable* at $u \in \mathcal{U}$ in case there exists an element $p \in \mathcal{U}^*$ such that

$$J(v) - J(u) - \langle p, v - u \rangle_{\mathcal{U}} \geq 0 \qquad \forall\, v \in \mathcal{U}\,. \tag{2.8}$$

Moreover, p is called *subgradient* at position u.
The collection of all subgradients at position u, i.e.

$$\partial J(u) := \{p \in \mathcal{U}^* | J(v) - J(u) - \langle p, v - u \rangle_{\mathcal{U}} \geq 0, \forall\, v \in \mathcal{U}\} \subset \mathcal{U}^*\ ,$$

is called *subdifferential* of J at u.

In the special case that J is Fréchet-differentiable the subdifferential coincides with the classical definition, i.e. $\partial J(u) = \{J'(u)\}$.

Example 2.3 (Subdifferential of the ℓ^1-Norm)**.**
Let $J : \mathbb{R} \longrightarrow \mathbb{R}^+$ be the absolute value function, i.e. $J(u) = |u|$. Since $J(u)$ is differentiable for $u \neq 0$ the subdifferential of the absolute value function in this case is obviously given by

$$\partial J(u) = \begin{cases} \{1\}\,, & \text{for } u > 0 \\ \{-1\}\,, & \text{for } u < 0 \end{cases}\,.$$

By using Definition 2.30, we see that $\partial|0| = [-1, 1]$ holds. Therefore, the subdifferential of the ℓ^1-norm can be characterized componentwise like the absolute value function, i.e.

$$p \in \partial \|u\|_1 \Leftrightarrow \begin{cases} |p_i| \leq 1, & \text{if } u_i = 0\ , \\ p_i = \operatorname{sign}(u_i), & \text{else}\ . \end{cases}$$

A subgradient of the absolute value function is displayed in Figure 2.8.

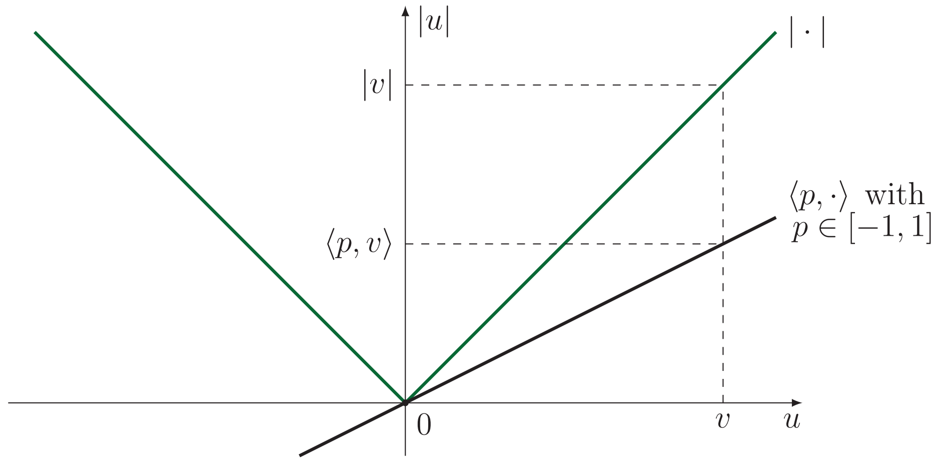


Figure 2.8.: Visualization of a subgradient $p \in \partial|0|$ of the absolute value function.

In variational methods the functionals, which shall be minimized, usually consist of the sum of a data fidelity term and at least one regularization functional. Computing the subdifferential of such functionals is not always straight forward. To simplify the work with those types of minimization problems, we additionally recall the following properties of the subdifferential:

Lemma 2.4 (Additivity of the Subdifferential)**.**
Let $\mathcal{U}$ be a real Banach space and let $J_1, J_2 : \mathcal{U} \longrightarrow \mathbb{R} \cup \{\infty\}$ be proper convex functionals, which are furthermore lower semi-continuous. If there exists a point $\bar{u} \in \mathrm{dom} J_1 \cap \mathrm{dom} J_2$, where J_1 is continuous, it holds that

$$\partial \left(J_1 + J_2\right)(u) = \partial J_1(u) + \partial J_2(u) \quad \forall\, u \in \mathcal{U} \ .$$

A proof can be found in Ekeland and Témam (1999, Chapter 1, Proposition 5.6). Since we are quite often confronted with 1-homogeneous functionals, we characterize their subdifferential at this point.

Lemma 2.5 (Subdifferential of 1-Homogeneous Functionals)**.**
Let $\mathcal{U}$ be a real Banach space with dual space $\mathcal{U}^*$ and let the proper convex functional $J : \mathcal{U} \longrightarrow \mathbb{R} \cup \{\infty\}$ be 1-homogeneous, i.e. $J(\alpha u) = \alpha J(u), \ \forall\ \alpha > 0$. Then the subdifferential of J reads as follows:

$$\partial J(u) = \{p \in \mathcal{U}^* |\, \langle p, u\rangle_{\mathcal{U}} = J(u),\ \langle p, v\rangle_{\mathcal{U}} \leq J(v)\ \forall\ v \in \mathcal{U}\} \ .$$

Proof.
Using the definition of the subgradient (2.8) yields

$$\langle p, v-u\rangle_{\mathcal{U}} \leq J(v) - J(u) \quad \forall\, v \in \mathcal{U} \,.$$

If we choose $v = 0$, we obtain

$$\langle p, u\rangle_{\mathcal{U}} \geq J(u) \,.$$

By choosing $v = 2u$ and using the one-homogenity of J, we furthermore obtain

$$\langle p, u\rangle_{\mathcal{U}} \leq J(2u) - J(u) = 2J(u) - J(u) = J(u) \,.$$

Hence we have $\langle p, u\rangle_{\mathcal{U}} = J(u)$. Inserting this into the definition of the subgradient (2.8) we finally obtain

$$\langle p, v\rangle_{\mathcal{U}} \leq J(v) \quad \forall\, v \in \mathcal{U} \,.$$

□

The subdifferential is very useful to compute optimality conditions for general convex variational problems like (1.4). We gain some insight about its optimal solution by using the following theorem.

Theorem 2.11 (Subdifferential and Optimality)**.**
Let $\mathcal{U}$ be a Banach space and let $J\colon\mathcal{U} \longrightarrow \mathbb{R} \cup \{\infty\}$ be a proper convex functional. An element $u \in \mathcal{U}$ is a minimizer of J if and only if $0 \in \partial J(u)$.

Proof.
By Definition 2.30 of the subdifferential we have

$$0 = \langle 0, v-u\rangle_{\mathcal{U}} \leq J(v) - J(u) \qquad \forall\, v \in \mathcal{U}$$

for $0 \in \partial J(u)$. Therefore, u is a global minimizer of J.
In case that $0 \notin \partial J(u)$ holds, there exists at least one $v \in \mathcal{U}$ with

$$J(v) - J(u) < \langle 0, v-u\rangle_{\mathcal{U}} = 0 \,.$$

Hence u cannot be a minimizer of J. □

Since our functionals are convex, the first-order optimality condition is not only necessary, but also sufficient.

2.3.2. Legendre-Fenchel Duality

In the analysis of variational problems like (1.4) the concept of duality, which we will recall shortly in this subsection, is very useful. More details on this subject can be found in Rockafellar (1970), Rockafellar and Wets (1997, Chapter 11) or Boyd and Vandenberghe (2004, Section 3.3 and 5).
After primarily introducing the concept of the convex conjugate, we are going to give some results and examples. Finally we end this subsection by stating *Fenchel's duality theorem*, which is one of the main results in convex analysis.

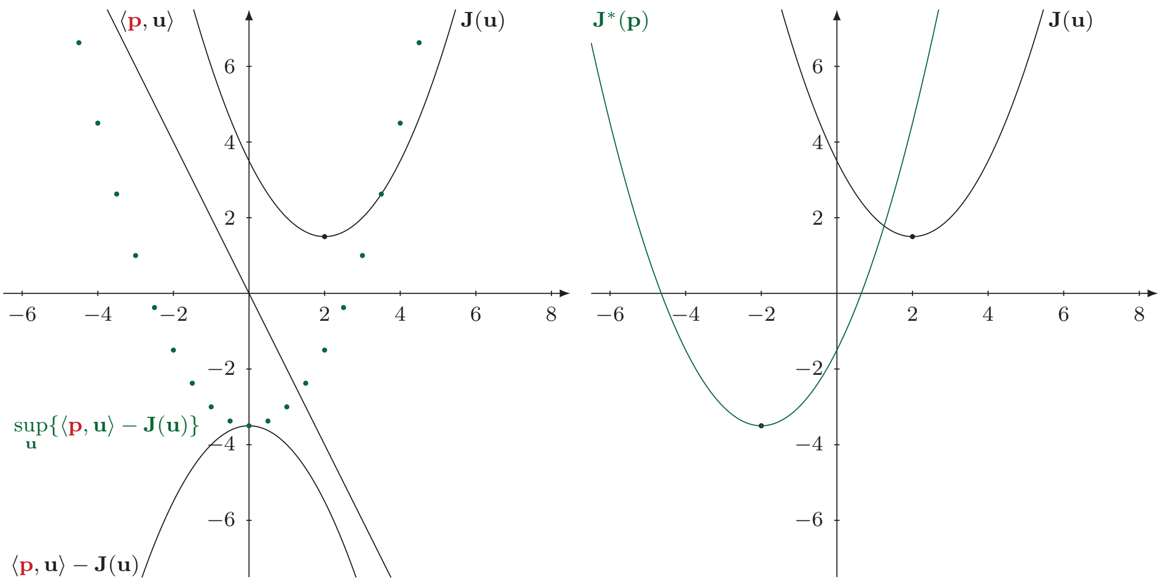
Definition 2.31 (Convex Conjugate)**.**
Let $\mathcal{U}$ be a Banach space with dual space $\mathcal{U}^*$. The *convex conjugate* $J^* : \mathcal{U}^* \longrightarrow \overline{\mathbb{R}}$ of a functional $J : \mathcal{U} \longrightarrow \overline{\mathbb{R}}$ is defined by

$$J^*(p) := \sup_{u \in \mathcal{U}} \left\{ \langle p, u \rangle_{\mathcal{U}} - J(u) \right\} \ .$$

The *biconjugate* $J^{**} : \mathcal{U} \longrightarrow \overline{\mathbb{R}}$ of J is defined as

$$J^{**}(u) := \sup_{p \in \mathcal{U}^*} \left\{ \langle u, p \rangle_{\mathcal{U}^*} - J^*(p) \right\} \ .$$

Note that the convex conjugate is also known as *Legendre-Fenchel transform*, which was named after Adrien-Marie Legendre and Werner Fenchel.



(a) In this example $h_p(u) = \langle p, u \rangle - J(u)$ is plotted for $p = -2$. To obtain the convex conjugate, we have to vary p (red). The green dots indicate the maxima of $h_p(u)$ for different values of p, however, they are are still plotted against u.

(b) In green we see the maxima of $h_p(u)$ as a function of p, which is equal to the convex conjugate $J^*(p)$.

Figure 2.9.: Construction of the convex conjugate $J^*(p)$ of a function $J(u)$

In Figure 2.9 we see how the convex conjugate can be constructed. We especially see that the distance to the y-axis of $\min_u J(u)$ as well as of $\min_p J^*(p)$ is equal except for the sign. Furthermore, the negative y-value of the minimum of $J^*(p)$ equals the value, where $J(u)$ intersects the y-axis.

Note that the biconjugate J^{**} is also the *closed convex hull* of J (cf. Figure 2.10), i.e. the largest lower semi-continuous convex functional with $J^{**} \leq J$. Furthermore, the Fenchel-Moreau theorem (cf. Ioffe and Tihomirov (1979, Section 3.3)) states that for proper functionals J holds $J = J^{**}$ if and only if J is convex and lower semi-continuous.

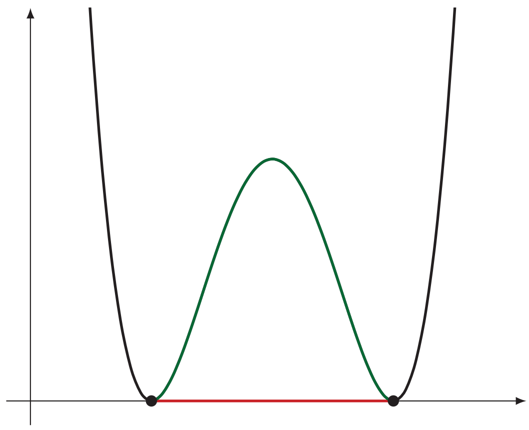


Figure 2.10.: A nonconvex function can be seen in black and green, whereas its convex hull consists of the black and red parts

The following theorem, which can be found in Rockafellar (1970, Theorem 31.1, p. 327), is a very important result in convex analysis, since it allows the reformulation of a minimization problem into an equivalent dual formulation, which might in some cases be easier to handle than the original primal problem.

Theorem 2.12 (Fenchel's Duality Theorem)**.**
Let $J_1 : \mathcal{V} \longrightarrow \overline{\mathbb{R}}$ and $J_2 : \mathcal{W} \subset \mathcal{U} \longrightarrow \overline{\mathbb{R}}$ be proper, lower semi-continuous and convex functionals for reflexive Banach spaces $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$, such that $\mathrm{dom} J_1 \cap \mathrm{dom} J_2 \neq \emptyset$. Furthermore, let $\mathcal{A} : \mathcal{U} \longrightarrow \mathcal{V}$ be a continuous linear operator. Moreover, let be $\alpha \in \mathbb{R}$ and $w \in \mathcal{V}$.
Then the following equality holds:

$$\inf_{z \in \mathcal{W}} \left\{ \frac{1}{\alpha} J_1(\mathcal{A}z - w) + J_2(z) \right\} = \sup_{p \in \mathcal{V}^*} \left\{ -\frac{1}{\alpha} J_1^*(\alpha p) - J_2^*(-\mathcal{A}^* p) - \langle p, w \rangle_{\mathcal{V}} \right\} . \tag{2.9}$$

Proof.
In this thesis we only want to sketch the proof like it was done in BENNING (2011). For a detailed proof see EKELAND AND TÉMAM (1999, Chapter 3, Sections 1-4).
Starting with

$$\inf_{z\in\mathcal{W}} \left\{ \frac{1}{\alpha} J_1(\mathcal{A}z - w) + J_2(z) \right\} ,$$

we can rewrite this expression in terms of a Lagrange functional to

$$\inf_{z\in\mathcal{W}, v\in\mathcal{V}} \sup_{p\in\mathcal{V}^*} \left\{ \frac{1}{\alpha} J_1(v) + J_2(z) + \langle p, \mathcal{A}z - w - v\rangle_{\mathcal{V}} \right\} .$$

Due to the assumptions made above, we can exchange infimum and supremum and obtain

$$\sup_{p\in\mathcal{V}^*} \left\{ \underbrace{\inf_{v\in\mathcal{V}} \left\{ \frac{1}{\alpha} \left(J_1(v) - \langle \alpha p, v\rangle_{\mathcal{V}} \right) \right\}}_{-\frac{1}{\alpha}J_1^*(\alpha p)} + \underbrace{\inf_{z\in\mathcal{W}} \left\{ J_2(z) - \langle -\mathcal{A}^* p, z\rangle_{\mathcal{W}} \right\}}_{-J_2^*(-\mathcal{A}^* p)} - \langle p, w\rangle_{\mathcal{V}} \right\} .$$

Thus (2.9) holds. □

For further illustration of the concept of duality, we cite an interesting example, which can be found in MÖLLER (2012, Section 2.1, Example 2.1.6).

Example 2.4.
Consider $\mathcal{U} = \mathbb{R}^n$, $\mathcal{V} = \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, $z \in \mathcal{U}$ and data $w \in \mathcal{V}$. Let the characteristic function of the corresponding set be denoted by $\chi_{\|\cdot\|\leq\alpha}$. In Table 2.1 we display some examples of primal minimization problems with their corresponding dual problems.

Primal Problem	Dual Problem
$\frac{1}{2}\|Az - w\|_2^2 + \alpha\|z\|_1$	$\frac{1}{2}\|p - w\|_2^2 + \chi_{\|\cdot\|_\infty\leq\alpha}(A^T p)$
$\frac{1}{2}\|Az - w\|_2^2 + \chi_{\|\cdot\|_1\leq\alpha}(z)$	$\frac{1}{2}\|p - w\|_2^2 + \alpha\|A^T p\|_\infty$
$\frac{1}{2}\|Az - w\|_2^2 + \alpha\|z\|_\infty$	$\frac{1}{2}\|p - w\|_2^2 + \chi_{\|\cdot\|_1\leq\alpha}(A^T p)$
$\frac{1}{2}\|Az - w\|_2^2 + \chi_{\|\cdot\|_\infty\leq\alpha}(z)$	$\frac{1}{2}\|p - w\|_2^2 + \alpha\|A^T p\|_1$
$\frac{1}{2}\|z - w\|_2^2 + \alpha\|Az\|_1$	$\frac{1}{2}\|A^T p - w\|_2^2 + \chi_{\|\cdot\|_\infty\leq\alpha}(p)$
$\chi_{Az=w}(z) + \mathcal{R}(z)$	$-\langle w, p\rangle + \mathcal{R}^*(A^T p)$

Table 2.1.: Examples of minimization problems and their corresponding duals

2.4. Introduction to Sparsity

In variational methods such as the one proposed in (1.4), ℓ^p-norms can be used as regularization functionals in order to favor solutions with certain properties.
There are many applications, in which the solution may be expressed by only a few basis vectors. The advantage of this property is that their solution can be recovered with only few information. On these grounds, algorithms may be accelerated, data can be highly compressed and thus less storage is needed. For these reasons, we motivate the usage of certain ℓ^p-norms as regularization functionals, which promote solutions with only few nonzero entries, i.e. *sparse* solutions. The general theory behind this idea is called *compressed sensing* or *compressive sensing*, see for instance Donoho (2006). Recovering sparse solutions has been a widely researched topic over the last years and found many real life applications, such as the image compression format JPEG-2000 and the audio compression format MP3, cf. Mallat (2008). These formats use the fact that images or audio signals can often be well approximated by using sparse representations of a suitable dictionary.
In this section we introduce sparsity-promoting functionals. We start by presenting the so-called ℓ^0-"norm", which on the one hand enforces sparsity, but on the other hand it is highly nonconvex. In order to avoid problems with non-convexity, we moreover propose the convex relaxation of the ℓ^0-"norm", namely the ℓ^1-norm. Finally we end this section by introducing regularization with Radon measures, which results from considering the ℓ^1-regularization in a continuous manner.

2.4.1. ℓ^0-Regularization

The basic idea behind promoting sparsity in reconstruction problems is to consider the so-called ℓ^0*-norm* of a vector $z \in \mathbb{R}^M$, i.e.

$$\|z\|_0 := \sum_{i=1}^{M} |z_i|^0 \tag{2.10}$$

with the convention $0^0 := 0$, which is not a norm in the classical sense and not even a quasi-norm. Nevertheless, since (2.10) can be interpreted as the limit of the ℓ^p-norm for $p \to 0$, it has been named "norm" anyway. The ℓ^0-norm is a measure for sparsity, since it counts the number of nonzero entries of a signal z. Thus it promotes solutions with only few nonzero entries, in case that it is used as a penalty term. In this context we consider the following property of a signal:

Definition 2.32 (s-Sparsity)**.**
A signal z is called *s-sparse* if it holds

$$\|z\|_0 = s \;.$$

Ideally we would like to consider problems like

$$\min_z \|z\|_0 \qquad \text{s. t.} \qquad Az = w \tag{2.11}$$

or formulated in an unconstrained way as

$$\min_z \frac{1}{2} \|Az - w\|_2^2 + \alpha \|z\|_0 \;. \tag{2.12}$$

By case differentiation we can see that for problem (2.12) with A being the identity the optimal solution is the *hard shrinkage* of w with threshold $\sqrt{2\alpha}$, i.e. it can be characterized componentwise as

$$\hat{z}_i = \begin{cases} w_i, & \text{if } |w_i| \geq \sqrt{2\alpha} \;, \\ 0, & \text{else.} \end{cases}$$

An illustration of the hard shrinkage operator can be seen in Figure 2.11.

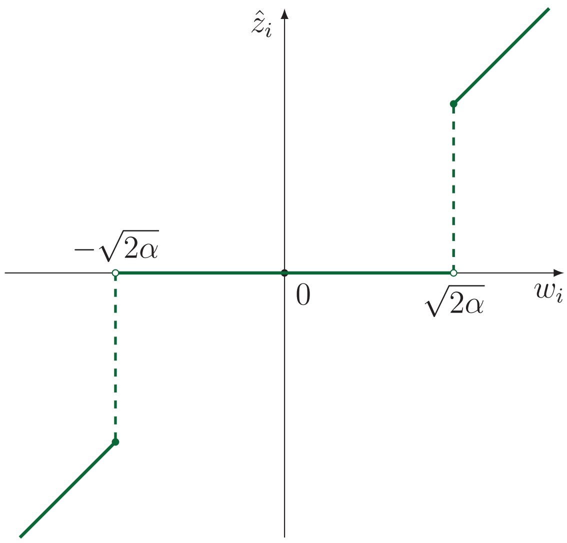


Figure 2.11.: Visualization of the hard shrinkage operator as a function of w_i with fixed regularization parameter α

Unfortunately it turns out that the ℓ^0-norm is extremely non-convex and hence the computation of global minima for an arbitrary matrix A is very difficult. In fact the solution of problems (2.11) and (2.12) is even NP-hard (cf. Natarajan (1995) and Davis et al. (1997)) and thus this approach is not really applicable in real life.

2.4.2. ℓ^1-Regularization

In the last subsection we have seen that the minimization of problems (2.11) and (2.12) raises difficulties due to the nonconvexity of the ℓ^0-norm.
Basically there are two ways to overcome this problem, either one has to use *greedy algorithms* (cf. DONOHO et al. 2006, GILBERT AND TROPP 2007, KUNIS AND RAUHUT 2008, NEEDELL AND VERSHYNIN 2009, RAUHUT 2008, TROPP 2004) or *convex relaxation* (cf. EKELAND AND TÉMAM 1999) of (2.11), which results in the ℓ^1-minimization problem

$$\min_z \|z\|_1 \qquad \text{s. t.} \qquad Az = w \; . \tag{2.13}$$

We concentrate on convex relaxation and develop some regularizations, which are based upon the fact that the ℓ^1-norm is the convex relaxation of the ℓ^0-norm.
There can be found several sufficient conditions for a solution of (2.11) being the unique minimizer of (2.13), such as the nullspace property (NSP) (cf. COHEN et al. 2009, DONOHO AND HUO 2001), the restricted isometry property (RIP) (cf. CANDÈS AND TAO 2005), the mutual incoherence property (MIP) (cf. DONOHO AND HUO 2001), the exact recovery condition (ERC) (cf. TROPP 2004) and the property based on the characteristic $\gamma_k(A)$ (cf. JUDITSKY AND NEMIROVSKI 2011). We will not go into the details of these properties for exact recovery and refer the reader to FORNASIER (2010) for further information.

Comparing the ℓ^1-norm with the ℓ^2-norm of a signal, one can discover signals with similar values of the ℓ^2-norm, even if the structure of the signals differs considerably. For instance a sparse signal can have the same ℓ^2-norm as a dense signal. However, the ℓ^1-norm of a sparse and a dense signal differs significantly. This provides another indication for the usage of the ℓ^1-norm in case sparse signals can be expected.
A widely-used variational scheme in compressed sensing consists of an ℓ^2-data fidelity and an ℓ^1-regularization term, i.e.

$$\min_z \frac{1}{2} \|Az - w\|_2^2 + \alpha \|z\|_1 \; , \tag{2.14}$$

which is an unconstrained version of the constrained problem (2.13). Let us again consider the case where A is the identity. Fortunately in this case the solution of (2.14) can be easily computed componentwise via the *soft shrinkage* operator, i.e.

$$\hat{z}_i = \text{shrink}(w_i, \alpha) := \text{sign}\,(w_i) \max(|w_i| - \alpha, 0) \; , \tag{2.15}$$

which can be seen in Figure 2.12.

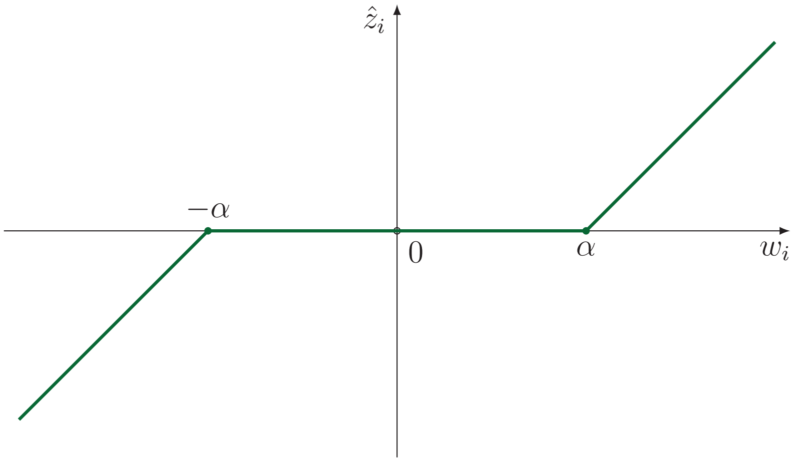


Figure 2.12.: Visualization of the soft shrinkage operator as a function of w_i with fixed regularization parameter α

2.4.3. Regularization with Radon Measures

In this subsection we introduce another important type of regularization method. Recent research makes significant progress in analyzing the asymptotics of certain regularization energies. In so doing, Radon measures (as proposed in Subsection 2.2.6) prove to be a useful and versatile tool.

In order to contemplate the ℓ^1-regularization in a continuous manner, one might first approach this problem by using the L^1-norm as a continuous counterpart to the ℓ^1-norm, i.e.

$$\mathcal{R}(u) = \|u\|_{L^1(\Omega)} \ ,$$

for a function $u\colon\Omega \longrightarrow \mathbb{R}$ with $\Omega \subset \mathbb{R}^d$ and $d \in \mathbb{N}$. One might also want to consider the L^1-norm not only of a signal itself, but also of a certain operator applied to the signal such as

$$\mathcal{R}(u) = \|\mathcal{A}u\|_{L^1(\Theta)} \ ,$$

with $\mathcal{A}\colon\mathcal{U}(\Omega) \longrightarrow L^1(\Theta)$ for a suitable Banach space $\mathcal{U}(\Omega)$ and suitable sets Ω and Θ. However, in Subsection 2.2.5 we observed that the space $L^1(\Omega)$ might be too restrictive for certain applications. Choosing $\mathcal{A} = \nabla$ for instance permits only solutions $u \in W^{1,1}(\Omega)$, which has not proved satisfactory, since $W^{1,1}(\Omega)$ does not contain functions with discontinuities.

In order to overcome this difficulties Bredies and Pikkarainen (2013) studied inverse problems in the space of finite Radon measures. They consider linear inverse problems of the type

$$\mathcal{K}\mu = \hat{w} \,, \tag{2.16}$$

where the exact data $\hat{w}$ is contained in a Hilbert space $\mathcal{H}$ and μ is an element of $\mathcal{M}(\Omega, \mathbb{R}^m)$, which denotes the space of finite vector-valued Radon measures. $\mathcal{K}$ is supposed to map continuously from $\mathcal{M}(\Omega, \mathbb{R}^m)$ into $\mathcal{H}$. Note that the set Ω is allowed to be a continuum. Due to the representation proposed in Theorem 2.10, it turns out to be useful to consider linear mappings, which are adjoints of linear and continuous maps $\mathcal{A}\colon \mathcal{H} \longrightarrow C_0(\Omega, \mathbb{R}^m)$, i.e. $\mathcal{K} = \mathcal{A}^*$. The usage of finite Radon measures proved to be a decisive advantage, since $\mathcal{M}(\Omega, \mathbb{R}^m)$ allows for finite linear combinations of delta peaks. With respect to the possibly continuous index set Ω, this fact can be interpreted as continuous counterpart to discrete sparsity. Therefore, Bredies and Pikkarainen analyze the variational scheme

$$\min_{\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)} \frac{1}{2} \left\| \mathcal{K}\mu - w \right\|_{\mathcal{H}}^2 + \alpha \left\| \mu \right\|_{\mathcal{M}(\Omega, \mathbb{R}^m)} \tag{2.17}$$

with $\mathcal{K} = \mathcal{A}^*$ for noisy data w.
There exists a minimizer for this variational model as we can see in the following theorem, which we cite from Bredies and Pikkarainen (2013, Proposition 3.1).

Theorem 2.13 (Existence of a Minimizer)**.**
Let $\mathcal{A} : \mathcal{H} \longrightarrow C_0(\Omega, \mathbb{R}^m)$ be linear and continuous and let $w \in \mathcal{H}$ and $\alpha > 0$ hold. Moreover, let be $\mathcal{K} = \mathcal{A}^*$. Then there exists a solution $\mu^* \in \mathcal{M}(\Omega, \mathbb{R}^m)$ for the minimization problem (2.17). In case that $\mathcal{K}$ is injective, μ^* is the unique solution of the problem.

Due to compactness we omit the proof, which can be found in the above-mentioned paper.
On the basis of this entire approach we will analyze the asymptotics of spatial sparsity priors in Chapter 7.

3

Mixed Norms and Their Ability to Promote Structured Sparsity

In this chapter we introduce mixed $\ell^{p,q}$-norms as a generalization of usual ℓ^p-norms (cf. Subsection 2.2.3). By using mixed norms as regularization functionals for variational methods, where the unknown is considered to be a matrix, various structures of the minimizer can be promoted by considering different values for p and q. Therefore, using mixed norms as regularization functional for the solution of inverse problems can be a useful tool.

In the first section we give a general introduction to the topic of mixed norms and emphasize their relationship to operator norms, which are already well established in literature. We not only state their similarities, but also explain some differences between these two matrix norms. Afterwards, we demonstrate how mixed norms can be used to promote sparsity for matrices. Certainly, promoting usual sparsity by choosing p and q equal to one is an option. However, the versatility of mixed norms rather emerges from choosing p unequal to q. In particular, choosing either p or q to be equal to one and its corresponding other as larger than one turns out to be beneficial, since those regularizations promote differently structured sparsities. As an example thereof we shortly explain the idea of *joint sparsity*. Subsequently, we introduce *local sparsity* via the $\ell^{1,\infty}$-norm, which shall be further analyzed in the course of this thesis.

3.1. Introduction to Mixed $\ell^{p,q}$-Norms

The concept of ℓ^p-norms as proposed in Subsection 2.2.3 can be generalized to $\ell^{p,q}$-norms, where we shall consider finite and infinite matrices (i.e. two dimensional sequences) instead of one dimensional sequences. There are mainly two different definitions of mixed norms. We shall give a brief introduction of both of them and show some relations between these two definitions in conclusion.

First we go along the lines of Rakbud and Ong (2011, Section 1) in order to introduce operators between $\ell^p(\Omega)$ and $\ell^q(\Omega)$. This will yield the first definition of mixed $\ell^{p,q}$-norms based on the operator norm. Afterwards, we will introduce mixed norms as a composition of ℓ^p- and ℓ^q-norms, which shall be used throughout this thesis.

Definition 3.1 (Infinite Matrix)**.**
Let Ω be a field. We call A an *infinite matrix* if $A := \big\{\{a_{mn}\}_{m\in\mathbb{N}}\big\}_{n\in\mathbb{N}}$ is a sequence of sequences with $a_{mn} \in \Omega$ for all $m, n \in \mathbb{N}$, i.e.

$$A = \{a_{mn}\}_{m,n\in\mathbb{N}} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots \\ a_{21} & a_{22} & a_{23} & \dots \\ a_{31} & a_{32} & a_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

In the course of this thesis we work with sequences in $\ell^q(\Omega)$ of sequences in $\ell^p(\Omega)$. Thus we concentrate on these sequence spaces and their resulting infinite matrices. In order to do so, we first present some further definitions.

Definition 3.2 (Linear Operator between $\ell^p(\Omega)$ and $\ell^q(\Omega)$)**.**
Let be $p, q \in [1, \infty)$ and let $A = \{a_{mn}\}_{m,n\in\mathbb{N}}$ be an infinite matrix. A defines a linear operator from $\ell^p(\Omega)$ to $\ell^q(\Omega)$ if

$$\sum_{n=1}^{\infty} a_{mn} x_n < \infty \quad \forall\, m \in \mathbb{N}$$

holds for all $\{x_n\}_{n\in\mathbb{N}} \in \ell^p(\Omega)$ and we have

$$Ax = \left\{\sum_{n=1}^{\infty} a_{mn} x_n\right\}_{m\in\mathbb{N}} \in \ell^q(\Omega) \qquad \text{i.e.} \qquad \sum_{m=1}^{\infty} \left|\sum_{n=1}^{\infty} a_{mn} x_n\right|^q < \infty\,.$$

Then we call $A\colon \ell^p(\Omega) \longrightarrow \ell^q(\Omega)$ with $x \mapsto Ax$ the *linear operator defined by A*.
Note that the cases $p = \infty$ and $q = \infty$ can be obtained by replacing the corresponding norm by the supremum.

Definition 3.3 (Space of Infinite Matrices $\ell^{\star\, p,q}(\Omega)$)**.**
Let Ω be a field and let $\ell^p(\Omega)$ and $\ell^q(\Omega)$ be two sequence spaces as defined in Definition 2.17 for $p, q > 1$. We define $\ell^{\star\, p,q}(\Omega)$ as the vector space of all infinite matrices defining continuous linear operators from $\ell^p(\Omega)$ into $\ell^q(\Omega)$. We see that $\ell^{\star\, p,q}(\Omega)$ is naturally a normed vector space with the operator norm

$$\|A\|^{\star}_{p,q} := \sup_{x \in \ell^p(\Omega)} \frac{\|Ax\|_q}{\|x\|_p} = \sup_{\|x\|_p = 1} \|Ax\|_q \ . \tag{3.1}$$

Since $\ell^p(\Omega)$ and $\ell^q(\Omega)$ are Banach spaces, $\ell^{\star\, p,q}(\Omega)$ is also a Banach space.

There exist several different definitions of $\ell^{p,q}$-norms in literature (cf. for instance Kowalski 2009, Tropp 2006a), especially in combination with finite matrices. For our purposes we consider $\ell^{p,q}(\Omega)$ with another type of mixed norms, mainly the one defined in Kowalski (2009, Definition 1), which can also be found in Samarah et al. (2005, Definition 2). It is consistent with the composition of two ℓ^p-norms, i.e. $\|\cdot\|_{p,q} = \|\cdot\|_q \circ \|\cdot\|_p = \left\| \|\cdot\|_p \right\|_q$, cf. also Kowalski and Torrésani (2009).
Before going into further details, it might be interesting to know that there exists a general theory about mixed norms as composition of several L^p-norms, ℓ^p-norms respectively. The interested reader might gain further information on this topic in Benedek and Panzone (1961), where also many properties of mixed norms are proven.
However, for our purposes we restrict this introduction to the case of composing an ℓ^q-norm with an ℓ^p-norm.

Definition 3.4 (Mixed $\ell^{p,q}$-Norms)**.**
Let the infinite matrix $\mathbf{w}$ with entries $w_{mn} \in \mathbb{R}^+ \backslash \{0\}$ for all $m, n \in \mathbb{N}$ be a weight matrix. Furthermore, let A be an infinite matrix and let be $p \geq 1$ and $q \geq 1$.
Then the *mixed $\ell^{p,q}$-norm* is defined as

$$\|A\|_{\mathbf{w};p,q} := \left(\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} w_{mn} \, |a_{mn}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \ . \tag{3.2}$$

For the sake of simplicity, we define $\|\cdot\|_{p,q} := \|\cdot\|_{\mathbb{1}^2;p,q}$, where $\mathbb{1}^2$ denotes the corresponding infinite matrix of ones.
The cases where we have $p = \infty$ and $q = \infty$ may again be obtained by replacing the corresponding norm by the supremum. Note that the above mentioned definition holds for finite matrices as well.

Lemma 3.1.
The mixed $\ell^{p,q}$-norm as defined in Definition 3.4 is truly a norm.

This property can be seen rather easily, since the mixed $\ell^{p,q}$-norm is obviously a composition of ℓ^p- and ℓ^q-norms.

Proof.
Let A and B be two infinite matrices and let be $\alpha \in \mathbb{C}$. Then we obtain the following properties of a norm:

1. Definiteness:
$$\|A\|_{p,q} = 0 \quad \Rightarrow \quad A \equiv 0$$

2. Absolute homogeneity:
$$\|\alpha A\|_{p,q} = \left\| \|\alpha A\|_p \right\|_q = \left\| |\alpha| \, \|A\|_p \right\|_q = |\alpha| \left\| \|A\|_p \right\|_q = |\alpha| \, \|A\|_{p,q}$$

3. Triangle inequality:
$$\begin{aligned} \|A + B\|_{p,q} &= \left\| \|A + B\|_p \right\|_q \leq \left\| \|A\|_p + \|B\|_p \right\|_q \\ &\leq \left\| \|A\|_p \right\|_q + \left\| \|B\|_p \right\|_q = \|A\|_{p,q} + \|B\|_{p,q} \end{aligned}$$

□

Now we can define another space of infinite matrices using the norm (3.2), cf. Samarah et al. (2005, Definition 2).

Definition 3.5 (Weighted mixed-norm space $\ell^{p,q}_{\mathbf{w}}(\Omega)$)**.**
Let the infinite matrix $\mathbf{w}$ with $w_{mn} \in \mathbb{R}^+ \backslash \{0\}$ for all $m, n \in \mathbb{N}$ be a weight matrix.
The space $\ell^{p,q}_{\mathbf{w}}(\Omega)$ with $0 < p, q \leq \infty$ consists of all infinite matrices $A = \{a_{mn}\}$, for which the $\ell^{p,q}$-norm as defined in (3.2) is finite.
Again we simplify the notation if $\mathbf{w} = \mathbb{1}^2$, i.e. we define $\ell^{p,q}(\Omega) := \ell^{p,q}_{\mathbb{1}^2}(\Omega)$.

Note that from now on we refer to Definition 3.4 when we use the terms *$\ell^{p,q}$-norm* or *mixed norms* if not stated otherwise.

Remark 3.1.
For finite matrices, the norm $\|\cdot\|^{\star}_{p,q}$ given by (3.1) and the norm $\|\cdot\|_{p,q}$ defined by (3.2) are equivalent.

Similarly to the definition of $\ell^{p,q}(\Omega)$, one can define mixed Lebesgue spaces on the basis of usual Lebesgue spaces, cf. Subsection 2.2.2.

Definition 3.6 (Mixed Lebesgue Spaces $L^{p,q}(\Omega)$)**.**
Let be p, $q \geq 1$ and let Ω and $\widetilde{\Omega}$ be two fields. The space $L^{p,q}(\Omega \times \widetilde{\Omega})$ is defined as the set of all measurable functions $u\colon \Omega \times \widetilde{\Omega} \longrightarrow \mathbb{R}$, for which the $L^{p,q}$-norm

$$\|u\|_{L^{p,q}(\Omega\times\widetilde{\Omega})} := \left(\int_{\widetilde{\Omega}} \left(\int_{\Omega} |u(x,y)|^p \,\mathrm{d}x\right)^{\frac{q}{p}} \mathrm{d}y\right)^{\frac{1}{q}}$$

is finite. Note that the $L^{p,q}$-norm is the composition of the L^p- and the L^q-norm, i.e. $\|u\|_{L^{p,q}(\Omega\times\widetilde{\Omega})} = \|\cdot\|_{L^q(\widetilde{\Omega})} \circ \|\cdot\|_{L^p(\Omega)} = \left\| \|\cdot\|_{L^p(\Omega)} \right\|_{L^q(\widetilde{\Omega})}$.

However, in this thesis we will predominantly use their discrete counterparts. In the following we therefore present some properties of $\ell^{p,q}$-norms.
Some special types of $\ell^{p,q}$-norms can be represented by operator norms (cf. Definition 3.3). As an example we consider the $\ell^{1,\infty}$-norm, which will be of particular importance in the course of this thesis. Further representations of mixed norms as operator norms can be found in Table 3.1. However, not every mixed $\ell^{p,q}$-norm has a representation as operator norm as we shall see in Example 3.2.

Example 3.1 (Operator Norm Representation of the Mixed $\ell^{1,\infty}$-Norm)**.**
Let A be an infinite matrix. Then we have

$$\begin{aligned}\|A\|^{\star}_{\infty,\infty} &= \sup_{\|x\|_\infty=1} \|Ax\|_\infty = \sup_{\|x\|_\infty=1} \sup_{i\in\mathbb{N}} \left|\sum_{j=1}^{\infty} a_{ij}x_j\right| = \sup_{i\in\mathbb{N}} \sup_{\|x\|_\infty=1} \left|\sum_{j=1}^{\infty} a_{ij}x_j\right| \\ &= \sup_{i\in\mathbb{N}} \sum_{j=1}^{\infty} |a_{ij}| = \|A\|_{1,\infty} \quad ,\end{aligned}$$

which uses the fact that the sum over j becomes maximal in case that $x_j = \operatorname{sgn}(a_{ij})$ holds.

Thus we see that a matrix with finite $\ell^{1,\infty}$-norm induces an operator between $\ell^\infty(\Omega)$ and $\ell^\infty(\Omega)$.

Mixed Norm	Operator Norm
$\|A\|_{1,\infty}$	$\|A\|^{\star}_{\infty,\infty}$
$\|A\|_{\infty,\infty}$	$\|A\|^{\star}_{1,\infty}$
$\|A\|_{1,1}$	$\|A\|^{\star}_{\infty,1}$
$\|A\|_{\infty,1}$	$\|A\|^{\star}_{1,1}$

Table 3.1.: Equalities of the mixed norms (3.1) and (3.2).

Example 3.2 (Spectral Norm & Frobenius Norm).
Let $A \in \mathbb{R}^{M \times N}$ be a finite matrix.
Then the spectral norm $\|A\|_{2,2}^{\star}$, which is the largest singular value of A, and the Frobenius norm $\|A\|_F := \|A\|_{2,2}$ are equivalent, i.e.

$$\|A\|_{2,2}^{\star} \leq \|A\|_{2,2} \leq \sqrt{\min\{M,N\}}\, \|A\|_{2,2}^{\star} \ .$$

These inequalities are consequences of the representation of the Frobenius norm via its singular value decomposition, i.e.

$$\|A\|_{2,2}^{\star\,2} = \sigma_{\max}^2 \leq \underbrace{\sigma_1^2 + \ldots + \sigma_r^2}_{\|A\|_{2,2}^2} \leq r\sigma_{\max}^2 = r\, \|A\|_{2,2}^{\star\,2} \ ,$$

where $\sigma_1, \ldots, \sigma_r$ are the singular values of A with $r \leq \min\{M,N\}$ and $\sigma_{\max}$ their maximum. Obviously for infinite matrices this is no longer true, since in this case r can go to infinity.
Moreover, every operator norm has to fulfill

$$\|I_{M \times N}\|_{p,q}^{\star} = 1 \ .$$

However, for the Frobenius norm we obtain

$$\|I_{M \times N}\|_{2,2} = \sqrt{\min\{M,N\}} > 1 \ .$$

Thus we have found a mixed $\ell^{p,q}$-norm, which can not be represented by an operator norm.

3.2. Sparsity and Mixed Norms

Promoting "usual" sparse solutions as proposed in Section 2.4 is not the only kind of prior information available in practice. Strong recent directions of research are related to unknowns being matrices i.e. inverse problems like (1.3). In this context mixed norms become a useful and versatile tool. One may consider minimization problems like

$$\min_Z \|Z\|_{p,q} \qquad \text{s. t.} \qquad AZ = W$$

for specific $p > 0$ and $q > 0$, where $A \in \mathbb{R}^{L \times M}$, $Z \in \mathbb{R}^{M \times T}$ and $W \in \mathbb{R}^{L \times T}$. Depending on the choice of p and q these minimization problems may be used to incorporate structured sparsity.

Example 3.3 (Usual Sparsity on Matrices)**.**
Let for instance $p = q = 1$ holds. By considering the optimization problem

$$\min_{Z} \|Z\|_{1,1} \qquad \text{s. t.} \qquad AZ = W \ ,$$

we promote usual sparsity as proposed in Subsection 2.4.2 on the vectorized matrix $\operatorname{vec}(Z)$, since we obviously have

$$\|Z\|_{1,1} = \|\operatorname{vec}(Z)\|_1 \ .$$

Thus we see that we should consider either $p \neq 1$ or $q \neq 1$ in order to incorporate more prior information concerning structured sparsity.

3.2.1. Joint Sparsity via $\ell^{p,1}$-Regularizations

By presenting sparsity-related mixed norms, we now shortly introduce the idea of *joint sparsity*, which has been studied for instance by Fornasier and Rauhut (2008) and Teschke and Ramlau (2007). Afterwards, we continue with the motivation of *local sparsity*, which shall be examined in detail during the course of this thesis.

Multi-channel signals as proposed in (1.3) frequently appear in real life applications. These signals often permit the usage of sparse frame expansions for each channel individually. Furthermore, the different channels may also possess common sparsity patterns. By expanding this idea, Baron et al. (2005), Gilbert et al. (2006) and Tropp (2006a) introduced new sparsity measures, which promote such coupling of the non-vanishing components through different channels. Usually these measures are weighted ℓ^1-norms of ℓ^p-norms of the channels with $p > 1$, i.e.

$$\|Z\|_{\mathbf{w};p,1} := \sum_{m=1}^{\infty} w_m \left(\sum_{n=1}^{\infty} |z_{mn}|^p \right)^{\frac{1}{p}} \ . \tag{3.3}$$

Note that in this case the weights only belong to the ℓ^1-norm. The cases, which are of particular interest, are given by $p = 2$ or $p = \infty$. This is reasonable, since (3.3) reduces to the usual weighted ℓ^1-norm for $p = 1$ and on the other hand for $2 < p < \infty$ the expression becomes too complicated. Minimizing (3.3) favors that in the "interchannel" vector z_m all entries may become significant in case that at least one of the components $|z_{mn}|$ is large.

As an example for a real life application, we can consider color image reconstruction, for instance art restoration. Color images are indeed non-trivial multivariate and multi-channel signals. Motivated by the idea that discontinuities may appear in all channels at the same locations, one may want to search for a solution with small total variation in each channel and moreover demand from the channels to possess a common subgradient and thus a common edge set (cf. Moeller et al. 2014). In our context this could also be interpreted as joint gradient sparsity.

3.2.2. Local Sparsity via $\ell^{1,q}$-Regularizations

In imaging sciences inverse problems frequently appear with some spatial dimensions and at least one additional dimension such as time or spectral information. These applications motivate the usage of *local sparsity*, which promotes that for instance every pixel should be a sparse representation with respect to a certain dictionary including the local behaviour in the additional dimension. Considering such a known dictionary $B \in \mathbb{R}^{T \times N}$ and an unknown coefficient matrix $U \in \mathbb{R}^{M \times N}$ the desired matrix $Z \in \mathbb{R}^{M \times T}$ can be written as

$$Z = UB^T \ .$$

Inserting this into the matrix equation (1.3) yields the new inverse problem

$$AUB^T = W \ , \tag{3.4}$$

with $A \in \mathbb{R}^{L \times M}$ and $W \in \mathbb{R}^{L \times T}$.

Thus the challenge is to incorporate usual sparsity on every row vector $u_{i\cdot}$ of the unknown coefficient matrix U, which means ideally we would like to minimize

$$\|U\|_{0,\infty} = \max_i \|u_{i\cdot}\|_0$$

subject to (3.4). Considering the fact that the ℓ^1-norm is the convex relaxation of the ℓ^0-norm (cf. Section 2.4), a natural relaxation is to consider the minimization problem

$$\min_U \|U\|_{1,\infty} \qquad \text{s. t.} \qquad AUB^T = W \ . \tag{3.5}$$

We shall analyze problem (3.5) and other $\ell^{1,\infty}$-related formulations extensively in Chapter 4.

Instead of minimizing the $\ell^{1,\infty}$-norm one may also consider minimizing $\ell^{1,q}$-norms with $1 < q < \infty$, which can be seen as a relaxation of the $\ell^{1,\infty}$-norm. However, choosing $q > 2$ rather complicates the problem. Thus the remaining case, which may be considered as an alternative to $\ell^{1,\infty}$-minimization, is the one, where $q = 2$ holds. This case has indeed been studied extensively in literature. Kowalski and Torrésani (2009) for instance compare the $\ell^{1,2}$-norm with the $\ell^{2,1}$-norm. Furthermore, the solution of the $\ell^{1,2}$-regularized problem is deduced in Kowalski (2009) as a part of a more general result.

4 Analysis of $\ell^{1,\infty}$- Related Formulations

In Subsection 3.2.2 of the previous chapter we have motivated the usage of the $\ell^{1,\infty}$-norm as regularization functional. This chapter engages further in this topic and largely consists of the theoretical part of our preprint Heins et al. (2014).

Unfortunately the minimization of the $\ell^{1,\infty}$-norm

$$\|U\|_{1,\infty} := \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} |u_{ij}|$$

over the unknown matrix $U \in \mathbb{R}^{M\times N}$ in combination with the data constraint (3.4) is not quite trivial. While attempting to examine this regularization functional, we perceive that the analysis and the computational realization of the $\ell^{1,\infty}$-minimization problem proves to be rather difficult. Therefore, we derive and analyze different formulations of this functional in order to make the analysis and especially the computation of the $\ell^{1,\infty}$-regularization better realizable. Moreover, we investigate some basic properties. In this new setting we additionally analyze what happens in the special case, where the regularization parameter tends to infinity and obtain a very simple representation of the solution. Furthermore, we shall investigate the potential to exactly reconstruct locally one-sparse signals by convex optimization techniques.

This analysis concerning exact recovery, cf. Section 4.2, is essential for the idea that it may be beneficial to consider a combination of minimizing the $\ell^{1,\infty}$-norm with classical sparsity. For $A \in \mathbb{R}^{L\times M}$, $U \in \mathbb{R}^{M\times N}$, $B \in \mathbb{R}^{T\times N}$, $W \in \mathbb{R}^{L\times T}$ and $\alpha, \beta \in \mathbb{R}^+$ we therefore also investigate the more general problem

$$\min_{U \in \mathbb{R}^{M \times N}} \alpha \|U\|_{1,\infty} + \beta \|U\|_{1,1} \quad \text{s.t.} \quad AUB^T = W \ , \tag{4.1}$$

which neither is a restriction for the further analysis nor complicates the computational realization. Besides the constrained model (4.1), we shall also investigate the unconstrained variational model

$$\min_{U \in \mathbb{R}^{M \times N}} \frac{1}{2} \|AUB^T - W\|_F^2 + \alpha \|U\|_{1,\infty} + \beta \|U\|_{1,1} \ , \tag{4.2}$$

which is suited to deal with noisy data as well. Here we use the Frobenius norm for the data fidelity term, i.e. the $\ell^{2,2}$-norm. Due to the fact that in nearly every practical application one only looks for positive combinations of dictionary elements, we shall put a particular emphasis on the case of an *additional nonnegativity constraint* on U in (4.1), respectively (4.2).

After substantially analyzing the $\ell^{1,\infty}$-regularization, we examine the potential for further improvements. Since in many applications we can expect images to contain sharp edges, we discuss the possibility of including an additional regularization, which promotes sharp edges in the image, namely *total variation (TV) minimization*. We consider TV-regularization on the image itself as well as on the coefficient matrices in separate cases. The latter results from the assumption that sharp edges in the image yield sharp edges in the visualized coefficient matrices. However, adding another regularization to the main problem complicates the choice of the regularization parameters.

4.1. Basic Properties and Formulations

In this section we introduce some equivalent formulations of the main problems (4.1) and (4.2), which we use for the analysis later on. Additionally we point out some basic properties like convexity, existence and potential uniqueness. Furthermore, we propose the subdifferential and discuss a source condition. Moreover, we prove the equivalence of another reformulation, which improves the accessibility of the problem for numerical computation. Finally we investigate the limit for $\alpha \to \infty$ and observe what happens to the optimal solution in that case.

4.1.1. Problem Formulations

Since in most applications a nonnegativity constraint is reasonable, we first restrict (4.1) and (4.2) to this case. For the sake of simplicity, we define

$$G := \left\{ U \in \mathbb{R}^{M\times N} \mid u_{ij} \geq 0 \quad \forall\, i \in \{1, \dots M\},\ j \in \{1, \dots N\} \right\} .$$

Hence we have

$$\min_{U\in G} \left(\alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij} + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} \right) \quad \text{s.t.} \quad AUB^T = W \tag{4.3}$$

for the constrained problem and

$$\min_{U\in G} \left(\frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij} + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} \right) \tag{4.4}$$

for the unconstrained problem.
In order to make these problems more easily accessible, we reformulate the $\ell^{1,\infty}$-term in (4.3) and (4.4) via a linear constraint.

Theorem 4.1 (Non-Negative $\ell^{1,\infty}$-Regularization)**.**
Let $F : \mathbb{R}^{M\times N} \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a convex functional. Then

$$\min_{U\in G} \left(F(U) + \alpha \max_{i} \sum_{j=1}^{N} u_{ij} \right) \tag{4.5}$$

is equivalent to

$$\min_{U\in G,\, v\in\mathbb{R}^+} \ F(U) + v \quad \text{s.t.} \quad \alpha \sum_{j=1}^{N} u_{ij} \leq v . \tag{4.6}$$

Proof.
Introducing the constraint $v = \alpha \max_i \sum_{j=1}^{N} u_{ij}$ we obtain the equivalence

$$\begin{aligned} &\min_{U\in G} \ F(U) + \alpha \max_{i} \sum_{j=1}^{N} u_{ij} , \\ \Leftrightarrow &\min_{U\in G} \ F(U) + v \quad \text{s.t.} \quad \alpha \max_{i} \sum_{j=1}^{N} u_{ij} = v . \end{aligned}$$

Now consider the inequality-constrained problem

$$\min_{U \in G,\, v \in \mathbb{R}^+} F(U) + v \quad \text{s.t.} \quad \alpha \sum_{j=1}^{N} u_{ij} \leq v \ .$$

If $\alpha \max_i \sum_{j=1}^{N} u_{ij} < v$ holds, then (U, v) is not a minimizer, since we can choose $\bar{v} < v$, which is still feasible and reduces the objective. Thus in the optimal case the inequality constraint yields the equality constraint and the problems (4.5) and (4.6) coincide. □

Using Proposition 4.1 and defining F as the sum of the characteristic function for the data constraint and the non-negative $\ell^{1,1}$-term, we are able to reformulate problem (4.3) as

$$\min_{U \in G,\, v \in \mathbb{R}^+} \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} + v \quad \text{s.t.} \quad \alpha \sum_{j=1}^{N} u_{ij} \leq v,\ AUB^T = W \ . \tag{4.7}$$

Likewise we obtain the unconstrained problem from (4.4) as

$$\min_{U \in G,\, v \in \mathbb{R}^+} \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} + v \quad \text{s.t.} \quad \alpha \sum_{j=1}^{N} u_{ij} \leq v \tag{4.8}$$

by using the sum of the data fidelity and the non-negative $\ell^{1,1}$-term as functional F. In order to understand the potential exactness of sparse reconstructions, we focus on the analysis of (4.7) in Section 4.2, however, (4.8) is clearly more useful in practical situations when the data are not exact. Thus it builds the basis for most of the further analysis and in particular for computational investigations.

However, we first propose another formulation, which shall make the problem more easily accessible for the numerical solution. For this reformulation we show in Subsection 4.1.4 that $\max_{i=1,\dots,M} \sum_{j=1}^{N} u_{ij}(\alpha)$ depends continuously on the regularization parameter α in problem (4.5). Then we prove that $\alpha \mapsto \max_{i=1,\dots,M} \sum_{j=1}^{N} u_{ij}(\alpha)$ is monotonically decreasing and finally we analyze its limits for α going to zero and infinity. We can then show that under certain circumstances the support of the minimizers of (4.5) and

$$\min_{U \in G} F(U) \quad \text{s.t.} \quad \sum_{j=1}^{N} u_{ij} \leq \tilde{v} \tag{4.9}$$

coincide for a certain fixed $\tilde{v}$. Thus instead of regularizing with α we can now use $\tilde{v}$ as a regularization parameter.

4.1.2. Existence and Uniqueness

Let us now discuss some basic properties of $\ell^{1,\infty}$-regularized variational problems. We show that there exists a minimizer for these problems and discuss potential uniqueness.

Existence

Since the $\ell^{1,\infty}$-regularization functional is a norm it is also convex. We want to show that this holds for the non-negative formulation (4.10) as well.

Lemma 4.1 (Convexity of the Non-Negative $\ell^{1,\infty}$-Functional)**.**
Let be $U \in \mathbb{R}^{M\times N}$. The functional

$$\mathcal{R}(U) := \begin{cases} \max\limits_i \sum\limits_{j=1}^{N} u_{ij}, & \text{if } \; u_{ij} \geq 0, \\ \infty, & \text{else,} \end{cases} \tag{4.10}$$

is convex.

Proof.
Let be $U \neq V \in \mathbb{R}^{M\times N}$ and $\omega \in (0,1)$. Then we have

$$\begin{aligned} &\mathcal{R}(\omega U + (1-\omega)\,V) \\ &= \max_i \sum_{j=1}^{N} (\omega u_{ij} + (1-\omega)\,v_{ij}) + \chi_{\{U,V\in\mathbb{R}^{M\times N}\,|\,\omega u_{ij}+(1-\omega)v_{ij}\geq 0\}} \\ &\leq \omega \max_i \sum_{j=1}^{N} u_{ij} + (1-\omega)\max_i \sum_{j=1}^{N} v_{ij} + \chi_{\{U\in\mathbb{R}^{M\times N}\,|\,\omega u_{ij}\geq 0\}} + \chi_{\{V\in\mathbb{R}^{M\times N}\,|\,(1-\omega)v_{ij}\geq 0\}} \\ &= \omega\left(\max_i \sum_{j=1}^{N} u_{ij} + \chi_{\{U\in\mathbb{R}^{M\times N}\,|\,u_{ij}\geq 0\}}\right) + (1-\omega)\left(\max_i \sum_{j=1}^{N} v_{ij} + \chi_{\{V\in\mathbb{R}^{M\times N}\,|\,v_{ij}\geq 0\}}\right) \\ &= \omega\mathcal{R}(U) + (1-\omega)\,\mathcal{R}(V) \end{aligned}$$

Thus the functional (4.10) is convex. □

Lemma 4.2 (Lower Semi-Continuity)**.**
Let $F:\mathbb{R}^{M\times N} \longrightarrow \mathbb{R}\cup\{+\infty\}$ be a convex functional. Then (4.5) and (4.6) are lower semi-continuous.

Proof.
Since $F(U)$ is convex, we can conclude that (4.5) is convex by using Proposition 4.1. Due to Theorem 4.1 we deduce that (4.6) is convex as well.

Our problem is finite dimensional and hence all norms are equivalent. In addition it contains only linear inequalities. Therefore, we can deduce lower semi-continuity directly from convexity, which we already have. □

Let us now analyze the existence of minimizers of the different problems.

Theorem 4.2 (Existence of a Minimizer of the Constrained Problem).
Let there be at least one $\widetilde{U} \in G$ that satisfies $A\widetilde{U}B^T = W$. Then there exists a minimizer of the constrained problems (4.3) and (4.7).

Proof.
Since we only have linear parts, we see that

$$F(U) := \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} + C(U) \qquad \text{with} \qquad C(U) := \begin{cases} 0, & \text{if } AUB^T = W, \\ \infty, & \text{else,} \end{cases}$$

is convex. Then Proposition 4.2 leads to lower semi-continuity of (4.3) and (4.7).
We still need to show that there exists a ξ such that the sublevel set

$$\mathcal{S}_\xi = \left\{ U \in G \;\middle|\; \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} + \alpha \max_{i \in \{1,\ldots,M\}} \sum_{j=1}^{N} u_{ij} + C(U) \leq \xi \right\}$$

is compact and not empty.
With $\widetilde{U}$ we have a feasible element and we can define

$$\xi := \beta \sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{u}_{ij} + \alpha \max_{i \in \{1,\ldots,M\}} \sum_{j=1}^{N} \tilde{u}_{ij} \,.$$

Due to $\widetilde{U} \in \mathcal{S}_\xi$ we see that $\mathcal{S}_\xi$ is not empty and since for every $U \in \mathcal{S}_\xi$ holds that

$$\|U\|_{1,\infty} \leq \frac{\xi}{\alpha}\,, \quad \text{if} \quad \alpha \neq 0 \qquad \text{or} \qquad \|U\|_{1,1} \leq \frac{\xi}{\beta}\,, \quad \text{if} \quad \beta \neq 0$$

and $u_{ij} \geq 0$ for all $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N\}$ the sublevel set $\mathcal{S}_\xi$ is bounded. Our functional is finite dimensional, hence $\mathcal{S}_\xi$ is bounded in all norms. Furthermore, boundedness of $\mathcal{S}_\xi$ in combination with lower semi-continuity of (4.3) and (4.7) yields compactness of $\mathcal{S}_\xi$. Finally we obtain the existence of a minimizer of the constrained problems (4.3) and (4.7). □

Theorem 4.3 (Existence of a Minimizer of the Unconstrained Problem).
Let be $\alpha > 0$. Then there exists a minimizer of (4.4) and (4.8).

Proof.
Obviously

$$F(U) := \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij}$$

is convex. Using Proposition 4.2 we obtain that (4.4) and (4.8) are lower semi-continuous. The sublevel set

$$\mathcal{S}_\xi = \left\{ U \in G \;\middle|\; \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} + \alpha \max_{i \in \{1,\dots,M\}} \sum_{j=1}^{N} u_{ij} \leq \xi \right\}$$

with

$$\xi := \frac{1}{2} \|W\|_F^2$$

is not empty, since we obviously have $0 \in \mathcal{S}_\xi$.
Analogously to the proof of Theorem 4.2, we see that $\mathcal{S}_\xi$ is bounded. Due to the finite dimensionality of the problem, we have compactness of the bounded sublevel set $\mathcal{S}_\xi$. Together with semi-continuity we obtain existence of a minimizer of the unconstrained problems (4.4) and (4.8). □

Uniqueness

Let us now shortly discuss potential uniqueness of the solutions of (4.7) and (4.8).

Theorem 4.4 (Restriction of the Solution Set)**.**
There exists a solution $\left(\bar{U},\ \bar{v}\right)$ of (4.7) and (4.8) with $\bar{v}$ minimal, i.e. $\bar{v} \leq v$ for all minimizers $(U,\ v)$. Furthermore, $\bar{v}$ is unique.

Proof.
Obviously $\bar{v}$ can be defined as

$$\bar{v} := \inf \{ v \mid (U,\ v) \text{ is a minimizer of (4.6)} \}$$

with F as in (4.7), (4.8) respectively. Due to Theorem 4.2 and 4.3, we know that $\bar{v} < \infty$ has to hold. We proof the assumption via contradiction.
Assume there exists no $\bar{U}$ with $(\bar{U},\ \bar{v})$ being a minimizer of (4.7), (4.8) respectively. We can find a sequence of minimizers $(U_k,\ v_k)$ with $v_k \to \bar{v}$. U_k is bounded, since $U_k \in \mathcal{S}_\xi$ for all k. Thus there exists a subsequence $(U_{k_l},\ v_{k_l})$. Finally lower semi-continuity provides us with the limit $(\bar{U},\ \bar{v})$ being a minimizer, which is a contradiction to the assumption. Furthermore, since we have $\bar{v} \in \mathbb{R}^+$, it is obviously unique. □

In Theorem 4.4 we have seen that we can reduce the solution set to those solutions with optimal v, i.e.

$$\mathcal{S} := \left\{ (\bar{U},\ \bar{v}) \in G \times \mathbb{R}^+ \text{ is a minimizer of (4.6)} \mid \bar{v} \text{ minimal} \right\} ,$$

with F as in (4.7), (4.8) respectively. There always exists a unique $\bar{v} \in \mathbb{R}^+$, however, in general we are not able to deduce uniqueness for $(U,\ \bar{v}) \in G \times \mathbb{R}^+$.

4.1.3. Subdifferentials and Source Conditions

In this subsection we characterize the subdifferentials of the $\ell^{1,\infty}$-norm and its nonnegative counterpart. Furthermore, we discuss what kind of solutions $\widehat{U}$ to $A\widehat{U}B^T = W$ are likely to meet a source condition for the $\ell^{1,\infty}$-regularization.

The Subdifferential of $\ell^{1,\infty}$

We start by computing the subdifferential of the $\ell^{1,\infty}$-norm.

Theorem 4.5 (Subdifferential of the $\ell^{1,\infty}$-Functional)**.**
Let be $U,\ P \in \mathbb{R}^{M\times N}$. The subdifferential of $\|U\|_{1,\infty}$ can be characterized as follows: Let I be the set of indices, where U attains its maximum row-ℓ^1-norm, i.e.

$$I = \left\{ i \in \{1, \dots, M\} \;\middle|\; \sum_{j=1}^{N} |u_{ij}| = \max_{m \in \{1,\dots,M\}} \sum_{j=1}^{N} |u_{mj}| \right\}.$$

Then the following equivalence holds:

$$P \in \partial \|U\|_{1,\infty} \quad \Leftrightarrow \quad \begin{cases} p_{ij} = w_i \operatorname{sign}(u_{ij}), & \text{if } i \in I, \\ p_{ij} = 0, & \text{if } i \notin I, \end{cases} \tag{4.11}$$

with weights $w_i \geq 0$ such that $\sum\limits_{i\in I} w_i = 1$ holds if $u \neq 0$ and $\sum\limits_{i\in I} w_i \leq 1$ holds if $u \equiv 0$.

Proof.
First let P be chosen according to the characterization on the right hand side of (4.11). Since $\|U\|_{1,\infty}$ is 1-homogeneous, it is sufficient to verify the following two conditions (cf. Lemma 2.5):

1. $\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{N} p_{ij} u_{ij} = \|U\|_{1,\infty}$,
2. $\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{N} p_{ij} v_{ij} \leq \|V\|_{1,\infty} \quad \text{for all} \quad V \in \mathbb{R}^{M\times N}$.

Regarding 1., we have

$$\sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}u_{ij} = \sum_{i\in I}\sum_{j=1}^{N} w_i \operatorname{sign}(u_{ij})u_{ij} = \sum_{i\in I} w_i \sum_{j=1}^{N} |u_{ij}| = \|U\|_{1,\infty} \sum_{i\in I} w_i = \|U\|_{1,\infty} .$$

Regarding 2., we have

$$\begin{aligned} \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij} &\leq \left|\sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij}\right| \leq \sum_{i\in I}\sum_{j=1}^{N} |p_{ij}v_{ij}| \leq \sum_{i\in I} w_i \sum_{j=1}^{N} |v_{ij}| \\ &\leq \sum_{i\in I} w_i \max_m \sum_{j=1}^{N} |v_{mj}| \leq \max_m \sum_{j=1}^{N} |v_{mj}| = \|V\|_{1,\infty} . \end{aligned}$$

Thus we see that P chosen in this way belongs to the subdifferential of $\|U\|_{1,\infty}$.

Now let $P \in \partial\|U\|_{1,\infty}$, such that conditions 1. and 2. hold.
We will prove the characterization of the subdifferential in five steps:

1. We show that $p_{mn} = 0$ holds for all $m \notin I$.
   Let be $m \notin I$ and let n be arbitrary. Consider V with $v_{ij} = u_{ij}$ for all (i,j) except for $v_{mn} = u_{mn} + \varepsilon\operatorname{sign}(p_{mn})$. Since we have $m \notin I$, we can choose $\varepsilon > 0$ small enough such that $\|U\|_{1,\infty} = \|V\|_{1,\infty}$ holds. Thus we have

$$\begin{aligned} \|V\|_{1,\infty} - \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij} &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}u_{ij} - \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij} \\ &= p_{mn}(u_{mn} - v_{mn}) = -\varepsilon|p_{mn}| . \end{aligned}$$

   Since $\|V\|_{1,\infty} - \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij} \geq 0$ holds by condition 2., we have shown that $p_{mn} = 0$ for all $m \notin I$.

2. Furthermore, we show that $u_{mn}p_{mn} \geq 0$ for $m \in I$ and $u_{mn} \neq 0$.
   Let V be an element with $v_{ij} = u_{ij}$ for all $(i,j) \neq (m,n)$ and $v_{mn} = -u_{mn}$ for an arbitrary (m,n) with $m \in I$ and $u_{mn} \neq 0$. Clearly we have $\|U\|_{1,\infty} = \|V\|_{1,\infty}$, which yields

$$\begin{aligned} \|U\|_{1,\infty} = \|V\|_{1,\infty} \geq \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij} &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}u_{ij} - 2u_{mn}p_{mn} \\ &= \|U\|_{1,\infty} - 2u_{mn}p_{mn} \end{aligned}$$

   and hence we obtain $u_{mn}p_{mn} \geq 0$.

3. Now we prove that for $m \in I$ we have $|p_{mn}| = |p_{md}| = \text{const}$ for all (m,n) and (m,d) with $u_{mn} \neq 0$ and $u_{md} \neq 0$.
   Let V be an element with $v_{ij} = u_{ij}$ for all $(i,j) \notin \{(m,n),(m,d)\}$ and $v_{mn} = u_{mn} - \text{sign}(u_{mn})\varepsilon$, $v_{md} = u_{md} + \text{sign}(u_{md})\varepsilon$. Let $\varepsilon \neq 0$ be chosen small enough such that $\text{sign}(u_{mn}) = \text{sign}(v_{mn})$ and $\text{sign}(u_{md}) = \text{sign}(v_{md})$. Clearly $\|U\|_{1,\infty} = \|V\|_{1,\infty}$ holds such that we have

$$\begin{aligned}\|U\|_{1,\infty} = \|V\|_{1,\infty} &\geq \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}v_{ij} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}u_{ij} + (\text{sign}(u_{md})p_{md} - \text{sign}(u_{mn})p_{mn})\varepsilon \\ &= \|U\|_{1,\infty} + (\text{sign}(u_{md})p_{md} - \text{sign}(u_{mn})p_{mn})\varepsilon\ .\end{aligned}$$

   Thus we obtain $(\text{sign}(u_{md})p_{md} - \text{sign}(u_{mn})p_{mn})\varepsilon \leq 0$. Since the sign of ε is arbitrary, the reverse inequality holds as well and we obtain $\text{sign}(u_{mn})p_{mn} = \text{sign}(u_{md})p_{md}$. Since $u_{mn} \neq 0$ and $u_{md} \neq 0$ holds, we have $|\text{sign}(u_{mn})| = |\text{sign}(u_{md})| = 1$. Hence we obtain $|p_{mn}| = |p_{md}| = \text{const}$.
   In the following we will denote this constant (depending on k) as

$$w_m := |p_{mn}| \quad \forall\, n \in \{1,\dots,N\},\ m \in I\ .$$

4. Let us now show that $\sum_{i=1}^{M} w_i = 1$ has to hold if we have $u \neq 0$.
   For this purpose we compute

$$\begin{aligned}\|U\|_{1,\infty} &= \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}p_{ij} = \sum_{i\in I}\sum_{\{j|u_{ij}\neq 0\}} u_{ij}p_{ij} = \sum_{i\in I}\sum_{\{j|u_{ij}\neq 0\}} w_i u_{ij}\text{sign}(u_{ij}) \\ &= \sum_{i\in I} w_i \max_{m\in\{1,\dots,M\}} \sum_{j=1}^{N} |u_{mj}| = \|U\|_{1,\infty}\sum_{i\in I} w_i\end{aligned}$$

   in order to see that the above claim holds.
   Note that from 2. we obtain $\text{sign}(u_{mn}) = \text{sign}(p_{mn})$ and thus we have $p_{mn} = w_m\text{sign}(u_{mn})$ for $m \in I$.

5. Finally we show that $\sum_{i=1}^{M} w_i \leq 1$ holds in case we have $U = 0$.
   In case that we have $U \equiv 0$, let V be such that $v_{ij} = \text{sign}(p_{ij})$ holds at exactly

one index j per row i, where $|p_{ij}| = \max_{n\in\{1,\dots,N\}} |p_{in}|$ and $v_{ij} = 0$ else. We have

$$1 = \|V\|_{1,\infty} \geq \sum_{i=1}^{M}\sum_{j=1}^{N} v_{ij}p_{ij} = \sum_{i=1}^{M} \max_{n\in\{1,\dots,N\}} |p_{in}|.$$

This shows that $\sum_{i=1}^{M} \max_{n\in\{1,\dots,N\}} |p_{in}| \leq 1$ has to be true and we can again denote $w_i = \max_{n\in\{1,\dots,N\}} |p_{in}|$.

We can see that the collection of these five results show that P indeed meets all requirements on the right hand side of (4.11). □

One particular thing we can see from the proof of Theorem 4.5 is that $P \in \partial\|U\|_{1,\infty}$ for an arbitrary U meets

$$\|P\|_{\infty,1} = \sum_{i=1}^{M} \max_{j\in\{1,\dots,N\}} |p_{ij}| \leq 1$$

and

$$\|P\|_{\infty,1} = \sum_{i=1}^{M} \max_{j\in\{1,\dots,N\}} |p_{ij}| = 1 \qquad \text{for} \qquad u \neq 0\ .$$

Thus we see that the $\ell^{\infty,1}$-norm is the dual norm to the $\ell^{1,\infty}$-norm, which has already been observed by Tropp (2006a).

The Subdifferential of the Nonnegative $\ell^{1,\infty}$-Formulation

Let us now consider the nonnegative $\ell^{1,\infty}$-functional (4.10).

Theorem 4.6 (Subdifferential of the Nonnegative $\ell^{1,\infty}$-Functional).
Let $P^{1,\infty} \in \mathbb{R}^{M\times N}$ be the subdifferential of the $\ell^{1,\infty}$-norm characterized as before in (4.11) and let $p_{ij}^{1,\infty}$ be its entries for $i \in \{1,\dots,M\}$ and $j \in \{1,\dots,N\}$. Then the subdifferential of the nonnegative $\ell^{1,\infty}$-functional

$$\mathcal{R}(U) := \begin{cases} \max_i \sum_{j=1}^{N} u_{ij}, & \text{if } u_{ij} \geq 0, \\ \infty, & \text{else,} \end{cases}$$

can be characterized as

$$P \in \partial\mathcal{R}(U) \quad \Leftrightarrow \quad p_{ij} = p_{ij}^{1,\infty} + \mu_{ij}\ ,$$

where μ_{ij} for all $i \in \{1, \ldots, M\},\ j \in \{1, \ldots, N\}$ are the Lagrange parameters with

$$\mu_{ij} \begin{cases} = 0, & \text{if } u_{ij} \neq 0, \\ \leq 0, & \text{if } u_{ij} = 0. \end{cases}$$

Proof.
$\mathcal{R}(U)$ can be written using the characteristic function, i.e.

$$\mathcal{R}(U) = \|U\|_{1,\infty} + \chi_{\{U \in \mathbb{R}^{M \times N} \,|\, u_{ij} \geq 0 \ \ \forall\, i,\, j\}} \ .$$

In case that the subdifferential is additive, we have

$$\partial \mathcal{R}(U) = \partial \|U\|_{1,\infty} + \partial \chi_{\{U \in \mathbb{R}^{M \times N} \,|\, u_{ij} \geq 0 \ \ \forall\, i,\, j\}}$$

and directly obtain

$$P \in \partial \mathcal{R}(U) \quad \Leftrightarrow \quad p_{ij} = p_{ij}^{1,\infty} + \mu_{ij} \ .$$

In order to prove that in this case the subdifferential is additive, we have to show that the following conditions hold (cf. Lemma 2.4):

1. $\|U\|_{1,\infty}$ and $\chi_{\{U \in \mathbb{R}^{M \times N} \,|\, u_{ij} \geq 0 \ \ \forall\, i,\, j\}}$ are proper, convex and lower semi-continuous,
2. there exists a $\bar{U} \in \operatorname{dom} \|U\|_{1,\infty} \cap \operatorname{dom} \chi_{\{U \in \mathbb{R}^{M \times N} \,|\, u_{ij} \geq 0 \ \ \forall\, i,\, j\}}$, where one of the two functionals is continuous.

This is quite easy to see:

1. Since $\|U\|_{1,\infty}$ is a norm, it is obviously proper and convex. Furthermore, we see that $\chi_{\{U \in \mathbb{R}^{M \times N} \,|\, u_{ij} \geq 0 \ \ \forall\, i,\, j\}}$ is obviously proper. It is also convex, because the characteristic function of a convex set is also convex. Both functionals are lower semi-continuous, since we are in a finite dimensional setting.
2. Let $\bar{u}_{ij} > 0$ for all $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N\}$. Then we have

$$\bar{U} \in \operatorname{dom} \|U\|_{1,\infty} \cap \operatorname{dom} \chi_{\{U \in \mathbb{R}^{M \times N} \,|\, u_{ij} \geq 0 \ \ \forall\, i,\, j\}} \ .$$

   Furthermore, both functionals are continuous at $\bar{U}$.

Thus we see that the subdifferential is additive and we obtain the assumption. □

Remark 4.1.
Obviously $P \in \partial\mathcal{R}(U)$ can be characterized as follows:

$$p_{ij} \begin{cases} = 0, & \text{if } i \notin I \text{ and } u_{ij} > 0, \\ \leq 0, & \text{if } i \notin I \text{ and } u_{ij} = 0, \\ = w_i, & \text{if } i \in I \text{ and } u_{ij} > 0, \\ \leq w_i, & \text{if } i \in I \text{ and } u_{ij} = 0. \end{cases}$$

Source Conditions

Knowing the characterization of the subgradient, we can state a condition, which allows us to determine whether a certain solution to $AUB^T = W$ is $\ell^{1,\infty}$-minimizing. We will call this condition a *source condition* as used in the inverse problem and error estimate literature e.g. in Burger and Osher (2004), Engl et al. (1996), Schuster et al. (2012). However, we would like to point out that similar conditions have been called *dual certificate* in the compressed sensing literature (c.f. Candès et al. (2011), Candès and Plan (2010), Hsu et al. (2011), Zhang and Zhang (2012)).

Definition 4.1.
We say that a solution $\widehat{U}$ of $A\widehat{U}B^T = W$ meets a source condition with respect to a proper, convex regularization functional J if there exists a Q such that $P = A^TQB \in \partial J(\widehat{U})$.

The source condition of some $\widehat{U}$ with respect to J is nothing but the optimality condition for $\widehat{U}$ being a J-minimizing solution to $A\widehat{U}B^T = W$.

Lemma 4.3 (cf. Burger and Osher (2004))**.**
Let $\widehat{U}$ with $A\widehat{U}B^T = W$ meet a source condition with respect to J. Then $\widehat{U}$ is a J-minimizing solution.

Considering this, the next question naturally emerges for our characterization of the subdifferential, i.e what kind of solutions $\widehat{U}$ to $A\widehat{U}B^T = W$ are likely to meet a source condition for $\ell^{1,\infty}$-regularization. Particularly, we are interested in investigating how likely $\ell^{0,\infty}$-minimizing solutions are to meet a source condition.
Due to the similarity between the subgradient of the ℓ^1-norm and the subgradient of $\ell^{1,\infty}$-norm at rows with index $i \in I$, we can make the following simple observation:

Lemma 4.4.
Let $\widehat{U}$ be an ℓ^1-minimizing solution to $A\widehat{U}B^T = W$ for which $\sum_{j=1}^N |\hat{u}_{ij}| = \sum_{j=1}^N |\hat{u}_{mj}|$ for all $i, m \in \{1, \dots, M\}$, then $\widehat{U}$ also is an $\ell^{1,\infty}$-minimizing solution.

Proof.
The ℓ^1-subgradient divided by the number of rows is an $\ell^{1,\infty}$-subgradient. □

The above lemma particularly shows that exact recovery criteria for the properties of the sensing matrix kron(B, A) (like the Restricted Isometry Property, the Null Space Property, or the Mutual Incoherence Property), are sufficient for the exact recovery of sparse solutions with the same ℓ^1-norm in each row.
Of course, we do expect to recover more $\ell^{0,\infty}$-minimizing solutions than just the ones with the same ℓ^1-row-norm. Looking at the characterization of the subdifferential (cf. Remark 4.1), we can observe that there are two cases that pose much more severe restrictions, i.e. the equality constraints, than the two other cases (which only lead to inequality constraints). Thus we generally expect solutions, which require only a few of the equality constraints to be more likely to meet a source condition. As we can see equality constraints need to be met for nonzero elements, such that the $\ell^{1,\infty}$-regularization prefers sparse solutions. Additionally the constraints are less restrictive if the corresponding row has maximal ℓ^1-norm. Thus we expect those solutions to be likely to meet a source condition that reach a maximal ℓ^1-norm in as many rows as possible while being sparse. Naturally, these solutions will be row-sparse, which further justifies the idea that the $\ell^{1,\infty}$-norm can be used as a convex approximation of the $\ell^{0,\infty}$-problem.

4.1.4. Equivalence of Formulations

In this subsection we want to show that the minimizers of (4.5) and (4.9) coincide under certain circumstances. In order to do so, we examine how the regularization parameter $\alpha > 0$ is connected to the minimizer of (4.5). For this purpose we first prove the continuity and monotonicity of

$$\alpha \mapsto \max_{i=1,\dots,M} \sum_{j=1}^{N} u_{ij}(\alpha) \; . \tag{4.12}$$

Afterwards, we shall investigate the meaning of $\tilde{v}$ in (4.9) and its connection to the minimizer of $F(U)$ with minimal $\ell^{1,\infty}$-norm. Analyzing the limits of (4.12) leads us to the main result of this subsection and the connection between the two problems (4.5) and (4.9).

Remark 4.2.
For most of the proofs in this subsection we require $F(U)$ to be continuous. Thus most of the results are not useful for the constrained problem (4.3), since

$$F(U) = \begin{cases} 0, & \text{if } AUB^T = W, \\ \infty, & \text{else,} \end{cases}$$

is not continuous. Nevertheless, in reality we have to deal with noisy data anyway and thus we only want to implement the unconstrained problem (4.4). This will become easier, since we can simply use its reformulation (4.9), which we will summarize in Theorem 4.7.

For this subsection we define the functional $J_\alpha : G \longrightarrow \mathbb{R}^+$ via

$$J_\alpha(U) := F(U) + \alpha \max_{i \in \{1,\dots,M\}} \sum_{j=1}^{N} u_{ij} \ . \tag{4.13}$$

Lemma 4.5 (Continuity of (4.12))**.**
Let $F : \mathbb{R}^{M \times N} \longrightarrow \mathbb{R}^+$ be a convex continuous functional with bounded sublevel sets. Let $U(\alpha) \in G$ be a minimizer of J_α with $\|U(\alpha)\|_{1,\infty}$ minimal.
Then $\alpha \mapsto \max_{i \in \{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha)$ is a continuous function.

Proof.
Let U_k be a minimizer of J_{α_k} and let $\alpha_k \to \alpha$ be a sequence of regularization parameters. Due to boundedness, we are able to find subsequences $U_{k_l} \to U$. Because of the convergence of the subsequences, the limits of all subsequences are equal and we obtain convergence of the whole sequence $U_k \to U$. We prove by contradiction and claim that U is not a minimizer of J_α. In this case there would exist a $\widetilde{U}$ with

$$J_\alpha(\widetilde{U}) < J_\alpha(U) \ . \tag{4.14}$$

By defining $U^\dagger := \frac{\alpha_k}{\alpha}\widetilde{U}$, we obtain

$$\begin{aligned} J_\alpha(U^\dagger) &= F(U^\dagger) + \alpha \max_{i \in \{1,\dots,M\}} \sum_{j=1}^{N} u^\dagger_{ij} \\ &= F\left(\frac{\alpha_k}{\alpha}\widetilde{U}\right) + \alpha_k \max_{i \in \{1,\dots,M\}} \sum_{j=1}^{N} \tilde{u}_{ij} \\ &= J_{\alpha_k}(\widetilde{U}) + F\left(\frac{\alpha_k}{\alpha}\widetilde{U}\right) - F(\widetilde{U}) \ . \end{aligned}$$

The continuity of F yields

$$F\left(\frac{\alpha_k}{\alpha}\widetilde{U}\right) - F(\widetilde{U}) \to 0$$

for $k \to \infty$ and thus we obtain

$$J_\alpha(U^\dagger) = J_{\alpha_k}(\widetilde{U}) \, .$$

Since J is continuous, we obtain for $k \to \infty$ that

$$\begin{aligned} J_{\alpha_k}(\widetilde{U}) &\to J_\alpha(\widetilde{U}) \qquad \text{and} \\ J_{\alpha_k}(U_k) &\to J_\alpha(U) \end{aligned}$$

hold. By using (4.14), we see that the inequality

$$J_{\alpha_k}(\widetilde{U}) < J_{\alpha_k}(U_k)$$

has to hold as well. This is a contradiction to the assumption that U_k is a minimizer of J_{α_k}. Hence U has to be a minimizer of J_α and we see that $\alpha \mapsto \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha)$ is continuous. Thus we also now that (4.12) is continuous on $(0, \infty)$. □

Lemma 4.6 (Monotonicity of (4.12))**.**
Let $F : \mathbb{R}^{M\times N} \longrightarrow \mathbb{R}^+$ be a convex continuous functional with bounded sublevel sets.
Let $U(\alpha) \in G$ be a minimizer of J_α with $\|U(\alpha)\|_{1,\infty}$ minimal.
Then $\alpha \mapsto \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha)$ is a monotonically decreasing function.

Proof.
Let be $0 < \alpha < \beta$. $U(\beta)$ is feasible for J_α, i.e. $u_{ij}(\beta) \geq 0$. Thus we obtain

$$\begin{aligned} &F\left(U\left(\alpha\right)\right) + \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \\ &\leq F\left(U\left(\beta\right)\right) + \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\beta) \\ &= F\left(U\left(\beta\right)\right) + \beta \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\beta) + (\alpha-\beta) \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\beta) \, . \end{aligned} \tag{4.15}$$

On the other hand we have

$$F\left(U(\beta)\right) + \beta \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\beta) \leq F\left(U(\alpha)\right) + \beta \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \, ,$$

since $U(\alpha)$ is feasible for J_β. Inserting this in (4.15) yields

$$\begin{aligned} & F\left(U(\alpha)\right) + \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\alpha) \le F\left(U(\alpha)\right) + \beta \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\alpha) \\ & \qquad\qquad + (\alpha-\beta) \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\beta) \,, \\ \Leftrightarrow \quad & (\alpha-\beta) \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\alpha) \le (\alpha-\beta) \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\beta) \,, \\ \Leftrightarrow \quad & \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\alpha) \ge \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\beta) \,, \end{aligned}$$

since we have $\alpha < \beta$. We conclude that $\alpha \to \max_{i\in\{1,\dots,M\}} \sum_{j=1}^N u_{ij}(\alpha)$ is monotonically decreasing. □

Lemma 4.7.
Let $F : \mathbb{R}^{M\times N} \longrightarrow \mathbb{R}^+$ be a convex continuous functional with bounded sublevel sets. Let $\bar{U} \in G$ be a solution of (4.9) such that $\sum_{j=1}^N \bar{u}_{ij} < \tilde{v}$ holds for all $i \in [1,\dots,M]$. Then we have $\tilde{v} > \|\widehat{U}\|_{1,\infty}$, where $\widehat{U} \in G$ is a minimizer of $F(U)$ with $\|U\|_{1,\infty}$ minimal and vice versa.

Proof.
Let be $\tilde{v} > \|\widehat{U}\|_{1,\infty}$, then $\widehat{U}$ is feasible for (4.9) and obviously a minimizer as well.
Let be $\tilde{v} \le \|\widehat{U}\|_{1,\infty}$. Then in case that $\sum_{j=1}^N \bar{u}_{ij} < \tilde{v}$ holds for all i, we obviously have $\bar{U} \ne \widehat{U}$ and thus we obtain

$$F(\bar{U}) > F(\widehat{U}) \,,$$

since $\widehat{U}$ is a minimizer of F with $\|\widehat{U}\|_{1,\infty}$ minimal. Due to convexity, we obtain

$$F(\varepsilon\widehat{U} + (1-\varepsilon)\bar{U}) \le \varepsilon F(\widehat{U}) + (1-\varepsilon)F(\bar{U}) < F(\bar{U})$$

and for small ε we have

$$\varepsilon \sum_{j=1}^N \hat{u}_{ij} + (1-\varepsilon) \sum_{j=1}^N \bar{u}_{ij} \le \tilde{v} \,.$$

Thus we have found an element with smaller value of F as the minimizer $\bar{U}$, which is a contradiction. □

Lemma 4.8 (Limits of (4.12))**.**
Let $F : \mathbb{R}^{M\times N} \longrightarrow \mathbb{R}^+$ be a convex continuous functional with bounded sublevel sets. Then we have

$$\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \to 0 \quad \text{for} \quad \alpha \to \infty \quad \text{and}$$
$$\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \to \|\widehat{U}\|_{1,\infty} \quad \text{for} \quad \alpha \to 0$$

with $\widehat{U} \in G$ being a minimizer of F with $\|U\|_{1,\infty}$ minimal.

Proof.
Let $U(\alpha)$ be a minimizer of J_α as proposed in (4.13).

1. Consider the case of $\alpha \to \infty$. $U = 0$ is feasible for J_α, thus we obtain

$$F\left(U(\alpha)\right) + \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \leq F(0)\ .$$

Therefore, $\alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha)$ is bounded by $F(0)$ and we have

$$\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \to 0 \quad \text{for} \quad \alpha \to \infty\ .$$

2. Now consider the case of $\alpha \to 0$. We can find a subsequence $\alpha_k \to 0$ such that

$$U\left(\alpha_k\right) \to \widehat{U}$$

holds, where $\widehat{U}$ is a minimizer of F with $\|U\|_{1,\infty}$ minimal. Obviously $\widehat{U}$ is feasible for J_{α_k}. Hence we obtain

$$F\left(U(\alpha_k)\right) + \alpha_k \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha_k) \leq F(\widehat{U}) + \alpha_k \|\widehat{U}\|_{1,\infty}\ .$$

Since $\widehat{U}$ is a minimizer of F, it has to hold that $F(U(\alpha_k)) \geq F(\widehat{U})$ and thus we obtain

$$\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha_k) \leq \|\widehat{U}\|_{1,\infty}\ .$$

Obviously $U(\alpha)$ is feasible for (4.9) with

$$\tilde{v} = \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha_k) \ .$$

Then Lemma 4.7 yields

$$\sum_{j=1}^{N} u_{ij}(\alpha_k) = \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha_k) = \|U(\alpha_k)\|_{1,\infty} \quad \text{for some } i \in \{1,\dots,M\} \ .$$

Due to the lower semi-continuity of the norm, we obtain

$$\liminf_{\alpha_k\to 0} \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha_k) = \liminf_{\alpha_k\to 0} \|U(\alpha_k)\|_{1,\infty} \geq \|\widehat{U}\|_{1,\infty}$$

and finally we have

$$\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) \to \|\widehat{U}\|_{1,\infty} \quad \text{for} \quad \alpha \to 0 \ .$$

□

In Figure 4.1 we see an example on how the function

$$v(\alpha) := \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha)$$

could look, where $U(\alpha) \in G$ is a minimizer of J_α with $\|U(\alpha)\|_{1,\infty}$ minimal.

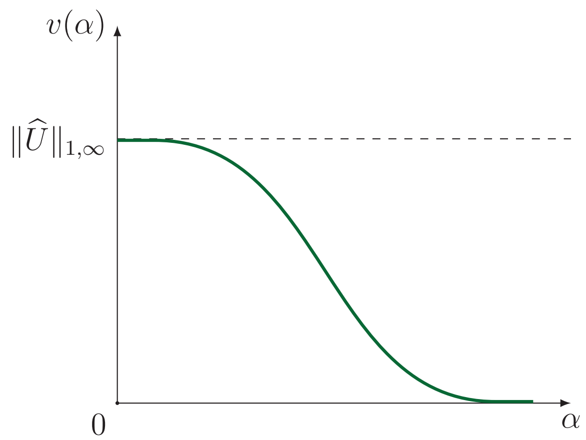


Figure 4.1.: Illustration of the relation between the regularization parameter α and the nonnegative $\ell^{1,\infty}$-norm $v(\alpha)$ of the corresponding minimizer as defined above

Remark 4.3.
In case that $\partial F(0) \neq \emptyset$ holds, we have

$$\max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij}(\alpha) = 0$$

already for $\alpha < \infty$, but large enough.
We see this by considering $p_0 \in \partial F(0)$ and then selecting $p = -\frac{1}{\alpha} p_0$. Here we choose α large enough that we have $\|p\|_{\infty,1} < 1$ and obtain

$$\langle p, v\rangle \leq \|p\|_{\infty,1} J(v) \leq J(v) \ .$$

Thus p is a subgradient at $U(\alpha) = 0$ and we see that the optimality condition is fulfilled.

Finally we can conclude the following essential statement:

Theorem 4.7 (Connection of the Solutions of (4.6) and (4.9))**.**
Let $F: \mathbb{R}^{M\times N} \longrightarrow \mathbb{R}^+$ be a convex continuous functional with bounded sublevel sets. Let be $\tilde{v} \in (0, \|\widehat{U}\|_{1,\infty})$, where $\widehat{U}$ is a minimizer of $F(U)$ with $\|U\|_{1,\infty}$ minimal. Let $\bar{U} \in G$ be a solution of

$$\min_{U\in G} \ F(U) \quad \text{s. t.} \quad \sum_{j=1}^{N} u_{ij} \leq \tilde{v} \ . \tag{4.9}$$

Then there exists an $\alpha > 0$ such that $\bar{U}$ is a solution of

$$\min_{U\in G} \ F(U) + \alpha \max_{i\in\{1,\dots,M\}} \sum_{j=1}^{N} u_{ij} \ . \tag{4.5}$$

Remark 4.4.
If U is a solution of (4.9), we can directly decide, whether there exists an α for this problem, i.e. in the case of

$$\tilde{v} = \|U\|_{1,\infty} \ .$$

Theorem 4.8 (Existence of a Solution of (4.9))**.**
Let $F: \mathbb{R}^{M\times N} \longrightarrow \mathbb{R}^+$ be a convex continuous functional with bounded sublevel sets. Then there exists a minimizer of (4.9).

Proof.
We can write (4.9) using the characteristic function, i.e.

$$\min_{U \in G} F(U) + \chi_{\left\{U \in \mathbb{R}^{M\times N} \,|\, \sum_{j=1}^{N} u_{ij} \leq \tilde{v} \,\forall i \in \{1,\dots,M\}\right\}} .$$

For $\xi \in \mathbb{R}$ we consider the sublevel set

$$S_\xi = \left\{ U \in G \,|\, F(U) + \chi_{\left\{U \in \mathbb{R}^{M\times N} \,|\, \sum_{j=1}^{N} u_{ij} \leq \tilde{v} \,\forall i \in \{1,\dots,M\}\right\}} \leq \xi \right\} .$$

Since F is continuous and $\chi(0) = 0$ holds, we have $0 \in S_\xi$ and thus S_ξ is not empty. Furthermore, S_ξ is bounded, since the norm of U is bounded. Additionally the functional stays lower semicontinuous and we obtain the existence of a minimizer of (4.9). □

Using Theorem 4.7 for problem (4.8), we obtain

$$\min_{U \in G} \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} \quad \text{s.t.} \quad \sum_{j=1}^{N} u_{ij} \leq \tilde{v} . \tag{4.16}$$

Note that we need to look for a suitable regularization parameter in the implementation anyway. Thus we can instead determine a suitable $\tilde{v}$ and obtain an easier optimization problem.

4.1.5. Asymptotic 1-Sparsity

Let us now consider the case of (4.16) with $\beta = 0$, i.e.

$$\min_{U \in G} \frac{1}{2} \left\| AUB^T - W \right\|_F^2 \qquad \text{s. t.} \qquad \sum_{j=1}^{N} u_{ij} \leq \tilde{v} . \tag{4.17}$$

We observe that we indeed obtain a special kind of sparsity in every row in the case that the regularizing parameter $\tilde{v}$ becomes sufficiently small.
In order to analyze Problem (4.17) asymptotically, we consider the rescaling $X := \tilde{v}^{-1}U \Leftrightarrow U = \tilde{v}X$ and obtain the new variational problem

$$\min_{X \in G} \frac{1}{2} \left\| \tilde{v}AXB^T - W \right\|_F^2 \qquad \text{s. t.} \qquad \sum_{j=1}^{N} x_{ij} \leq 1 . \tag{4.18}$$

Let us now analyze the structure of a solution of (4.18) for $\tilde{v} \to 0$.

Theorem 4.9.
Let k_i be the number of maxima in the ith row of $G := A^T W B$ and let $X(\tilde{v})$ be a minimizer of (4.18). Then the ith row of

$$\bar{X} := \lim_{\tilde{v} \to 0} X(\tilde{v})$$

is at most k_i-sparse.

Proof.
After simplifying the norm and dividing by $\tilde{v}$ in (4.18), we can equivalently consider

$$\min_{X \in G} \frac{\tilde{v}}{2} \left\| A X B^T \right\|_F^2 - \left\langle A X B^T, W \right\rangle \qquad \text{s. t.} \qquad \sum_{j=1}^{N} x_{ij} \leq 1 \, .$$

For the case that we have $\tilde{v} \to 0$, the first summand tends to zero and thus

$$\max_{X \in G} \left\langle X, A^T W B \right\rangle \qquad \text{s. t.} \qquad \sum_{j=1}^{N} x_{ij} \leq 1 \tag{4.19}$$

holds. With the above definition of G we shall now consider

$$\max_{X \in G} \sum_{i=1}^{M} \sum_{j=1}^{N} g_{ij} x_{ij} \qquad \text{s. t.} \qquad \sum_{j=1}^{N} x_{ij} \leq 1 \, .$$

Let J_i be the column index set at which the maximum of the ith row of G is reached, i.e.

$$J_i = \{ n \in \{1, \ldots, N\} \,|\, g_{in} \geq g_{ij} \;\forall\, j \in \{1, \ldots, N\} \}$$

for every $i \in \{1, \ldots, M\}$. Since we have the constraint that the row sum of X should not exceed 1, it has to hold that

$$\sum_{j=1}^{N} g_{ij} x_{ij} \leq g_{in} \qquad \forall\, n \in J_i$$

for every $i \in \{1, \ldots, M\}$. Hence we see that for a solution of (4.19) has to hold

$$\sum_{n \in J_i} x_{in} = 1 \qquad \text{and} \qquad x_{ij} = 0 \quad \forall\, j \notin J_i \, .$$

This means that the ith row of the solution of (4.19) has at most k_i nonzero entries. Thus the ith row of $\bar{X}$ is at most k_i-sparse, maybe even sparser. □

Remark 4.5.
In case that the ith row of $\bar{X}$ is k_i-sparse, the asymptotic solution $\bar{X}$ has nonzero entries at the same positions as $G = A^T W B$ has its maxima in each row.
Note that the row-maxima are not necessarily unique. However, in the case that for every $i \in \{1, \ldots, M\}$ the index set J_i contains only one element, the rows of $\bar{X}$ are 1-sparse.

Theorem 4.9 raises the question, whether there exists a small regularization parameter $\tilde{v}$, for which $X(\tilde{v})$ is already k_i-sparse. In this case we could apply this knowledge to the original problem (4.17), which is not possible in the limit case, since then $\bar{U} := \lim_{\tilde{v} \to 0} U(\tilde{v})$ would be equal to zero.

Theorem 4.10.
Let the ℓ^2-norm of the columns of $A \in \mathbb{R}^{L \times M}$ and $B \in \mathbb{R}^{T \times N}$ be nonzero, i.e.

$$\|a_{\cdot i}\|_2 > 0 \quad \forall\, i \in \{1, \ldots, M\} \qquad \text{and} \qquad \|b_{\cdot j}\|_2 > 0 \quad \forall\, j \in \{1, \ldots, N\}\,.$$

Then there exists a regularization parameter $\tilde{v} > 0$ such that the solution of (4.17) has nonzero entries at the same positions as $G := A^T W B$ has row-maxima.

Proof.
Consider the rescaled problem (4.18). After simplifying the norm and dividing by $\tilde{v}$, we consider equivalently

$$\min_{X \in G} \frac{\tilde{v}}{2} \left\|AXB^T\right\|_F^2 - \left\langle X, A^T W B \right\rangle \qquad \text{s. t.} \qquad \sum_{j=1}^{N} x_{ij} \leq 1$$

and thus

$$\min_{X \in G} \frac{\tilde{v}}{2} \sum_{l=1}^{L} \sum_{k=1}^{T} \left(\sum_{i=1}^{M} \sum_{j=1}^{N} a_{li} x_{ij} b_{kj} \right)^2 - \sum_{i=1}^{M} \sum_{j=1}^{N} x_{ij} g_{ij} \qquad \text{s. t.} \qquad \sum_{j=1}^{N} x_{ij} \leq 1\,,$$

where we use again $G := A^T W B$. The Lagrange functional reads as follows:

$$\begin{aligned} \mathcal{L}(X; \lambda, \mu) = {} & \frac{\tilde{v}}{2} \sum_{l=1}^{L} \sum_{k=1}^{T} \left(\sum_{i=1}^{M} \sum_{j=1}^{N} a_{li} x_{ij} b_{kj} \right)^2 - \sum_{i=1}^{M} \sum_{j=1}^{N} x_{ij} g_{ij} \\ & + \sum_{i=1}^{M} \lambda_i \left(\sum_{j=1}^{N} x_{ij} - 1 \right) - \sum_{i=1}^{M} \sum_{j=1}^{N} \mu_{ij} x_{ij} \end{aligned}$$

with

$$\lambda_i \geq 0 \quad \text{and} \quad \lambda_i \left(\sum_{j=1}^{N} x_{ij} - 1 \right) = 0 \ ,$$
$$\mu_{ij} \geq 0 \quad \text{and} \quad \mu_{ij} x_{ij} = 0 \ .$$

Let us now consider the optimality condition

$$\begin{aligned} 0 = \partial_{x_{ij}} \mathcal{L} &= \tilde{v} x_{ij} \left\| a_{\cdot i} \right\|_2^2 \left\| b_{\cdot j} \right\|_2^2 + \tilde{v} h_{ij} - g_{ij} + \lambda_i - \mu_{ij} \\ \Leftrightarrow \quad \tilde{v} x_{ij} \left\| a_{\cdot i} \right\|_2^2 \left\| b_{\cdot j} \right\|_2^2 &= g_{ij} - \lambda_i + \mu_{ij} - \tilde{v} h_{ij} \ , \end{aligned}$$

where h_{ij} denotes the sum of the mixed terms resulting from the data term, which are independent from x_{ij}, i.e.

$$h_{ij} := \sum_{m \neq i} \sum_{n \neq j} \langle a_{\cdot i}, a_{\cdot m} \rangle \, x_{mn} \, \langle b_{\cdot n}, b_{\cdot j} \rangle + \left\| b_{\cdot j} \right\|_2^2 \sum_{m \neq i} \langle a_{\cdot i}, a_{\cdot m} \rangle \, x_{mj} + \left\| a_{\cdot i} \right\|_2^2 \sum_{n \neq j} x_{in} \, \langle b_{\cdot n}, b_{\cdot j} \rangle \ .$$

Let now $\tilde{v} > 0$ hold and let J_i be the index set for which the entries of the ith row of the solution of (4.17) are nonzero. We show that $g_{ij} > g_{in}$ holds for all $j \in J_i$ and for all $n \notin J_i$. In order to do so, we consider

$$0 = \partial_{x_{ij}} \mathcal{L} - \partial_{x_{in}} \mathcal{L} \qquad \forall \, j \in J_i, \ \forall \, n \notin J_i \ .$$

We have $x_{ij} > 0$ and $\mu_{ij} = 0$, since it is $j \in J_i$. Furthermore, it holds that $x_{in} = 0$ and $\mu_{in} \geq 0$, since we have $n \notin J_i$. Thus we obtain

$$0 \leq \mu_{in} = g_{ij} - g_{in} + \tilde{v} \left(h_{in} - h_{ij} \right) - \tilde{v} x_{ij} \left\| a_{\cdot i} \right\|_2^2 \left\| b_{\cdot j} \right\|_2^2$$

and further

$$0 < \tilde{v} x_{ij} \left\| a_{\cdot i} \right\|_2^2 \left\| b_{\cdot j} \right\|_2^2 \leq g_{ij} - g_{in} + \tilde{v} \left(h_{in} - h_{ij} \right) \ ,$$

due to the fact that $\tilde{v}$, x_{ij}, $\left\| a_{\cdot i} \right\|_2^2$ and $\left\| b_{\cdot j} \right\|_2^2$ are positive. Hence it is left to show that

$$g_{ij} > g_{in} + \tilde{v} d_{ijn} \tag{4.20}$$

holds, where we define $d_{ijn} := h_{ij} - h_{in}$. This statement is obvious for $d_{ijn} \geq 0$. Let now $d_{ijn} < 0$ hold.

For $n \notin J_i$ we assume that $g_{in} = \max\limits_{\nu \in \{1,\ldots,N\}} g_{i\nu}$ holds. In addition let be $g_{in} \geq g_{ij} + 2\tilde{v} |d_{ijn}|$ for all $j \in J_i$. This is always possible, since we can choose $\tilde{v} > 0$ small enough. With

(4.20) we have

$$g_{ij} > g_{in} - \tilde{v}|d_{ijn}|$$

and thus we obtain

$$g_{ij} + \tilde{v}|d_{ijn}| > g_{in} \geq g_{ij} + 2\tilde{v}|d_{ijn}| \qquad \forall\, j \in J_i\ ,$$

which is a contradiction. Hence we finally obtain that

$$g_{ij} > g_{in} \qquad \forall\, j \in J_i,\ \forall\, n \notin J_i$$

has to hold and we see that the solution X has nonzero entries at the same positions like $A^T W B$ has row-maxima, even for $\tilde{v} > 0$ but small enough. Then obviously the same holds for the solution of (4.17). □

4.2. Exact Recovery of Locally 1-Sparse Solutions

In this section we discuss the question of exact recovery for our model.

There already exist several conditions, which provide information about exact reconstruction using linearly independent subdictionaries, see for instance Fuchs (2004) and Tropp (2006b). Unlike the case, where the basis vectors are linearly independent, we consider the operator to be coherent, i.e. the *mutual incoherence parameter* (cf. Juditsky and Nemirovski 2011, p. 3)

$$\mu(B) := \max_{i \neq j} \frac{|\langle b_i, b_j \rangle|}{\|b_i\|_2^2}$$

for b_i, b_j being distant basis vectors, is large. In other words, the vectors are very similar. This is a reasonable assumption for many applications, see for instance Section 6.1.

In Appendix D we gain some understanding of necessary scaling conditions recovering locally 1-sparse solutions considering only one spacial dimension plus one additional dimension using problem (4.2). We learn that if the solution is 1-sparse in one spacial dimension plus the additional dimension, the matrix $B \in \mathbb{R}^{T \times N}$ has to meet the scaling condition

$$\|b_n\|_{\ell^2} = 1 \qquad \text{and} \qquad |\langle b_n, b_m \rangle| \leq 1 \quad \text{for} \quad n \neq m \tag{4.21}$$

in order to recover 1-sparse solutions.

4.2.1. Lagrange Functional and Optimality Conditions

In this subsection we introduce the Lagrange functional and optimality conditions of problem (4.7), which we will need in the further analysis.
We equivalently rewrite problem (4.7) by writing the data constraint for every l and k, i.e.

$$\min_{U\in G,\, v\in\mathbb{R}} \beta\sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij} + v \quad \text{s.t.} \quad \alpha\sum_{j=1}^{N} u_{ij} \le v,\ \sum_{i=1}^{M}\sum_{j=1}^{N} a_{li}u_{ij}b_{kj} = w_{lk} \tag{4.22}$$

with $l \in \{1,\dots,L\}$ and $k \in \{1,\dots,T\}$. For this problem the Lagrange functional reads as follows:

$$\begin{aligned}\mathcal{L}(v,u_{ij};\lambda,\mu,\eta) = {}& \beta\sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij} + v + \sum_{i=1}^{M}\lambda_i\left(\alpha\sum_{j=1}^{N}u_{ij} - v\right) - \sum_{i=1}^{M}\sum_{j=1}^{N}\mu_{ij}u_{ij}\\ &+ \sum_{l=1}^{L}\sum_{k=1}^{T}\eta_{lk}\left(w_{lk} - \sum_{i=1}^{M}\sum_{j=1}^{N} a_{li}u_{ij}b_{kj}\right) ,\end{aligned} \tag{4.23}$$

where λ, μ and η are Lagrange parameters. Now we are able to state the optimality conditions

$$0 = \partial_v\mathcal{L} \quad = 1 - \sum_{i=1}^{M}\lambda_i\ , \tag{OPT1}$$

$$0 = \partial_{u_{ij}}\mathcal{L} = \beta + \alpha\lambda_i - \mu_{ij} - \sum_{l=1}^{L}\sum_{k=1}^{T}\eta_{lk}a_{li}b_{kj}\ , \tag{OPT2}$$

with the complementary conditions (cf. Hiriart-Urruty and Lemaréchal 1993, p. 305-306, Theorem 2.1.4)

$$\lambda_i \ge 0 \quad \text{and} \quad \lambda_i\left(v - \alpha\sum_{j=1}^{N}u_{ij}\right) = 0\ , \tag{4.24}$$

$$\mu_{ij} \ge 0 \quad \text{and} \quad \mu_{ij}u_{ij} = 0\ . \tag{4.25}$$

4.2.2. Scaling Conditions for Exact Recovery of the Constrained Problem

On the basis of this, we examine under which assumptions a 1-sparse solution of the constrained $\ell^{0,\infty}$-problem can be reconstructed exactly by using the constrained $\ell^{1,\infty}$-$\ell^{1,1}$-minimization (4.22).
We will see that the scaling condition (4.21) in a slightly reformulated way is a sufficient condition for exact recovery.

Theorem 4.11 (Recovery of Locally 1-Sparse Data)**.**
Let be $c_i \in \mathbb{R}^+$ and let $J\colon\{1,...,M\} \longrightarrow \{1,...,N\}$ with $i \longmapsto J(i)$ be the function that maps every index i to the index of the corresponding basis vector, where the coefficient is unequal to zero. Let

$$\hat{u}_{ij} = \begin{cases} c_i, & \text{if } \; j = J(i)\,, \\ 0, & \text{if } \; j \neq J(i)\,, \end{cases}$$

be the exact solution of the constrained non-negative $\ell^{0,\infty}$-problem

$$\min_{U \in G} \left(\max_{i \in \{1,...,M\}} \sum_{j=1}^{N} u_{ij}^0 \right) \qquad \text{s. t.} \qquad AUB^T = W\ . \tag{4.26}$$

Let A^T be surjective and let the scaling condition

$$\left\| b_{J(i)} \right\|_2 = 1 \quad \text{and} \quad \left| \left\langle b_{J(i)}, b_j \right\rangle \right| \leq 1 \quad \forall\, j \in \{1, \ldots, N\} \tag{4.27}$$

hold for all $i \in \{1, \ldots, M\}$. Then $\left(\widehat{U},\ \alpha \max\limits_p c_p \right)$ is a solution of (4.22).

Proof.
In order to proof Theorem 4.11, we have to show that there exist Lagrange parameters $\lambda \in \mathbb{R}^M$, $\mu \in \mathbb{R}^{M \times N}$ and $\eta \in \mathbb{R}^{L \times T}$ such that $\widehat{U}$ fulfills the optimality conditions (OPT1) and (OPT2) with respect to the complimentary conditions (4.24) and (4.25).
We choose the Lagrange parameters for all $i \in \{1, \ldots, M\}$ as follows:

$$\lambda_i = \begin{cases} \frac{1}{m}, & \text{if } \; c_i = \frac{v}{\alpha}\,, \\ 0, & \text{if } \; c_i < \frac{v}{\alpha}\,, \end{cases}$$

with $\frac{v}{\alpha} = \max\limits_p c_p$ and m being the number of indices, for which holds $c_i = \frac{v}{\alpha}$,

$$\mu_{ij} = \begin{cases} 0, & \text{if } \; j = J(i)\,, \\ (\alpha\lambda_i + \beta) \left(1 - \sum\limits_{k=1}^{T} b_{kJ(i)} b_{kj} \right), & \text{if } \; j \neq J(i)\,, \end{cases}$$

and η as solution of

$$\sum_{l=1}^{L} a_{li} \eta_{lk} = (\alpha\lambda_i + \beta)\, b_{kJ(i)} \qquad \forall\, i \in \{1, \ldots, M\}\,,\ \forall\, k \in \{1, \ldots, T\}\ . \tag{4.28}$$

Note that (4.28) is solvable, since A^T is surjective.

1. Let us show that (OPT1) and (4.24) hold for $\widehat{U}$:

   a) Obviously we have

   $$\sum_{i=1}^{M} \lambda_i = \sum_{\substack{i \in \{1,\dots,M| \\ c_i = \frac{v}{\alpha}\}}} \frac{1}{m} = m\frac{1}{m} = 1\,.$$

   Thus (OPT1) is fulfilled.

   b) In case that $c_i < \frac{v}{\alpha}$ holds, we see that (4.24) is trivially fulfilled. Hence let be $c_i = \frac{v}{\alpha}$. We consider

   $$\lambda_i \left(v - \alpha \sum_{j=1}^{N} \hat{u}_{ij} \right) = \frac{1}{m}(v - \alpha c_i) = \frac{1}{m}(v - \alpha\frac{v}{\alpha}) = 0$$

   and observe that (4.24) is fulfilled as well.

2. Let us now show that (OPT2) and (4.25) hold for $\widehat{U}$:

   a) In case that $j = J(i)$ holds, we obtain $\hat{u}_{ij} = c_i$ and $\mu_{iJ(i)} = 0$. Thus (4.25) is obviously fulfilled. The other case, i.e. $j \neq J(i)$, yields $\hat{u}_{ij} = 0$ and $\mu_{ij} = (\alpha\lambda_i + \beta)\left(1 - \sum_{k=1}^{T} b_{kJ(i)} b_{kj}\right)$. Since (4.27) has to hold, we obtain $\mu_{ij} \geq 0$ and we observe that in this case (4.25) is fulfilled as well.

   b) Let again be $j = J(i)$. Then we obtain

   $$\alpha\lambda_i + \beta - \sum_{l=1}^{L} \sum_{k=1}^{T} \eta_{lk} a_{li} b_{kJ(i)} = \left(1 - \sum_{k=1}^{T} b_{kJ(i)}^2\right)(\alpha\lambda_i + \beta) = 0$$

   by using the definitions of η and μ and the scaling condition (4.27). In this case (OPT2) is fulfilled.
   Let us now consider $j \neq J(i)$. Then we have

   $$\begin{aligned} &\alpha\lambda_i + \beta - \sum_{l=1}^{L} \sum_{k=1}^{T} \eta_{lk} a_{li} b_{kJ(i)} - \mu_{ij} \\ &= \alpha\lambda_i + \beta - \sum_{l=1}^{L} \sum_{k=1}^{T} \eta_{lk} a_{li} b_{kJ(i)} - (\alpha\lambda_i + \beta)\left(1 - \sum_{k=1}^{T} b_{kJ(i)} b_{kj}\right) \\ &= 0\,, \end{aligned}$$

   where we used the definition of μ. Thus we see that in this case (OPT2) is fulfilled as well.

In summary we see that there exist Lagrange parameters such that $\widehat{U}$ fulfills the optimality conditions and complementary conditions of (4.22). Thus we obtain the assertion. □

All in all we found a condition for exact recovery of solutions of the constrained $\ell^{0,\infty}$-problem, which contain 1-sparse rows, using the constrained problem (4.22) for the reconstruction, i.e. (4.27) has to hold.

Remark 4.6.
We need to assume that A^T is injective in order to solve (4.28). Unfortunately, if A^T is injective, then A is surjective and thus we could easier consider $UB^T = A^\dagger W$ instead, where $A^\dagger$ is the pseudoinverse of A.

Let us now consider an example of the extremest under-determined case, i.e. where we have $L = 1$.

Theorem 4.12.
Let be $\beta = 0$ and $A \in \mathbb{R}^{1\times M}$ with $M > 1$ and $a_i \neq 0$ for every $i \in \{1, \dots, M\}$. Let

$$\hat{u}_{ij} = \begin{cases} c_i, & \text{if } \; j = J(i)\,, \\ 0, & \text{if } \; j \neq J(i)\,, \end{cases}$$

be the exact solution of the nonnegative $\ell^{0,\infty}$-problem (4.26) with $J : \{1, ..., M\} \longrightarrow \{1, ..., N\}$, $i \longmapsto J(i)$ mapping again every index i to the index of the corresponding basis vector, where the coefficient is unequal to zero. Furthermore, let $m \in \{1, \dots, M\}$ be a row-index, where $\widehat{U}$ reaches its maximum, i.e. we have $c_m = \max_p c_p = \frac{v}{\alpha}$.
In case the exact solution $\widehat{U}$ contains a row-vector u_i, which has its nonzero entry at the same position as u_m, i.e. $J(i) = J(m)$, but their entries differ, i.e. $c_i < c_m$, then exact recovery of $\widehat{U}$ using the nonnegative $\ell^{1,\infty}$-problem (4.22) for the reconstruction is not possible.

Proof.
Let us suppose exact recovery were possible. Then there exist a λ, which fulfills (OPT1) and (4.24), a μ, which fulfills (4.25) and an η such that (OPT2) is fulfilled.
Considering the complementary condition (4.24) for $i \in \{1, \dots, M \mid c_i < \max_p c_p\}$, we have

$$0 = \lambda_i \left(v - \alpha \sum_{j=1}^{N} u_{ij} \right) = \lambda_i \left(v - \alpha u_{iJ(i)} \right) = \lambda_i \underbrace{(v - \alpha c_i)}_{\neq 0}\,,$$

since it is $c_i < \frac{v}{\alpha}$. Thus $\lambda_i = 0$ holds for every $i \in \{1, \ldots, M \mid c_i < \max_p c_p\}$. On the other hand with (OPT1) we have

$$1 = \sum_{i=1}^{M} \lambda_i = \sum_{\substack{m \in \{1,\ldots,M \mid \\ c_m = \max_p c_p\}}} \lambda_m \ ,$$

which yields $\lambda_m > 0$ for every $m \in \{1, \ldots, M \mid c_m = \max_p c_p\}$.
Now let us consider (OPT2) for $j = J(i)$ and $j = J(m)$, which then reads as follows:

$$\sum_{k=1}^{T} \eta_k b_{kJ(m)} = \alpha \frac{\lambda_i}{a_i} \qquad \text{and}$$
$$\sum_{k=1}^{T} \eta_k b_{kJ(m)} = \alpha \frac{\lambda_m}{a_m} \ ,$$

since we have $J(i) = J(m)$. Therefore, we obtain

$$a_m \lambda_i = a_i \lambda_m \ .$$

This is a contradiction, since we have $\lambda_i = 0$, $\lambda_m > 0$ and a_i and a_m are unequal to zero. Thus we observe that the 1-sparse $\ell^{0,\infty}$-solution $\hat{u}_{ij}$ can not be the solution of the $\ell^{1,\infty}$-problem (4.22) and in this case exact recovery is not possible. □

Remark 4.7.
In the case of Theorem 4.12 there always exists a solution of (4.22) and (4.26), which has a nonzero element in just one row, i.e. $\widehat{U}$ itself is 1-sparse.

Note that Theorem 4.12 does not state that the reconstructed support is wrong. Hence we could still obtain important information from the nonnegative $\ell^{1,\infty}$-reconstruction. Furthermore, Theorem 4.12 does not apply for the case where we have $\beta > 0$, since in this case we obtain

$$(\alpha\lambda_i + \beta)a_m = (\alpha\lambda_m + \beta)a_i$$

and thus we do not obtain a contradiction in the last step of the proof. Thus Theorem 4.12 suggests the usage of $\beta > 0$.

4.3. Further Improvement of the Results by Including Total Variation

In order to further improve the results and incorporate even more prior knowledge, we include an additional TV-regularization in space (cf. Subsection 2.2.5). The regularization with total variation promotes sharp edges in the resulting images and thus is a reasonable assumption for many applications. Incorporating this knowledge can either be done by additionally minimizing the total variation of the images in every time step, i.e.

$$\min_U \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \alpha \left\| U \right\|_{1,\infty} + \beta \left\| U \right\|_{1,1} + \gamma \sum_{k=1}^{T} \mathrm{TV}\big(Ub_k^T\big) \ , \tag{4.29}$$

or by minimizing the total variation of the coefficient matrices corresponding to every basis vector, i.e.

$$\min_U \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \alpha \left\| U \right\|_{1,\infty} + \beta \left\| U \right\|_{1,1} + \gamma \sum_{j=1}^{N} \mathrm{TV}(u_j) \ . \tag{4.30}$$

The latter can be justified by the fact that sharp edges in the image should result in sharp edges in the visualized coefficient matrices.

In Section 5.2 we extend the previously deduced $\ell^{1,\infty}$-algorithm by including this advanced approach using total variation. We will see in Section 6.3 that problem (4.30) works very well. However, we face the fact that choosing the right regularization parameters for these problems is challenging.

5 Algorithms Promoting Local Sparsity

This chapter contains algorithms for the implementation of different problems proposed in Chapter 4.
In order to solve these problems numerically, we use the alternating direction method of multipliers (ADMM), which can be found for instance in Rockafellar (1976). ADMM traces back to works of Glowinski and Marrocco (1975) and Gabay and Mercier (1976). It was furthermore subject of many other books and papers, including Fortin and Glowinski (1983a), especially its chapters by Fortin and Glowinski (1983b) and Gabay (1983), as well as Glowinski and Tallec (1987), Tseng (1991), Fukushima (1992), Eckstein and Fukushima (1993) and Chen and Teboulle (1994).
However, in this chapter we base upon the deduction by Boyd et al. (2010). We first derive an algorithm for the $\ell^{1,\infty}$-$\ell^{1,1}$-regularized problem. Then we give an extension for an additional TV-regularization. This can be done by either including total variation regularization on the image or on the coefficient matrices.

5.1. Algorithm for $\ell^{1,\infty}$-$\ell^{1,1}$-Regularized Problems

In this section, which is based on our preprint Heins et al. (2014), we propose an algorithm for the solution of $\ell^{1,\infty}$-$\ell^{1,1}$-regularized problems and develop it on the basis of the reformulation proposed in Subsection 4.1.4. Instead of solving problem (4.2), we develop an algorithm for the numerical solution of its reformulated problem (4.16), i.e.

$$\min_U \frac{1}{2}\left\|AUB^T - W\right\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij} \quad \text{s. t.} \quad \sum_{j=1}^{N} u_{ij} \leq \tilde{v},\ \forall\, i,\ u_{ij} \geq 0 \quad \forall\, i, j\ , \tag{4.16}$$

where $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, N\}$. As already indicated at the beginning of this chapter, we use ADMM to solve this problem numerically. For the computation of reasonably simple sub-steps, we split the problem twice. In doing so, we obtain

$$\min_{U,\, Z,\, D} \frac{1}{2}\|AZ - W\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) \qquad \text{s. t.} \qquad D = U,\ Z = UB^T,$$

where we define

$$J(D) := \begin{cases} 0, & \text{if} \quad \sum\limits_{j=1}^{N} d_{ij} \leq \tilde{v} \ \ \forall i, \quad d_{ij} \geq 0 \ \ \forall i,\ j\ , \\ \infty, & \text{else}\ , \end{cases} \tag{5.1}$$

with $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, N\}$. By using the Lagrange functional, which reads as

$$\begin{aligned} \mathcal{L}\left(U, D, Z; \widetilde{P}, \widetilde{Q}\right) = {} & \frac{1}{2}\|AZ - W\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) \\ & + \left\langle \widetilde{P}, U - D\right\rangle + \left\langle \widetilde{Q}, UB^T - Z\right\rangle\ , \end{aligned}$$

we obtain the *unscaled* augmented Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{un}}^{\lambda,\mu}\left(U, D, Z; \widetilde{P}, \widetilde{Q}\right) = {} & \frac{1}{2}\|AZ - W\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) + \left\langle \widetilde{P}, U - D\right\rangle \\ & + \frac{\lambda}{2}\|U - D\|_F^2 + \left\langle \widetilde{Q}, UB^T - Z\right\rangle + \frac{\mu}{2}\left\|UB^T - Z\right\|_F^2 \end{aligned} \tag{5.2}$$

with Lagrange parameters λ, μ and dual variables $\widetilde{P}$ and $\widetilde{Q}$. Since its handling is much easier, we also state the *scaled* augmented Lagrangian, i.e.

$$\begin{aligned} \mathcal{L}_{\text{sc}}^{\lambda,\mu}(U, D, Z; P, Q) = {} & \frac{1}{2}\|AZ - W\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) \\ & + \frac{\lambda}{2}\|U - D + P\|_F^2 + \frac{\mu}{2}\left\|UB^T - Z + Q\right\|_F^2 \end{aligned}$$

with the new scaled dual variables $P := \frac{\widetilde{P}}{\lambda}$ and $Q := \frac{\widetilde{Q}}{\mu}$. By using ADMM (cf. Rockafellar 1976), we obtain the following algorithm:

$$\begin{aligned}
U^{k+1} &= \underset{U}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu}(U, D^k, Z^k; P^k, Q^k)\,,\\
D^{k+1} &= \underset{D}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu}(U^{k+1}, D, Z^k; P^k, Q^k)\,,\\
Z^{k+1} &= \underset{Z}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu}(U^{k+1}, D^{k+1}, Z; P^k, Q^k)\,,\\
P^{k+1} &= P^k - \left(D^{k+1} - U^{k+1}\right)\,,\\
Q^{k+1} &= Q^k - \left(Z^{k+1} - U^{k+1}B^T\right)\,.
\end{aligned}$$

For faster convergence we use a standard extension of ADMM in Subsection 5.1.3, i.e. an adaptive parameter choice as proposed in Boyd et al. (2010, Subsection 3.4.1) with its derivation in Boyd et al. (2010, Section 3.3), which we adapt to our problem. Another advantage of this extension is that the performance becomes less dependent on the initial choice of the penalty parameter. In order to do so, we first propose the optimality conditions.

5.1.1. Optimality Conditions

We obtain the following primal feasibility conditions:

$$0 = \partial_P \mathcal{L} = U - D\,, \tag{5.3}$$
$$0 = \partial_Q \mathcal{L} = UB^T - Z\,. \tag{5.4}$$

Moreover, the following dual feasibility conditions can be derived:

$$0 = \partial_U \mathcal{L} = \lambda P + \mu Q B\,, \tag{5.5}$$
$$0 \in \partial_D \mathcal{L} = \beta \mathbb{1}_{M\times N} + \partial J(D) - \lambda P\,, \tag{5.6}$$
$$0 = \partial_Z \mathcal{L} = A^T\left(AZ - W\right) - \mu Q\,. \tag{5.7}$$

Since U^{k+1} minimizes $\mathcal{L}_{\text{sc}}^{\lambda,\mu}(U, D^k, Z^k; P^k, Q^k)$ by definition, we obtain

$$\begin{aligned}
0 \in \partial_U \mathcal{L}_{\text{sc}}^{\lambda,\mu} &= \lambda\big(U^{k+1} - D^k + P^k\big) + \mu\big(U^{k+1}B^T - Z^k + Q^k\big)B\\
&= \lambda\big(P^k - \big(D^{k+1} - U^{k+1}\big) + D^{k+1} - D^k\big)\\
&\quad + \mu\big(Q^k - \big(Z^{k+1} - U^{k+1}B^T\big) + Z^{k+1} - Z^k\big)B\\
&= \lambda P^{k+1} + \lambda\big(D^{k+1} - D^k\big) + \mu Q^{k+1}B + \mu\big(Z^{k+1} - Z^k\big)B
\end{aligned}$$

by using the definitions of P^{k+1} and Q^{k+1}. This is equivalent to

$$\lambda\big(D^k - D^{k+1}\big) + \mu\big(Z^k - Z^{k+1}\big)B \in \lambda P^{k+1} + \mu Q^{k+1}B \ ,$$

where the right hand side is the first dual feasibility condition (5.5). Therefore,

$$S^{k+1} := \lambda\big(D^k - D^{k+1}\big) + \mu\big(Z^k - Z^{k+1}\big)B \tag{5.8}$$

can be seen as a dual residual for (5.5). Equivalently we consider

$$\begin{aligned} 0 \in \partial_D \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu} &= \beta \mathbb{1}_{m\times n} + \partial J\left(D^{k+1}\right) - \lambda\big(U^{k+1} - D^{k+1} + P^k\big) \\ &= \beta \mathbb{1}_{m\times n} + \partial J\left(D^{k+1}\right) - \lambda P^{k+1} \ , \end{aligned}$$

which means that P^{k+1} and D^{k+1} always satisfy (5.6). The same applies for Q^{k+1} and Z^{k+1}, where we have

$$\begin{aligned} 0 \in \partial_Z \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu} &= A^T\big(AZ^{k+1} - W\big) - \mu\big(U^{k+1}B^T - Z^{k+1} + Q^k\big) \\ &= A^T\big(AZ^{k+1} - W\big) - \mu\big(Q^k - Z^{k+1} + U^{k+1}B^T\big) \\ &= A^T\big(AZ^{k+1} - W\big) - \mu Q^{k+1} \end{aligned}$$

and thus (5.7) is always satisfied as well. In addition we refer to

$$R_1^{k+1} := D^{k+1} - U^{k+1} \qquad \text{and} \tag{5.9}$$

$$R_2^{k+1} := Z^{k+1} - U^{k+1}B^T \tag{5.10}$$

as the primal residuals at iteration $k+1$.

Obviously we obtain five optimality conditions (5.3) - (5.7). We see that (5.6) and (5.7) are always satisfied. The other three, i.e. (5.3) - (5.5), lead to the primal residuals (5.9) and (5.10) and to the dual residual (5.8), which all converge to zero as ADMM proceeds (cf. Boyd et al. 2010, Appendix A, p. 106 et seqq.).

5.1.2. Stopping Criteria

In an analogous way to Boyd et al. (2010, Section 3.3.1) we derive the stopping criteria for the algorithm. As shown in Appendix B the primal and dual residuals can be related to a bound on the objective suboptimality of the current point κ^*. Hence we obtain

$$\begin{aligned} &\frac{1}{2}\|AZ - W\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) - \kappa^* \\ &\leq \left\langle P^k, R_1^k \right\rangle + \left\langle Q^k, R_2^k \right\rangle + \left\langle U^k - U^*, S^k \right\rangle \ . \end{aligned} \tag{5.11}$$

We see that the residuals should be small in order to obtain small objective suboptimality. Since we want to obtain a stopping criterion but the exact solution U^* is unknown, we estimate that $\|U^k - U^*\|_F \leq \nu$ holds. Thus we obtain

$$\begin{aligned} &\frac{1}{2}\|AZ - W\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) - \kappa^* \\ &\leq \|P^k\|_F\|R_1^k\|_F + \|Q^k\|_F\|R_2^k\|_F + \nu\|S^k\|_F \ . \end{aligned}$$

It stands to reason that the primal and dual residual must be small, i.e.

$$\|R_1^k\|_F \leq \varepsilon_1^{\text{pri}} \ , \qquad \|R_2^k\|_F \leq \varepsilon_2^{\text{pri}} \ , \qquad \|S^k\|_F \leq \varepsilon^{\text{dual}} \ ,$$

with tolerances $\varepsilon_{1,2}^{\text{pri}} > 0$ and $\varepsilon^{\text{dual}} > 0$ for the feasibility conditions (5.3 - 5.5), respectively. Boyd et al. suggest that these can be chosen via an absolute and relative criterion, i.e.

$$\begin{aligned} \varepsilon_1^{\text{pri}} &= \sqrt{MN}\ \varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\|U^k\|_F, \|D^k\|_F, 0\right\} \ , \\ \varepsilon_2^{\text{pri}} &= \sqrt{MT}\ \varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\|U^k B^T\|_F, \|Z^k\|_F, 0\right\} \ , \\ \varepsilon^{\text{dual}} &= \sqrt{MN}\ \varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\|\lambda P^k + \mu Q^k B\|_F \ , \end{aligned}$$

where $\varepsilon^{\text{rel}} = 10^{-3}$ or 10^{-4} is a relative tolerance and the absolute tolerance ε^{abs} depends on the scale of the typical variable values. Note that the factors $\sqrt{MN}$ and $\sqrt{MT}$ result from the fact that the Frobenius norms are in $\mathbb{R}^{M\times N}$ and $\mathbb{R}^{M\times T}$, respectively.

5.1.3. Adaptive Parameter Choice

In order to extend the standard ADMM and to improve its convergence rate, we vary the penalty parameters λ^k and μ^k in each iteration as proposed in Boyd et al. (2010, Section 3.4.1). This extension has been analyzed in Rockafellar (1976) in the context of the method of multipliers. There it has been shown that if the penalty parameters go to infinity, superlinear convergence may be reached. If we consider λ and μ to become fixed after a finite number of iterations, the fixed penalty parameter theory still applies, i.e. we obtain convergence of the ADMM.

The following scheme is proposed in He et al. (2000),Wang and Liao (2001) and others and often works well:

$$\lambda^{k+1} = \begin{cases} \tau_1^{\text{incr}} \lambda^k, & \text{if } \|R_1^k\|_F > \eta_1 \|S^k\|_F, \\ \frac{\lambda^k}{\tau_1^{\text{decr}}}, & \text{if } \|S^k\|_F > \eta_1 \|R_1^k\|_F, \\ \lambda^k, & \text{otherwise,} \end{cases} \quad \text{and } P^{k+1} = \begin{cases} \frac{P^k}{\tau_1^{\text{incr}}}, & \text{if } \|R_1^k\|_F > \eta_1 \|S^k\|_F, \\ P^k \tau_1^{\text{decr}}, & \text{if } \|S^k\|_F > \eta_1 \|R_1^k\|_F, \\ P^k, & \text{otherwise,} \end{cases}$$

as well as

$$\mu^{k+1} = \begin{cases} \tau_2^{\text{incr}} \mu^k, & \text{if } \|R_2^k\|_F > \eta_2 \|S^k\|_F, \\ \frac{\mu^k}{\tau_2^{\text{decr}}}, & \text{if } \|S^k\|_F > \eta_2 \|R_1^k\|_F, \\ \mu^k, & \text{otherwise,} \end{cases} \quad \text{and } Q^{k+1} = \begin{cases} \frac{Q^k}{\tau_2^{\text{incr}}}, & \text{if } \|R_2^k\|_F > \eta_2 \|S^k\|_F, \\ Q^k \tau_2^{\text{decr}}, & \text{if } \|S^k\|_F > \eta_2 \|R_2^k\|_F, \\ Q^k, & \text{otherwise,} \end{cases}$$

where $\eta_{1,2} > 1$, $\tau_{1,2}^{\text{incr}} > 1$, $\tau_{1,2}^{\text{decr}} > 1$. Typical choices are $\eta_{1,2} = 10$ and $\tau_{1,2}^{\text{incr}} = \tau_{1,2}^{\text{decr}} = 2$. Note that the dual variables P^k and Q^k only have to be updated in the scaled form. In Figure 5.1 we see a schema of the previously deduced algorithm, which we subsume in the next section.

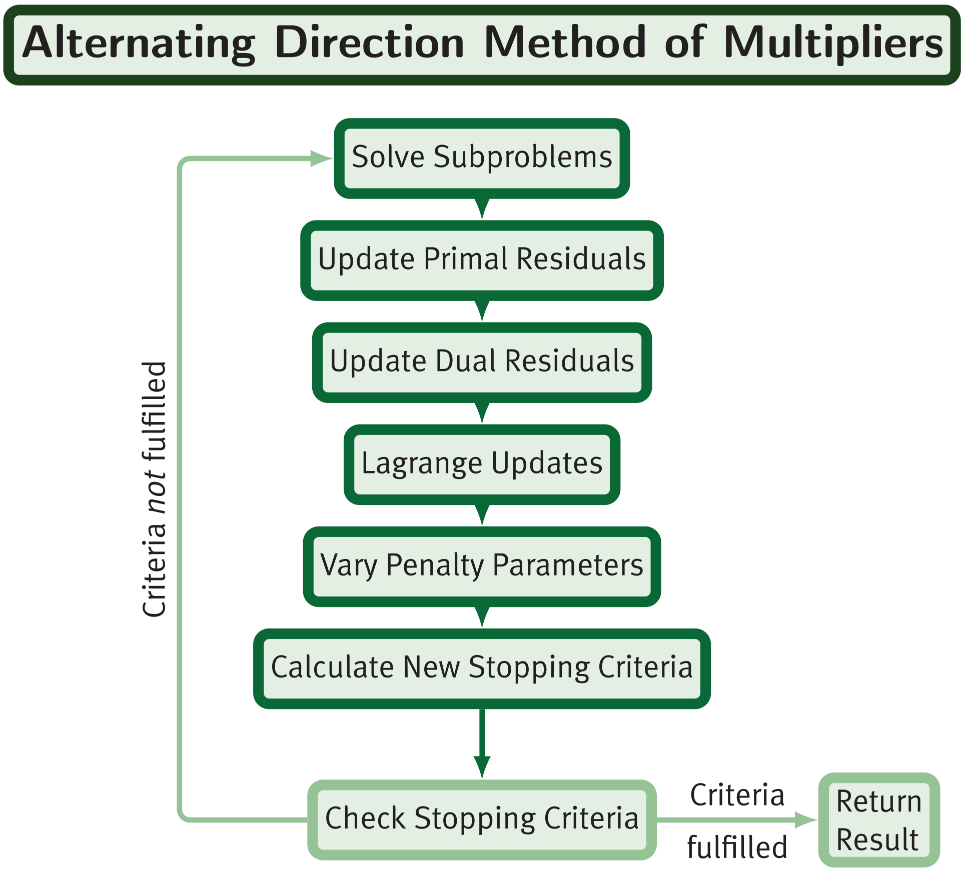


Figure 5.1.: Overview of the alternating direction method of multipliers (ADMM) including the adaptive parameter choice as proposed in Boyd et al. (2010)

5.1.4. Solving the $\ell^{1,\infty}$-$\ell^{1,1}$-Regularized Problem

Let us now state the whole algorithm for the solution of problem (4.16).

Algorithm 1 $\ell^{1,\infty}$-$\ell^{1,1}$-regularized problem via ADMM with double splitting

Parameters: $v, \beta > 0,\ \lambda, \mu > 0,\ A \in \mathbb{R}^{L\times M},\ B \in \mathbb{R}^{T\times N}, W \in \mathbb{R}^{L\times T},\ \eta_{1,2} > 1,$ $\tau_{1,2}^{\text{incr}} > 1,\ \tau_{1,2}^{\text{decr}} > 1,\ \varepsilon^{\text{rel}} = 10^{-3}$ or $10^{-4},\ \varepsilon^{\text{abs}} > 0$

Initialization: $U, Z, D, P, Q, S, R_1, R_2 = 0,\ \varepsilon_1^{\text{pri}} = \sqrt{MN}\,\varepsilon^{\text{abs}}, \varepsilon_2^{\text{pri}} = \sqrt{MT}\,\varepsilon^{\text{abs}},$ $\varepsilon^{\text{dual}} = \sqrt{MN}\,\varepsilon^{\text{abs}}$

while $\|R_1\|_F > \varepsilon_1^{\text{pri}}\ \ \&\ \ \|R_2\|_F > \varepsilon_2^{\text{pri}}\ \ \&\ \ \|S\|_F > \varepsilon^{\text{dual}}$ **do**

  $D^{\text{old}} = D;$

  $Z^{\text{old}} = Z;$

  ▷ Solve Subproblems

  $U = (\lambda\,(D - P) + \mu\,(Z - Q)\,B)\left(\lambda I + \mu B^T B\right)^{-1};$

  $D = \underset{D\in G}{\operatorname{argmin}}\ \frac{\lambda}{2}\,\|D - U + P\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij}\ \text{ s.t. }\ \sum_{j=1}^{N} d_{ij} \le v;$ ▷ see Appendix A

  $Z = \left(A^T A + \mu I\right)^{-1}\left(A^T W + \mu\left(U B^T + Q\right)\right);$

  ▷ Update Primal Residuals

  $R_1 = D - U;$

  $R_2 = Z - U B^T;$

  ▷ Update Dual Residual

  $S = \lambda\left(D^{\text{old}} - D\right) + \mu\left(Z^{\text{old}} - Z\right)B;$

  ▷ Lagrange Updates

  $P = P - (D - U);$

  $Q = Q - \left(Z - U B^T\right);$

  ▷ Varying Penalty/Lagrange Parameters

  if $\|R_1\|_F > \eta_1\|S\|_F$ **then**

    $\lambda = \lambda\tau_1^{\text{incr}};$

    $P = \frac{P}{\tau_1^{\text{incr}}};$

  else if $\|S\|_F > \eta_1\|R_1\|_F$ **then**

    $\lambda = \frac{\lambda}{\tau_1^{\text{decr}}};$

    $P = P\tau_1^{\text{decr}};$

  end if

  if $\|R_2\|_F > \eta_2\|S\|_F$ **then**

    $\mu = \mu\tau_2^{\text{incr}};$

    $Q = \frac{Q}{\tau_2^{\text{incr}}};$

  else if $\|S\|_F > \eta_2\|R_2\|_F$ **then**

    $\mu = \frac{\mu}{\tau_2^{\text{decr}}};$

```
        Q = Q τ_2^decr;
    end if
                                                              ▷ Stopping Criteria
    ε_1^pri = √(MN) ε^abs + ε^rel max{‖U^k‖_F, ‖D^k‖_F, 0};
    ε_2^pri = √(MT) ε^abs + ε^rel max{‖U^k B^T‖_F, ‖Z^k‖_F, 0};
    ε^dual = √(MN) ε^abs + ε^rel ‖λP^k + μQ^k B‖_F;
end while
return U                                                      ▷ Solution of (4.16)
```

5.2. Algorithms for $\ell^{1,\infty}$-$\ell^{1,1}$-TV-Regularized Problems

In Section 5.1 we proposed an algorithm for the solution of $\ell^{1,\infty}$-$\ell^{1,1}$-regularized problems. As we will see in Section 6.2 the results are very nice, however, they still offer room for improvement. In order to do so, we include further a-priori knowledge, i.e. we suppose the images to have sharp edges for every time step. This is a reasonable assumption especially in imaging and image processing. On this account we included an additional TV-regularization in Section 4.3.
In this section we first deduce an algorithm for solving $\ell^{1,\infty}$-$\ell^{1,1}$-regularized problems including an additional total variation minimization on the images in every time step. However, since sharp edges in the images are reflected in their corresponding coefficient matrices, they usually result in sharp edges of the visualized coefficient matrices as well. Thus we also want to propose an algorithm for the solution of $\ell^{1,\infty}$-$\ell^{1,1}$-regularized problems including TV-regularization on the coefficient matrices.

5.2.1. Additional TV-Regularization on the Image

Let us now include an additional TV-regularization on the image for every time step independently. Similarly to Section 5.1 we deduce the algorithm on the basis of the reformulation proposed in Subsection 4.1.4, i.e. we want to solve

$$
\begin{aligned}
&\min_U \frac{1}{2}\left\|AUB^T - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij} + \gamma \sum_{t=1}^{T} \left\|\nabla\left(Ub_t^T\right)\right\|_1 \\
&\text{s. t.} \quad \sum_{j=1}^{N} u_{ij} \le \tilde{v} \quad \forall\, i \quad \text{and} \quad u_{ij} \ge 0 \quad \forall\, i,\ j\ ,
\end{aligned}
\tag{5.12}
$$

where $i \in \{1,\dots,M\}$ and $j \in \{1,\dots,N\}$. Analogously to the previous section we use ADMM for the solution of the problem. Note that we have to include TV-regularization in (5.12) for each time step $t \in \{1,\dots,T\}$ separately, since we want total variation minimization to be active in the spatial dimension but not in time. Thus we include

the sum over all TV-regularizations for every time step.
In order to compute reasonably simple sub-steps, we now have to split the problem three times, i.e.

$$\min_{U,\,D,\,Z,\,G} \frac{1}{2}\|AZ-W\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{t=1}^{T} \|g_t\|_1 + J(D)$$
$$\text{s. t. } \; D = U,\; Z = UB^T,\; g_t = \nabla z_t \quad \forall\, t \in \{1,\ldots,T\},$$

with J as defined in (5.1). The Lagrange functional for this case reads as follows:

$$\begin{aligned} \mathcal{L}\left(U,D,Z,G;\widetilde{P},\widetilde{Q},\widetilde{R}\right) = & \frac{1}{2}\|AZ-W\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{t=1}^{T} \|g_t\|_1 \\ & + J(D) + \left\langle \widetilde{P}, U-D \right\rangle + \left\langle \widetilde{Q}, UB^T - Z \right\rangle \\ & + \sum_{t=1}^{T} \langle \tilde{r}_t, \nabla z_t - g_t \rangle \;. \end{aligned} \tag{5.13}$$

Then we obtain the *unscaled* augmented Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{un}}^{\lambda,\mu,\eta}\left(U,D,Z,G;\widetilde{P},\widetilde{Q},\widetilde{R}\right) = & \frac{1}{2}\|AZ-W\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{t=1}^{T} \|g_t\|_1 + J(D) \\ & + \left\langle \widetilde{P}, U-D \right\rangle + \frac{\lambda}{2}\|U-D\|_F^2 \\ & + \left\langle \widetilde{Q}, UB^T - Z \right\rangle + \frac{\mu}{2}\left\|UB^T - Z\right\|_F^2 \\ & + \sum_{t=1}^{T} \langle \tilde{r}_t, \nabla z_t - g_t \rangle + \sum_{t=1}^{T} \frac{\eta_t}{2}\|\nabla z_t - g_t\|_2^2 \end{aligned}$$

with Lagrange parameters λ, μ, η and dual variables $\widetilde{P}$, $\widetilde{Q}$ and $\widetilde{R}$.
We have seen in Section 5.1 that the usage of the *scaled* augmented Lagrangian offers the advantage of easier handling. Therefore, we will use this approach again, i.e. we consider

$$\begin{aligned} \mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta}(U,D,Z,G;P,Q,R) = & \frac{1}{2}\|AZ-W\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{t=1}^{T} \|g_t\|_1 + J(D) \\ & + \frac{\lambda}{2}\|U-D+P\|_F^2 + \frac{\mu}{2}\left\|UB^T - Z + Q\right\|_F^2 \\ & + \sum_{t=1}^{T} \frac{\eta_t}{2}\|\nabla z_t - g_t + r_t\|_2^2 \end{aligned}$$

with the scaled dual variables $P := \frac{\widetilde{P}}{\lambda}$, $Q := \frac{\widetilde{Q}}{\mu}$ and $r_t := \frac{\tilde{r}_t}{\eta_t}$ for every $t \in \{1, \dots, T\}$. We now obtain an algorithm, which is a little bit more complex, i.e.

$$\begin{aligned}
U^{k+1} &= \underset{U}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta}(U, D^k, Z^k, G^k; P^k, Q^k, R^k)\,, \\
D^{k+1} &= \underset{D}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta}(U^{k+1}, D, Z^k, G^k; P^k, Q^k, R^k)\,, \\
z_t^{k+1} &= \underset{z_t}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta_t}(U^{k+1}, D^{k+1}, z_t, g_t^k; P^k, Q^k, r_t^k) \qquad \forall\, t \in \{1, \dots, T\}\,, \\
g_t^{k+1} &= \underset{g_t}{\operatorname{argmin}}\ \mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta_t}(U^{k+1}, D^{k+1}, z_t^{k+1}, g_t; P^k, Q^k, r_t^k) \qquad \forall\, t \in \{1, \dots, T\}\,, \\
P^{k+1} &= P^k - \left(D^{k+1} - U^{k+1}\right)\,, \\
Q^{k+1} &= Q^k - \left(Z^{k+1} - U^{k+1} B^T\right)\,, \\
r_t^{k+1} &= r_t^k - \left(g_t^{k+1} - \nabla z_t^{k+1}\right) \qquad \forall\, t \in \{1, \dots, T\}\,.
\end{aligned}$$

In analogy to the previous section we use again the adaptive parameter choice as proposed in Boyd et al. (2010, Subsection 3.4.1). For this purpose we start once more by proposing the optimality conditions.

Optimality Conditions

By deriving the partial derivatives of the Lagrange functional (5.13), we obtain the following primal feasibility conditions:

$$0 = \partial_P \mathcal{L} = U - D\,, \tag{5.14}$$

$$0 = \partial_Q \mathcal{L} = U B^T - Z\,, \tag{5.15}$$

$$0 = \partial_{r_t} \mathcal{L} = \nabla z_t - g_t \qquad \forall\, t \in \{1, \dots, T\}\,, \tag{5.16}$$

and the dual feasibility conditions

$$0 = \partial_U \mathcal{L} = \lambda P + \mu Q B\,, \tag{5.17}$$

$$0 \in \partial_D \mathcal{L} = \beta \mathbb{1}_{M \times N} + \partial J(D) - \lambda P\,, \tag{5.18}$$

$$0 = \partial_{z_t} \mathcal{L} = A^T (A z_t - w_t) B - \mu q_t - \eta_t \nabla \cdot r_t \qquad \forall\, t \in \{1, \dots, T\}\,, \tag{5.19}$$

$$0 \in \partial_{g_t} \mathcal{L} = \gamma \partial \left\| g_t \right\|_1 - \eta_t r_t \qquad \forall\, t \in \{1, \dots, T\}\,. \tag{5.20}$$

Since U^{k+1} minimizes $\mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta}(U, D^k, Z^k, G^k; P^k, Q^k, R^k)$ by definition, we obtain in analogy to Section 5.1

$$0 \in \partial_U \mathcal{L}_{\text{sc}}^{\lambda,\mu,\eta} = \lambda P^{k+1} + \lambda\left(D^{k+1} - D^k\right) + \mu Q^{k+1} B + \mu\left(Z^{k+1} - Z^k\right) B$$

by using the definitions of P^{k+1} and Q^{k+1}. This is equivalent to

$$\lambda\big(D^k - D^{k+1}\big) + \mu\big(Z^k - Z^{k+1}\big)B \in \lambda P^{k+1} + \mu Q^{k+1}B \ ,$$

where the right hand side of this statement is the first dual feasibility condition (5.17). Thus we see that

$$S^{k+1} := \lambda\big(D^k - D^{k+1}\big) + \mu\big(Z^k - Z^{k+1}\big)B \tag{5.21}$$

serves as a first dual residual for (5.17). In an analogous manner we consider

$$0 \in \partial_D \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta} = \beta \mathbb{1}_{M\times N} + \partial J\left(D^{k+1}\right) - \lambda P^{k+1}$$

and observe that P^{k+1} and D^{k+1} always satisfy (5.18).
Computing the derivative for z_t, i.e.

$$0 \in \partial_{z_t} \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta_t} = A^T\big(Az_t^{k+1} - w_t\big) - \mu q_t^{k+1} - \eta_t \nabla\cdot r_t^{k+1} - \eta_t \nabla\cdot \left(g_t^{k+1} - g_t^k\right) ,$$

which is equivalent to

$$\eta_t \nabla\cdot \left(g_t^{k+1} - g_t^k\right) \in A^T\big(Az_t^{k+1} - w_t\big) - \mu q_t^{k+1} - \eta_t \nabla\cdot r_t^{k+1} \ ,$$

we see that the right hand side matches the third dual feasibility condition (5.33). Thus we obtain another dual residual, i.e.

$$c_t^{k+1} := \eta_t \nabla\cdot \left(g_t^{k+1} - g_t^k\right) , \tag{5.22}$$

to which we refer as second dual residual.
Last but not least we see that r_t^{k+1} and g_t^{k+1} always satisfy (5.20) for all $t \in \{1, \dots, T\}$ by computing

$$0 \in \partial_{g_t} \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta_t} = \gamma \partial \left\|g_t^{k+1}\right\|_1 - \eta_t r_t^{k+1} \ .$$

Additionally we refer to

$$X_1^{k+1} := D^{k+1} - U^{k+1} \ , \tag{5.23}$$

$$X_2^{k+1} := Z^{k+1} - U^{k+1}B^T \qquad \text{and} \tag{5.24}$$

$$y_t^{k+1} := g_t^{k+1} - \nabla z_t^{k+1} \qquad \forall\, t \in \{1, \dots, T\} \tag{5.25}$$

as the primal residuals at iteration $k+1$.

With (5.14) - (5.20) we now have seven optimality conditions, where we have seen that (5.18) and (5.20) are always satisfied. The remaining five optimality conditions (5.14) - (5.17) and (5.19) yield the primal residuals (5.23) - (5.25) and the dual residuals (5.21) and (5.22), which converge to zero as ADMM proceeds (cf. BOYD et al. 2010, Appendix A, p. 106 et seqq.).

Stopping Criteria

Once more we derive the stopping criteria in analogy to BOYD et al. (2010, Section 3.3.1).

Similarly to Subsection 5.1.2 the primal and dual residuals can be related to a bound on the objective suboptimality of the current point κ^*. By estimating $\|U^k - U^*\|_F \leq \nu$ we obtain

$$\begin{aligned}&\frac{1}{2}\|AZ - W\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma\sum_{t=1}^{T}\|g_t\|_1 + J(D) - \kappa^* \\ \leq\; &\|P^k\|_F\|X_1^k\|_F + \|Q^k\|_F\|X_2^k\|_F + \sum_{t=1}^{T}\left\|r_t^k\right\|_2\left\|y_t^k\right\|_2 + \nu\left\|S^k\right\|_F + \nu\sum_{t=1}^{T}\left\|c_t^k\right\|_2 \;.\end{aligned}$$

Thus we see that the primal and dual residuals must be small, i.e.

$$\|X_1^k\|_F \leq \varepsilon_1^{\text{pri}}, \quad \|X_2^k\|_F \leq \varepsilon_2^{\text{pri}}, \quad \|y_t^k\|_2 \leq \zeta_t^{\text{pri}}\;\forall\, t, \quad \|S^k\|_F \leq \varepsilon^{\text{dual}}, \quad \left\|c_t^k\right\|_2 \leq \zeta_t^{\text{dual}}\;\forall\, t,$$

with positive tolerances $\varepsilon_1^{\text{pri}}$, $\varepsilon_2^{\text{pri}}$, ζ_t^{pri}, $\varepsilon^{\text{dual}}$ and ζ_t^{dual} for the feasibility conditions (5.28 - 5.31) and (5.33), respectively. We choose these tolerances in the same way as it was done in BOYD et al. (2010) via an absolute and relative criterion, i.e.

$$\begin{aligned}\varepsilon_1^{\text{pri}} &= \sqrt{MN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\max\left\{\|U^k\|_F, \|D^k\|_F, 0\right\}\;, \\ \varepsilon_2^{\text{pri}} &= \sqrt{MT}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\max\left\{\|U^kB^T\|_F, \|Z^k\|_F, 0\right\}\;, \\ \zeta_t^{\text{pri}} &= \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\max\left\{\|\nabla z_t^k\|_2, \|g_t^k\|_2, 0\right\} && \forall\, t\in\{1,\dots,T\}\,, \\ \varepsilon^{\text{dual}} &= \sqrt{MN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\|\lambda P^k + \mu Q^k B\|_2\,, \\ \zeta_t^{\text{dual}} &= \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\left\|\mu q_t^k + \eta_t\nabla\cdot r_t^k\right\|_2 && \forall\, t\in\{1,\dots,T\}\,,\end{aligned}$$

where $\varepsilon^{\text{rel}} = 10^{-3}$ or 10^{-4} is a relative tolerance. Moreover, the absolute tolerance ε^{abs} depends on the scale of the typical variable values. Note that the square roots once more result from the dimensions of the matrices used in the respective second part of the sums.

Adaptive Parameter Choice

Once again we use the adaptive parameter choice proposed in Boyd et al. (2010, Section 3.4.1) and vary the penalty parameters λ^k, μ^k and η^k in each iteration.
Similar to the previous section we make use of the following scheme:

$$\lambda^{k+1} = \begin{cases} \omega^{\text{incr}}\lambda^k, & \text{if } \|X_1^k\|_F > \delta\|S^k\|_F, \\ \frac{\lambda^k}{\omega^{\text{decr}}}, & \text{if } \|S^k\|_F > \delta\|X_1^k\|_F, \\ \lambda^k, & \text{otherwise,} \end{cases} \text{ and } P^{k+1} = \begin{cases} \frac{P^k}{\omega^{\text{incr}}}, & \text{if } \|X_1^k\|_F > \delta\|S^k\|_F, \\ P^k\omega^{\text{decr}}, & \text{if } \|S^k\|_F > \delta\|X_2^k\|_F, \\ P^k, & \text{otherwise,} \end{cases}$$

$$\mu^{k+1} = \begin{cases} \tau^{\text{incr}}\mu^k, & \text{if } \|X_2^k\|_F > \vartheta\|S^k\|_F, \\ \frac{\mu^k}{\tau^{\text{decr}}}, & \text{if } \|S^k\|_F > \vartheta\|X_2^k\|_F, \\ \mu^k, & \text{otherwise,} \end{cases} \text{ and } Q^{k+1} = \begin{cases} \frac{Q^k}{\tau^{\text{incr}}}, & \text{if } \|X_2^k\|_F > \vartheta\|S^k\|_F, \\ Q^k\tau^{\text{decr}}, & \text{if } \|S^k\|_F > \vartheta\|X_2^k\|_F, \\ Q^k, & \text{otherwise,} \end{cases}$$

$$\eta_t^{k+1} = \begin{cases} \sigma^{\text{incr}}\eta_t^k, & \text{if } \|y_t^k\|_2 > \rho\|c_t^k\|_2, \\ \frac{\eta_t^k}{\sigma^{\text{decr}}}, & \text{if } \|c_t^k\|_2 > \rho\|y_t^k\|_2, \\ \eta_t^k, & \text{otherwise,} \end{cases} \text{ and } r_t^{k+1} = \begin{cases} \frac{r_t^k}{\sigma^{\text{incr}}}, & \text{if } \|y_t^k\|_2 > \rho\|c_t^k\|_2, \\ r_t^k\sigma^{\text{decr}}, & \text{if } \|c_t^k\|_2 > \rho\|y_t^k\|_2, \\ r_t^k, & \text{otherwise,} \end{cases}$$

for all $t \in \{1, \dots, T\}$, where $\delta,\ \vartheta,\ \rho > 1$ and $\omega,\ \tau,\ \sigma > 1$. Typical choices are $\delta,\ \vartheta,\ \rho = 10$ and $\omega,\ \tau,\ \sigma = 2$. Note that once more the dual variables P^k, Q^k and R^k have to be updated, since we are using the scaled form.

Solving the $\ell^{1,\infty}$-$\ell^{1,1}$-TV-Regularized Problem Including TV on the Image

Algorithm 2 $\ell^{1,\infty}$-$\ell^{1,1}$-TV$\big(UB^T\big)$-regularized problem via ADMM with triple splitting

Parameters: $\upsilon > 0,\ \beta > 0,\ \gamma > 0,\ A \in \mathbb{R}^{L\times M},\ B \in \mathbb{R}^{T\times N}, W \in \mathbb{R}^{L\times T},$
$\delta, \vartheta, \rho > 1,\ \omega, \tau, \sigma > 1,\ \varepsilon^{\text{rel}} = 10^{-3}$ or $10^{-4},\ \varepsilon^{\text{abs}} > 0$

Initialization: $U, D, Z, G, P, Q, R, S, C, X_1, X_2, Y = 0,\ \varepsilon_1^{\text{pri}} = \sqrt{MN}\varepsilon^{\text{abs}},$
$\varepsilon_2^{\text{pri}} = \sqrt{MT}\varepsilon^{\text{abs}}, \zeta_t^{\text{pri}} = \sqrt{M}\varepsilon^{\text{abs}}, \varepsilon^{\text{dual}} = \sqrt{MN}\varepsilon^{\text{abs}}, \zeta_t^{\text{dual}} = \sqrt{M}\varepsilon^{\text{abs}}$

while $\|X_1\|_F > \varepsilon_1^{\text{pri}}\ \&\ \|X_2\|_F > \varepsilon_2^{\text{pri}}\ \&\ \|y_t\|_2 > \zeta_t^{\text{pri}}\ \forall t\ \&\ \|S\|_F > \varepsilon^{\text{dual}}\ \&\ \|c_t\|_2 > \zeta_t^{\text{dual}}\ \forall t$ **do**

$D^{\text{old}} = D;$

$Z^{\text{old}} = Z;$

$G^{\text{old}} = G;$

▷ Solve Subproblems I

$U = (\lambda(P - D) + \mu(Q - Z)B)\left(\lambda I + \mu B^T B\right)^{-1};$

$D = \underset{D\in G}{\operatorname{argmin}} \frac{\lambda}{2}\|D - U + P\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij}$ s.t. $\sum_{j=1}^{N} d_{ij} \leq \upsilon;$ ▷ see Appendix A

▷ Update Primal Residual I
$X_1 = D - U;$
▷ Lagrange Update I
$P = P - (D - U);$
▷ Stopping Criteria I
$\varepsilon_1^{\text{pri}} = \sqrt{MN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\|U\|_F, \|D\|_F, 0\};$
for all $t \in \{1, \dots, T\}$ **do**
▷ Solve Subproblem II
  $z_t = \left(A^T A + \mu I - \eta_t \Delta\right)^{-1} \left(A^T w_t + \mu\left(q_t + U b_t^T\right) + \eta_t \nabla \cdot (r_t - g_t)\right);$
  if $\eta_t \neq 0$ **then**
    $g_t = \operatorname{sign}(\nabla z_t + r_t) \max\left(|\nabla z_t + r_t| - \frac{\gamma}{\eta_t},\ 0\right);$
  end if
▷ Update Primal Residual II
  $y_t = g_t - \nabla z_t;$
▷ Lagrange Update II
  $r_t = r_t - (g_t - \nabla z_t);$
▷ Update Dual Residual I
  $c_t = \eta_t \nabla \cdot (g_t - g_t^{\text{old}});$
▷ Stopping Criteria II
  $\zeta_t^{\text{pri}} = \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\|\nabla z_t\|_2, \|g_t\|_2, 0\};$
end for
▷ Update Primal Residual III
$X_2 = Z - U B^T;$
▷ Lagrange Update III
$Q = Q - \left(Z - U B^T\right);$
▷ Update Dual Residual II
$S = \lambda(D^{\text{old}} - D) + \mu(Z^{\text{old}} - Z)B;$
▷ Stopping Criteria III
$\varepsilon_2^{\text{pri}} = \sqrt{MT}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\|U B^T\|_F, \|Z\|_F, 0\right\};$
▷ Varying Penalty/Lagrange Parameters I
if $\|X_1\|_F > \delta \|S\|_F$ **then**
  $\lambda = \lambda \omega^{\text{incr}};$
  $P = \frac{P}{\omega^{\text{incr}}};$
else if $\|S\|_F > \delta \|X_1\|_F$ **then**
  $\lambda = \frac{\lambda}{\omega^{\text{decr}}};$
  $P = P \omega^{\text{decr}};$
end if
if $\|X_2\|_F > \vartheta \|S\|_F$ **then**
  $\mu = \mu \tau^{\text{incr}};$
  $Q = \frac{Q}{\tau^{\text{incr}}};$
else if $\|S\|_F > \vartheta \|X_2\|_F$ **then**
  $\mu = \frac{\mu}{\tau^{\text{decr}}};$
  $Q = Q \tau^{\text{decr}};$
end if
▷ Stopping Criteria IV
$\varepsilon^{\text{dual}} = \sqrt{MN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \|\lambda P + \mu Q B\|_F;$

for all $t \in \{1, \dots, T\}$ **do**
▷ Varying Penalty/Lagrange Parameters II
  if $\|y_t\|_2 > \rho \|c_t\|_2$ **then**
    $\eta_t = \eta_t \sigma^{\text{incr}}$;
    $r_t = \frac{r_t}{\sigma^{\text{incr}}}$;
  else if $\|c_t\|_2 > \rho \|y_t\|_2$ **then**
    $\eta_t = \frac{\eta_t}{\sigma^{\text{decr}}}$;
    $r_t = r_t \sigma^{\text{decr}}$;
  end if
▷ Stopping Criteria V
  $\zeta_t^{\text{dual}} = \sqrt{M}\, \varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \|\mu q_t + \eta_t \nabla \cdot r_t\|_2$;
end for
end while
return U ▷ Solution of (5.12)

5.2.2. Additional TV-Regularization on the Coefficient Matrices

Let us now consider TV-minimization on the coefficient vectors for every basis function. As before we will develop an algorithm using ADMM. However, we will shorten this subsection a little bit, since the derivation of the algorithm works analogously to the derivations before.

We consider again the reformulation from Subsection 4.1.4, i.e.

$$\begin{aligned} \min_U \ & \frac{1}{2} \left\| AUB^T - W \right\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij} + \gamma \sum_{j=1}^{N} \|\nabla u_j\|_1 \\ \text{s. t.} \ & \sum_{j=1}^{N} u_{ij} \leq \tilde{v} \quad \forall\, i \quad \text{and} \quad u_{ij} \geq 0 \quad \forall\, i,\ j\ , \end{aligned} \tag{5.26}$$

where $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, N\}$. Note that we have to include TV-regularization in (5.26) for each basis vector $j \in \{1, \dots, N\}$ separately in order to use total variation only on the coefficient matrices for every basis vector and not as well on the additional dimension. Thus we include the sum over all TV-regularizations for every basis vector.

In order to compute reasonably simple sub-steps, we split the problem three times. However, we cannot split the way we have done it before, since that would leave us with an optimality condition, in which the variable has an operator on both sides. Hence it would not be easily possible to separate the variable from the operators. On account of this we propose another splitting. Instead of splitting $Z = UB^T$ in the data fidelity term, we rather split $Z = AU$. This yields the following optimization problem:

$$\min_{U,\,D,\,Z,\,G} \frac{1}{2}\left\|ZB^T - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{j=1}^{N} \|g_j\|_1 + J(D)$$
$$\text{s. t.}\ \ D = U,\ Z = AU,\ g_j = \nabla u_j \quad \forall\, j \in \{1,\dots,N\},$$

where we use J as defined in (5.1). Then we have the Lagrange functional

$$\begin{aligned}\mathcal{L}\left(U,D,Z,G;\widetilde{P},\widetilde{Q},\widetilde{R}\right) = {} & \frac{1}{2}\left\|ZB^T - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{j=1}^{N} \|g_j\|_1 \\ & + J(D) + \left\langle \widetilde{P}, U - D\right\rangle + \left\langle \widetilde{Q}, AU - Z\right\rangle \\ & + \sum_{j=1}^{N} \langle \tilde{r}_j, \nabla u_j - g_j\rangle\ ,\end{aligned} \tag{5.27}$$

the *unscaled* augmented Lagrangian

$$\begin{aligned}\mathcal{L}_{\mathrm{un}}^{\lambda,\mu,\eta}\left(U,D,Z,G;\widetilde{P},\widetilde{Q},\widetilde{R}\right) = {} & \frac{1}{2}\left\|ZB^T - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{j=1}^{N} \|g_j\|_1 + J(D) \\ & + \left\langle \widetilde{P}, U - D\right\rangle + \frac{\lambda}{2}\|U - D\|_F^2 \\ & + \left\langle \widetilde{Q}, AU - Z\right\rangle + \frac{\mu}{2}\|AU - Z\|_F^2 \\ & + \sum_{j=1}^{N} \langle \tilde{r}_j, \nabla u_j - g_j\rangle + \sum_{j=1}^{N} \frac{\eta_j}{2}\|\nabla u_j - g_j\|_2^2\ ,\end{aligned}$$

with Lagrange parameters λ, μ, η and dual variables $\widetilde{P}$, $\widetilde{Q}$ and $\widetilde{R}$ and the *scaled* augmented Lagrangian

$$\begin{aligned}\mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta}(U,D,Z,G;P,Q,R) = {} & \frac{1}{2}\left\|ZB^T - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{j=1}^{N} \|g_j\|_1 + J(D) \\ & + \frac{\lambda}{2}\|U - D + P\|_F^2 + \frac{\mu}{2}\|AU - Z + Q\|_F^2 \\ & + \sum_{j=1}^{N} \frac{\eta_j}{2}\|\nabla u_j - g_j + r_j\|_2^2\ ,\end{aligned}$$

with the scaled dual variables $P := \frac{\widetilde{P}}{\lambda}$, $Q := \frac{\widetilde{Q}}{\mu}$ and $r_j := \frac{\tilde{r}_j}{\eta_j}$ for every $j \in \{1,\dots,N\}$.

We now obtain the algorithm as follows:

$$
\begin{aligned}
u_j^{k+1} &= \operatorname*{argmin}_{u_j} \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta_j}(u_j, D^k, Z^k, g_j^k; P^k, Q^k, r_j^k) && \forall\, j \in \{1,\dots,N\}\,,\\
D^{k+1} &= \operatorname*{argmin}_{D} \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta}(U^{k+1}, D, Z^k, G^k; P^k, Q^k, R^k)\,,\\
Z^{k+1} &= \operatorname*{argmin}_{Z} \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta}(U^{k+1}, D^{k+1}, Z, G^k; P^k, Q^k, R^k)\,,\\
g_j^{k+1} &= \operatorname*{argmin}_{g_j} \mathcal{L}_{\mathrm{sc}}^{\lambda,\mu,\eta_j}(u_j^{k+1}, D^{k+1}, Z^{k+1}, g_j; P^k, Q^k, r_j^k) && \forall\, j \in \{1,\dots,N\}\,,\\
P^{k+1} &= P^k - \left(D^{k+1} - U^{k+1}\right)\,,\\
Q^{k+1} &= Q^k - \left(Z^{k+1} - U^{k+1}B^T\right)\,,\\
r_j^{k+1} &= r_j^k - \left(g_j^{k+1} - \nabla u_j^{k+1}\right) && \forall\, j \in \{1,\dots,N\}\,.
\end{aligned}
$$

We use the adaptive parameter choice from Boyd et al. (2010, Subsection 3.4.1) just like in the previous section. For this purpose we start once more by stating the optimality conditions.

Optimality Conditions

By deriving the partial derivatives of the Lagrange functional (5.27), we obtain the following primal feasibility conditions:

$$0 = \partial_P \mathcal{L} = U - D\,, \tag{5.28}$$

$$0 = \partial_Q \mathcal{L} = AU - Z\,, \tag{5.29}$$

$$0 = \partial_{r_j} \mathcal{L} = \nabla u_j - g_j \qquad \forall\, j \in \{1,\dots,N\}\,. \tag{5.30}$$

In addition we obtain the dual feasibility conditions

$$0 = \partial_{u_j} \mathcal{L} = \lambda p_j + \mu A^T q_j - \eta_j \nabla\cdot r_j \qquad \forall\, j \in \{1,\dots,N\}\,, \tag{5.31}$$

$$0 \in \partial_D \mathcal{L} = \beta \mathbb{1}_{M\times N} + \partial J(D) - \lambda P\,, \tag{5.32}$$

$$0 = \partial_Z \mathcal{L} = \left(ZB^T - W\right)B - \mu Q\,, \tag{5.33}$$

$$0 \in \partial_{g_j} \mathcal{L} = \gamma \partial \left\|g_j\right\|_1 - \eta_j r_j \qquad \forall\, j \in \{1,\dots,N\}\,. \tag{5.34}$$

In analogy to Section 5.1 and the previous subsection we obtain

$$\lambda\big(d_j^k - d_j^{k+1}\big) + \mu A^T\big(z_j^k - z_j^{k+1}\big) - \eta_j \nabla\cdot\big(g_j^k - g_j^{k+1}\big) \in \lambda p_j^{k+1} + \mu A^T q_j^{k+1} - \eta_j \nabla\cdot r_j^{k+1}$$

for all $j \in \{1,\ldots,N\}$, where the right hand side of this expression is the first dual feasibility condition (5.31). Thus we obtain the dual residual

$$s_j^{k+1} := \lambda\big(d_j^k - d_j^{k+1}\big) + \mu A^T\big(z_j^k - z_j^{k+1}\big) - \eta_j \nabla\cdot\big(g_j^k - g_j^{k+1}\big) \quad \forall\, j \in \{1,\ldots,N\} \tag{5.35}$$

for (5.31). In addition P^{k+1} and D^{k+1} always satisfy (5.32). The same applies for Q^{k+1} and Z^{k+1}, where (5.33) is also always satisfied. Furthermore, for all $j \in \{1,\ldots,N\}$ we see that r_j^{k+1} and g_j^{k+1} always satisfy (5.34). Additionally we refer to

$$C^{k+1} := D^{k+1} - U^{k+1}\,, \tag{5.36}$$

$$X^{k+1} := Z^{k+1} - AU^{k+1} \qquad \text{and} \tag{5.37}$$

$$y_j^{k+1} := g_j^{k+1} - \nabla u_j^{k+1} \qquad \forall\, j \in \{1,\ldots,N\} \tag{5.38}$$

as the primal residuals at iteration $k+1$.

With (5.28) - (5.34) we have seven optimality conditions, where we have seen that (5.32) - (5.34) are always satisfied. The remaining four optimality conditions (5.28) - (5.31) yield the primal residuals (5.36) - (5.38) and the dual residual (5.35), which converge to zero as ADMM proceeds (cf. Boyd et al. 2010, Appendix A, p. 106 et seqq.).

Stopping Criteria

Once more we derive the stopping criteria in analogy to Boyd et al. (2010, Section 3.3.1).

The primal and dual residuals can again be related to a bound on the objective suboptimality of the current point κ^*. By estimating $\|U^k - U^*\|_F \le \nu$ we obtain

$$\begin{aligned}
&\frac{1}{2}\left\|ZB^T - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + \gamma \sum_{j=1}^{N} \|g_j\|_1 + J(D) - \kappa^* \\
&\le \|P^k\|_F \|C^k\|_F + \|Q^k\|_F \|X^k\|_F + \sum_{j=1}^{N} \left\|r_j^k\right\|_2 \left\|y_j^k\right\|_2 + \nu \sum_{j=1}^{N} \|s_j^k\|_2\,.
\end{aligned}$$

Hence we have to limit the primal and dual residuals again, i.e.

$$\|C^k\|_F \le \varepsilon^{\text{pri}}\,, \qquad \|X^k\|_F \le \xi^{\text{pri}}\,, \qquad \|y_j^k\|_2 \le \zeta_j^{\text{pri}} \;\; \forall\, j\,, \qquad \|s_j^k\|_2 \le \varepsilon_j^{\text{dual}} \;\; \forall\, j\,,$$

with positive tolerances ε^{pri}, ξ^{pri}, ζ_j^{pri} and $\varepsilon_j^{\text{dual}}$ for the feasibility conditions (5.28) - (5.31), respectively. Once more we choose these tolerances as

$$\begin{aligned}
\varepsilon^{\text{pri}} &= \sqrt{MN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\|U^k\|_F, \|D^k\|_F, 0\right\}\,, \\
\xi^{\text{pri}} &= \sqrt{LN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\|AU^k\|_F, \|Z^k\|_F, 0\right\}\,, \\
\zeta_j^{\text{pri}} &= \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\|\nabla u_j^k\|_2, \|g_j^k\|_2, 0\right\} && \forall\, j \in \{1, \dots, N\}\,, \\
\varepsilon_j^{\text{dual}} &= \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \|\lambda p_j^k + \mu A^T q_j^k - \eta_j \nabla \cdot r_j^k\|_2 && \forall\, j \in \{1, \dots, N\}\,,
\end{aligned}$$

where $\varepsilon^{\text{rel}} = 10^{-3}$ or 10^{-4} is a relative tolerance and the absolute tolerance ε^{abs} depends on the scale of the typical variable values.

Adaptive Parameter Choice

The adaptive parameter choice from Boyd et al. (2010, Section 3.4.1) reads as follows:

$$\lambda^{k+1} = \begin{cases} \omega^{\text{incr}} \lambda^k, & \text{if } \|C^k\|_F > \delta \|S^k\|_F, \\ \frac{\lambda^k}{\omega^{\text{decr}}}, & \text{if } \|S^k\|_F > \delta \|C^k\|_F, \\ \lambda^k, & \text{otherwise,} \end{cases} \quad \text{and } P^{k+1} = \begin{cases} \frac{P^k}{\omega^{\text{incr}}}, & \text{if } \|C^k\|_F > \delta \|S^k\|_F, \\ P^k \omega^{\text{decr}}, & \text{if } \|S^k\|_F > \delta \|C^k\|_F, \\ P^k, & \text{otherwise,} \end{cases}$$

$$\mu^{k+1} = \begin{cases} \tau^{\text{incr}} \mu^k, & \text{if } \|X^k\|_F > \vartheta \|S^k\|_F, \\ \frac{\mu^k}{\tau^{\text{decr}}}, & \text{if } \|S^k\|_F > \vartheta \|X^k\|_F, \\ \mu^k, & \text{otherwise,} \end{cases} \quad \text{and } Q^{k+1} = \begin{cases} \frac{Q^k}{\tau^{\text{incr}}}, & \text{if } \|X^k\|_F > \vartheta \|S^k\|_F, \\ Q^k \tau^{\text{decr}}, & \text{if } \|S^k\|_F > \vartheta \|X^k\|_F, \\ Q^k, & \text{otherwise,} \end{cases}$$

$$\eta_j^{k+1} = \begin{cases} \sigma^{\text{incr}} \eta_j^k, & \text{if } \|y_j^k\|_2 > \rho \|s_j^k\|_2, \\ \frac{\eta_j^k}{\sigma^{\text{decr}}}, & \text{if } \|s_j^k\|_2 > \rho \|y_j^k\|_2, \\ \eta_j^k, & \text{otherwise,} \end{cases} \quad \text{and } r_j^{k+1} = \begin{cases} \frac{r_j^k}{\sigma^{\text{incr}}}, & \text{if } \|y_j^k\|_2 > \rho \|s_j^k\|_2, \\ r_j^k \sigma^{\text{decr}}, & \text{if } \|s_j^k\|_2 > \rho \|y_j^k\|_2, \\ r_j^k, & \text{otherwise,} \end{cases}$$

for all $j \in \{1, \dots, N\}$, where $\delta,\ \vartheta,\ \rho > 1$ and $\omega,\ \tau,\ \sigma > 1$ holds. Typical choices are $\delta,\ \vartheta,\ \rho = 10$ and $\omega,\ \tau,\ \sigma = 2$.

Solving the $\ell^{1,\infty}$-$\ell^{1,1}$-TV-Regularized Problem Including TV on the Coefficient Matrices

Let us now state the complete algorithm for the solution of problem (5.26).

Algorithm 3 $\ell^{1,\infty}$-$\ell^{1,1}$-TV(U)-regularized problem via ADMM with triple splitting

Parameters: $v > 0,\ \beta > 0,\ \gamma > 0,\ A \in \mathbb{R}^{L\times M},\ B \in \mathbb{R}^{T\times N}, W \in \mathbb{R}^{L\times T},$
$\delta, \vartheta, \rho > 1,\ \omega, \tau, \sigma > 1,\ \varepsilon^{\text{rel}} = 10^{-3}$ or $10^{-4},\ \varepsilon^{\text{abs}} > 0$

Initialization: $U, D, Z, G, P, Q, R, S, C, X, Y = 0,\ \varepsilon^{\text{pri}} = \sqrt{MN}\,\varepsilon^{\text{abs}},$
$\xi^{\text{pri}} = \sqrt{LN}\,\varepsilon^{\text{abs}},\ \zeta_j^{\text{pri}} = \sqrt{M}\,\varepsilon^{\text{abs}},\ \varepsilon_j^{\text{dual}} = \sqrt{M}\,\varepsilon^{\text{abs}}$

while $\|C\|_F > \varepsilon^{\text{pri}}$ & $\|X\|_F > \xi^{\text{pri}}$ & $\|y_j\|_2 > \zeta_j^{\text{pri}}\ \forall\, j$ & $\|s_j\|_2 > \varepsilon_j^{\text{dual}}\ \forall\, j$ **do**

  $D^{\text{old}} = D;$

  $Z^{\text{old}} = Z;$

  $G^{\text{old}} = G;$

  for all $j \in \{1, \ldots, N\}$ **do**

▷ Solve Subproblems I

    $u_j = \left(\lambda I + \mu A^T A - \eta_j \Delta\right)^{-1}\left(\lambda\,(d_j - p_j) + \mu A^T(z_j - q_j) - \eta_j \nabla\cdot(g_j - r_j)\right);$

    if $\eta_j \neq 0$ **then**

      $g_j = \operatorname{sign}(\nabla u_j + r_j)\max\left(|\nabla u_j + r_j| - \frac{\gamma}{\eta_j},\ 0\right)$

    end if

▷ Update Primal Residual I

    $y_j = g_j - \nabla u_j;$

▷ Lagrange Updates I

    $r_j = r_j - (g_j - \nabla u_j);$

▷ Stopping Criteria I

    $\zeta_j^{\text{pri}} = \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\max\{\|\nabla u_j\|_2, \|g_j\|_2, 0\};$

  end for

▷ Solve Subproblems II

  $D = \underset{D\in G}{\operatorname{argmin}}\ \frac{\lambda}{2}\|D - U + P\|_F^2 + \beta\sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij}\ \text{ s.t. }\ \sum_{j=1}^{N} d_{ij} \leq v;$ ▷ see Appendix A

  $Z = \left(WB + \mu\,(AU + Q)\right)\left(B^T B + \mu I\right)^{-1};$

▷ Update Primal Residuals II

  $C = D - U;$

  $X = Z - AU;$

▷ Lagrange Updates II

  $P = P - (D - U);$

  $Q = Q - (Z - AU);$

▷ Stopping Criteria II

$\varepsilon^{\text{pri}} = \sqrt{MN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\|U\|_F, \|D\|_F, 0\};$

$\xi^{\text{pri}} = \sqrt{LN}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\|AU\|_F, \|Z\|_F, 0\};$

▷ Update Dual Residual

for all $j \in \{1, \dots, N\}$ **do**

  $s_j = \lambda\left(d_j^{\text{old}} - d_j\right) + \mu A^T\left(z_j^{\text{old}} - z_j\right) - \eta_j \nabla\cdot\left(g_j^{\text{old}} - g_j\right);$

  ▷ Stopping Criteria III

  $\varepsilon_j^{\text{dual}} = \sqrt{M}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\|\lambda p_j + \mu A^T q_j - \eta_j \nabla\cdot r_j\|_2;$

  ▷ Varying Penalty/Lagrange Parameters I

  if $\|y_j\|_2 > \rho\|s_j\|_2$ **then**

    $\eta_j = \eta_j \sigma^{\text{incr}};$

    $r_j = \frac{r_j}{\sigma^{\text{incr}}};$

  else if $\|s_j\|_2 > \rho\|y_j\|_2$ **then**

    $\eta_j = \frac{\eta_j}{\sigma^{\text{decr}}};$

    $r_j = r_j \sigma^{\text{decr}};$

  end if

end for

▷ Varying Penalty/Lagrange Parameters II

if $\|C\|_F > \delta\|S\|_F$ **then**

  $\lambda = \lambda\omega^{\text{incr}};$

  $P = \frac{P}{\omega^{\text{incr}}};$

else if $\|S\|_F > \delta\|C\|_F$ **then**

  $\lambda = \frac{\lambda}{\omega^{\text{decr}}};$

  $P = P\omega^{\text{decr}};$

end if

if $\|X\|_F > \vartheta\|S\|_F$ **then**

  $\mu = \mu\tau^{\text{incr}};$

  $Q = \frac{Q}{\tau^{\text{incr}}};$

else if $\|S\|_F > \vartheta\|X\|_F$ **then**

  $\mu = \frac{\mu}{\tau^{\text{decr}}};$

  $Q = Q\tau^{\text{decr}};$

end if

end while

return U ▷ Solution of (5.26)

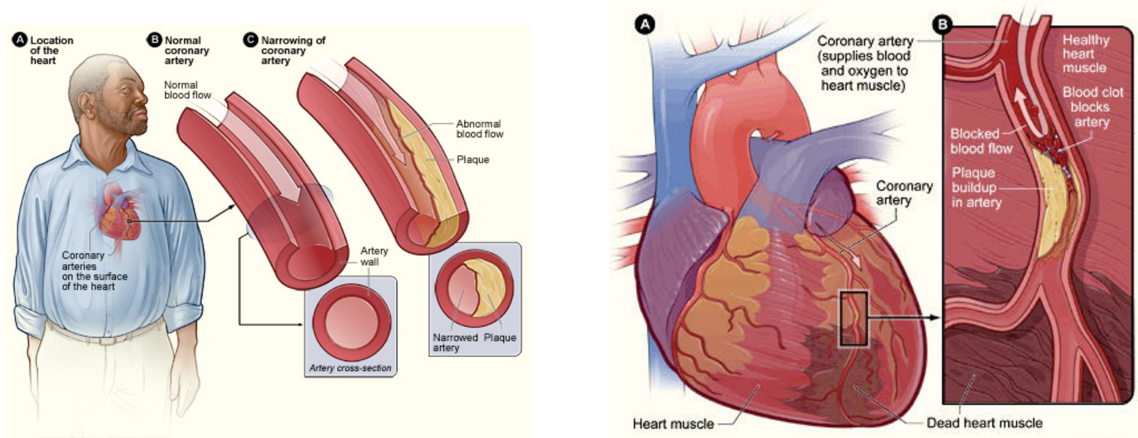
6

Computational Experiments

In the last chapters we have analyzed $\ell^{1,\infty}$-regularized variational models and its reformulations. Moreover, we have deduced different algorithms for the computation of a solution for the $\ell^{1,\infty}$-regularized minimization problem.
In this chapter we propose dynamic positron emission tomography for the visualization of myocardial perfusion as a possible application. To incorporate knowledge about this application, we include *kinetic modeling* in order to model the blood flow and tracer exchange in the heart muscle. After an introduction to the corresponding medical and mathematical background, we show some results for synthetic examples and discuss the quality of our approach. Afterwards, we present some results using total variation as proposed before in Sections 4.3 and 5.2. We distinguish between applying total variation to the image in every time step and to the coefficient matrices for every basis vector.

6.1. Application to Dynamic Positron Emission Tomography

Coronary heart disease (CHD), also known as atherosclerosis, is the most common form of heart disease. It often leads to heart attacks, which are life-threatening and in many cases deathly. In 2012 most cases of death in the world resulted from CHD, cf. for instance Finegold et al. (2013). It was even the major cause of hospital admissions. The slow buildup of plaque (cf. Figure 6.1) along the inner walls of the arteries of the heart is usually overlooked in the initial stage, since the perceivable symptoms arise not until an advanced state of the disease is reached. Many people with CHD do not exhibit symptoms for a long time. As the disease progresses, often a sudden heart attack arises and the patient is caught by surprise. Thus it is challenging to medicate the patient in time.



(a) The buildup of plaque limits the flow of oxygen-rich blood through the artery.

(b) Heart with a blocked artery, which caused a heart attack. We see the resulting muscle damage.

Figure 6.1.: U. S. National Heart, Lung and Blood Institute (2012)

In order to make a diagnosis and administer the right therapy, the medical doctors need to know how well the cardiac muscle is perfused and where exactly the problems are located. Thus it is necessary to visualize the perfusion of the cardiac muscle and to locate damaged areas after a heart attack. Apart from this, it would be a great advantage to find problematic arteries, to even prevent an infarction.
One way to visualize the myocardial perfusion is dynamic positron emission tomography. Positron Emission Tomography (PET) is an imaging technique used in nuclear medicine that visualizes the distribution of a radioactive tracer, which was applied to the patient, see for instance Bailey et al. (2005) and Vardi et al. (1985). Compared to computer tomography (CT), PET has the advantage of being a functional rather than a morphological imaging technique. Figure 6.2 demonstrates a schema of the reconstruction process of PET.

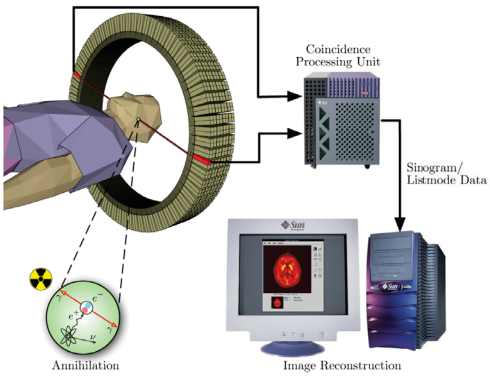


Figure 6.2.: Schema showing the different processing steps of PET by Langner (2003)

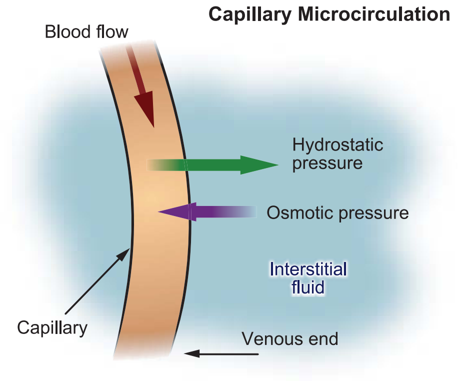
By using radioactive water ($H_2^{15}O$) as a tracer, it is possible to visualize blood flow. $H_2^{15}O$ has the advantage of being highly diffusible and the radiation exposure is low. Even dynamic images are possible. On the other hand the reconstructed images have poor quality due to the short radioactive half-life of $H_2^{15}O$.
Now let us consider the inverse problem of dynamic PET, i.e.

$$\mathcal{A}Z = W \,, \tag{6.1}$$

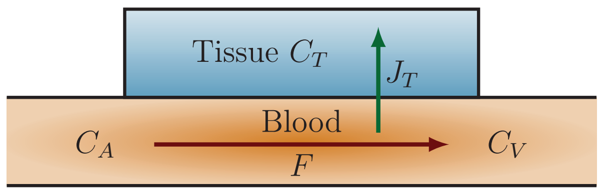
where the operator $\mathcal{A}$ linking the dynamic image Z with the measured data W is usually the *Radon operator*, but could also be another operator depending on the application. Using kinetic modeling (cf. Wernick and Aarsvold 2004, Chapter 23, p. 499 et seqq.) we are able to describe the unknown image Z as the tracer concentration in the tissue C_T, i.e.

$$C_T(x,t) = F(x) \int_0^t C_A(\tau) e^{-\frac{F(x)}{\lambda}(t-\tau)}\, \mathrm{d}\tau \,, \tag{6.2}$$

where $C_A(\tau)$ is the arterial tracer concentration, also called input curve, $F(x)$ refers to the perfusion and λ is the ratio between the tracer concentration in tissue and the venous tracer concentration resulting from Fick's principle.



(a) Capillary microcirculation, U. S. National Cancer Institute (2006)



(b) Illustration of kinetic modeling

Figure 6.3.: Capillary microcirculation and kinetic modeling

Kinetic modelling describes the tracer exchange with the tissue in the capillaries. The tracer is injected and flows from the arteries with concentration C_A to the veins with concentration C_V. While passing the capillaries between arteries and veins, a part of it moves across the vascular wall with flux J_T into the tissue, cf. Figure 6.3.
Expression (6.2) is an integral equation including the exponential factor $e^{-\frac{F(x)}{\lambda}(t-\tau)}$,

which depends on both input arguments, i.e. time t and space x. This expression is highly nonlinear and thus not easy to handle, especially in combination with inverse problems. Due to the fact that we have prior knowledge about $\frac{F(x)}{\lambda}$, we are able to provide a big pool of given perfusion values for this expression. Subsequently, we can consider a linearization, i.e.

$$\mathcal{B}(u, C_A) := \sum_{j=1}^{N} u_j(x) \underbrace{\int_0^t C_A(\tau) e^{-\tilde{b}_j(t-\tau)} \, \mathrm{d}\tau}_{b_j(t)} \ , \tag{6.3}$$

in which $\tilde{b}_j$ are the prior known perfusion values and the integral is now independent of space. Expression (6.3) is reasonable if there is at most one $u_j \neq 0$ for $j \in \{1, ..., N\}$, i.e. the coefficient u_j corresponding to the correct perfusion value $\tilde{b}_j$. In order to further simplify the work with this operator, we assume that the input curve C_A is predetermined.
Hereby we obtain the linear kinetic modeling operator

$$\mathcal{B}(u) = \sum_{j=1}^{N} u_j(x) b_j(t) \ , \tag{6.4}$$

which we use to describe the unknown image Z. The advantage of (6.4) over (6.2) is that we are able to compute the basis functions $b_j(t)$ in advance and thus we can provide many of those for the reconstruction process. Note that there exists another deduction of (6.4) by Reader (2007).
By considering a discretization of (6.4), we obtain

$$UB^T = \sum_{j=1}^{N} u_{ij} b_{kj} \ , \tag{6.5}$$

where i denotes the pixel and k the time step. After discretizing $\mathcal{A}$ as well, we can insert (6.5) for the image Z in (6.1) and obtain

$$AUB^T = W \ . \tag{6.6}$$

Hence (4.8) can be used for the reconstruction of the discretized coefficients u_{ij}.

6.2. Results for Synthetic Examples

In this section we present some numerical results. We are going to work on synthetic data to investigate the effectiveness of our approach. In order to do so, we use a simple 3D matrix $\widehat{U}$ containing the exact coefficients as ground truth, i.e. two spatial dimensions and one extra dimension referring to the number of basis vectors.

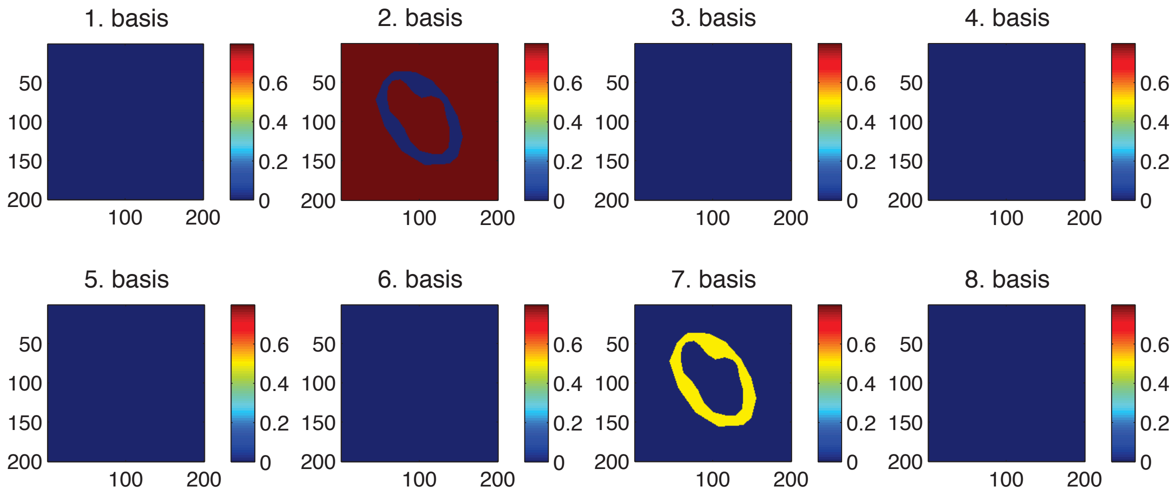


Figure 6.4.: Ground truth $\widehat{U} \in \mathbb{R}^{200\times200\times8}$ with 200^2 pixels and 8 basis vectors

Defining two regions, where the coefficients are nonzero for only one basis vector, yields the fact that the corresponding coefficients for most of the basis vectors are zero. Obviously our ground truth fulfills the prior knowledge, which we would like to promote in the reconstruction, i.e. there is only one coefficient per pixel, which is unequal to zero. In Figure 6.4 we see that the exact coefficients for the most basis vectors are zero. Only some coefficients corresponding to the second and seventh basis vectors are nonzero.

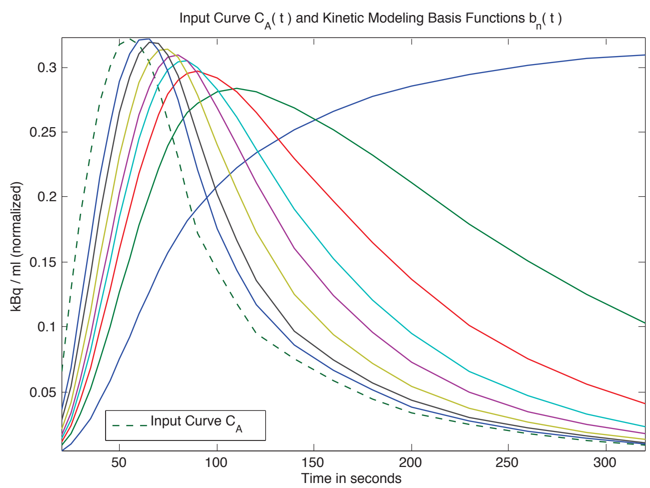


Figure 6.5.: Discretized kinetic modeling basis functions B^T plotted on a fine grid

In order to obtain the artificial data $W \in \mathbb{R}^{L\times T}$, we have to apply the matrices $A \in \mathbb{R}^{L\times M}$ and $B^T \in \mathbb{R}^{N\times T}$ to the ground truth $\widehat{U} \in \mathbb{R}^{M\times N}$. As an example for B we use kinetic modeling basis vectors as they are used in dynamic positron emission tomography (cf. Section 6.1 and Wernick and Aarsvold (2004, Chapter 23)), which are basically exponential functions with different parameters, cf. also Appendix C. In Figure 6.5 we see that those basis vectors are very similar, i.e. they are very coherent. For the verification of our model, we use as matrix A a simple 2D convolution in space.

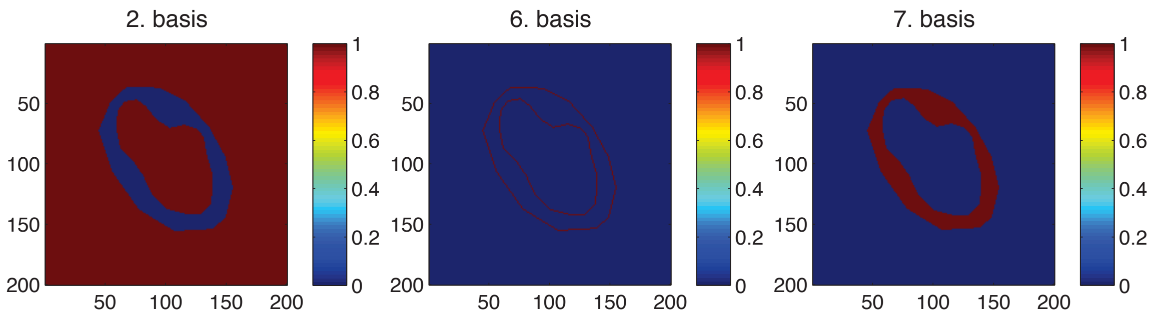


Figure 6.6.: 2nd, 6th and 7th reconstructed coefficient matrices using $\tilde{v} = 0.01$ and $\beta = 0.1$

By using Algorithm 1 on the so computed data W and including a strong $\ell^{1,\infty}$-regularization, i.e. $\tilde{v} = 0.01$, we obtain a very good reconstruction of the support. Figure 6.6 only shows the coefficient matrices to those basis vectors, which include reconstructed nonzero coefficients. For the sake of simplicity, we do not show the other reconstructed coefficient matrices, which are completely zero. Obviously we obtain a very good reconstruction of the support. Only a few coefficients, which actually correspond to the seventh basis vector, are reconstructed wrongly and show up in the sixth basis vector. This is due to the coherence of the basis vectors, i.e. the sixth and the seventh basis vector are very similar, compare for instance Figure 6.5.

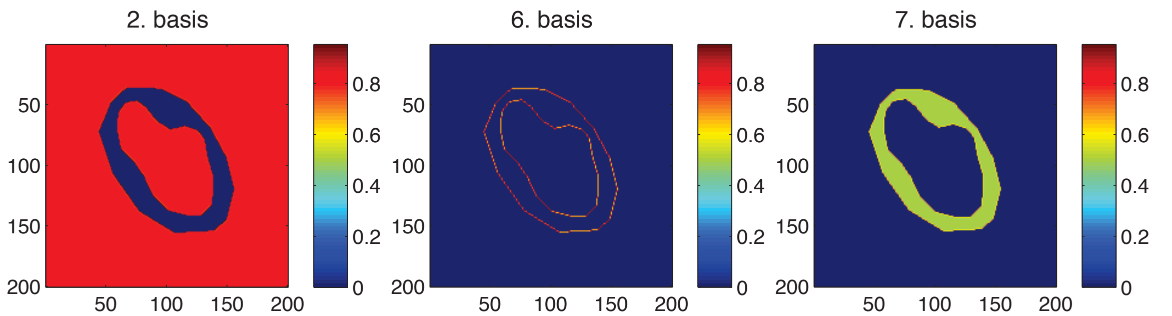


Figure 6.7.: 2nd, 6th and 7th reconstructed coefficient matrix using $\tilde{v} = 0.01$ and $\beta = 0.1$ including a second run only on the previously computed support; the other coefficient matrices are completely zero

However, we observe that in Figure 6.6 every value larger than $\tilde{v}$ is projected down to $\tilde{v}$ and we make a systematic error. This is due to the inequality constraint in problem

(4.16) and because of the fact that we chose $\tilde{v}$ smaller than the maximal value of the exact data $\widehat{U}$ (compare for instance Subsection 4.1.4 and especially Theorem 4.7). Thus we are not really close to the exact data. In order to overcome this problem, we first reconstruct the support including the $\ell^{1,\infty}$- and $\ell^{1,1}$-regularization and then perform a second run without regularization *only on the known support* to reduce the distance to the exact data.

In Figure 6.7 we see that this approach leads to very good results. We additionally reconstructed an example including some Gaussian noise. In Figure 6.8 we observe that the algorithm performs quite nicely, especially if we chose $\tilde{v}$ lower than before. For the example in Figure 6.8 we used $\tilde{v} = 10^{-5}$, which led to the best results. However, choosing $\tilde{v} = 10^{-3}$ already yields very similar results and needs less iterations, cf. Table 6.2.

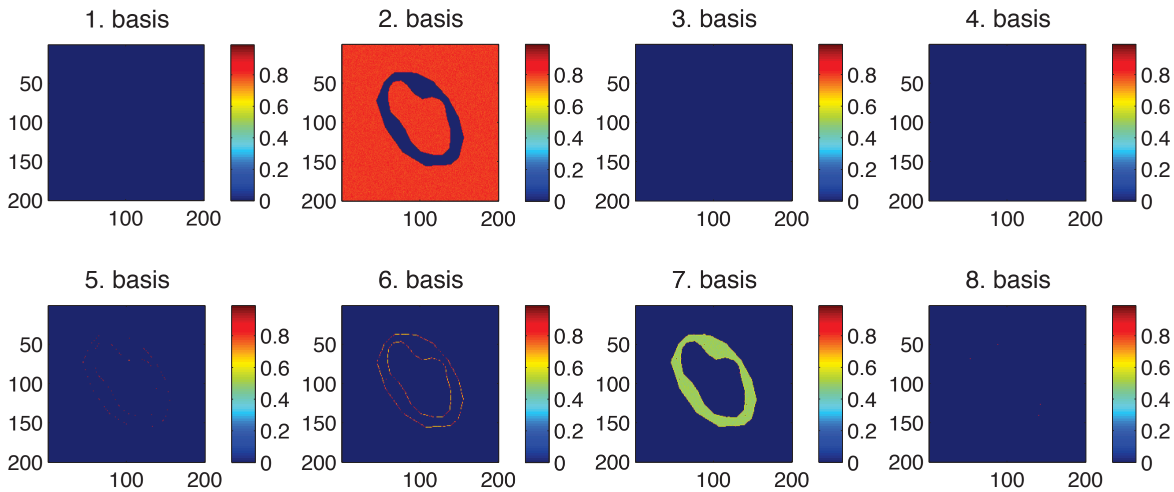


Figure 6.8.: Reconstruction using $\tilde{v} = 10^{-5}$ and $\beta = 0.1$ including Gaussian noise with standard deviation $\sigma = 0.01$

Let us now evaluate Algorithm 1 with respect to the quality of the reconstructed support. In order to do so, we compare the reconstructed support after the first run (including both regularizations) with the support of our ground truth and state how much percent of the true support is reconstructed wrongly depending on the $\ell^{1,\infty}$-regularization parameter $\tilde{v}$. We also include the distance of the wrongly picked basis vector in each pixel, for instance if the support of the ground truth picks basis vector number 7 and the reconstructed support picks basis vector number 5 instead, we double the influence of the error in this pixel if the reconstructed support picks basis vector number 4 instead of the correct number 7 we triple it and so on.

In Table 6.1 we see the evaluation of Algorithm 1 applied to the noiseless data W. When $\tilde{v}$ becomes smaller than 0.01 we observe that there is no further improvement. As we have seen in Figure 6.6 the boundary of the region is reconstructed wrongly and the algorithm selects the sixth instead of the seventh basis vector. However, the prior

$\tilde{v}$	Percentage of wrong pixel	Number of iterations
10^{-1}	0.6722 %	261
10^{-2}	0.1772 %	349
10^{-3}	0.1772 %	430
10^{-4}	0.1772 %	511
10^{-5}	0.1772 %	591
10^{-6}	0.1772 %	662
10^{-7}	0.1772 %	671

Table 6.1.: Evaluation of Algorithm 1 with $\beta = 0.1$

knowledge is already fulfilled, i.e. in every pixel there is only one basis vector active. This is the reason why there are still 0.1772% wrong pixel and we do not obtain further improvement.

$\tilde{v}$	Percentage of wrong pixel	Number of iterations
10^{-1}	3.7016 %	261
10^{-2}	0.2591 %	349
10^{-3}	0.1956 %	430
10^{-4}	0.1913 %	511
10^{-5}	0.1909 %	591
10^{-6}	0.1925 %	662
10^{-7}	0.1928 %	671

Table 6.2.: Evaluation of Algorithm 1 with $\beta = 0.1$ including Gaussian noise with standard deviation $\sigma = 0.01$

$\tilde{v}$	Percentage of wrong pixel	Number of iterations
10^{-1}	6.4450 %	260
10^{-2}	1.4628 %	349
10^{-3}	1.0153 %	430
10^{-4}	1.0078 %	511
10^{-5}	1.0172 %	591
10^{-6}	0.9694 %	662
10^{-7}	0.9834 %	671

Table 6.3.: Evaluation of Algorithm 1 with $\beta = 0.1$ including Gaussian noise with standard deviation $\sigma = 0.05$

In Table 6.2 and 6.3 we see the error measures for the same values of $\tilde{v}$ as in Table 6.1. But this time we included Gaussian noise on the data W with standard deviation 0.01 and 0.05, respectively. At first the error drops quickly. However, when $\tilde{v}$ becomes smaller, the error stagnates in a certain range similar to the noise free case.

In order to smartly choose $\tilde{v}$, we have to find a good tradeoff between a small error and a small number of iterations. Choosing $\tilde{v} \in [10^{-4}, \dots, 10^{-3}]$ seems to be a good choice.

6.3. Reconstruction Including Total Variation

In this section we include total variation into the reconstruction process. Using an additional total variation regularization should improve the results. However, it also complicates the process, since it becomes quite difficult to choose the right combination of regularization parameters. Furthermore, the algorithms become a bit slower than without using total variation, since the new methods are more complex than our first approach.

In order to validate Algorithms 2 and 3 we consider problems (5.12) and (5.26). For the reconstruction we use the same data $W \in \mathbb{R}^{L \times T}$ as in the previous Section 6.2. Moreover, we use the same discretized operators, i.e. $A \in \mathbb{R}^{L \times M}$ is a 2D-convolution in space for every basis vector and $B \in \mathbb{R}^{T \times N}$ is the discretized kinetic modeling operator as deduced in Section 6.1.

For both algorithms deduced in Section 5.2, we first of all want to obtain an overview, in which order of magnitude we should seek for a good combination of regularization parameters. Afterwards, we refine our search, in order to obtain the parameter combination, which yields the lowest percentage of wrongly reconstructed coefficients. Once more we use the same error measure as proposed before in Section 6.2, i.e. we compute the percentage of wrongly reconstructed pixels of the support in consideration of the distance of the wrongly picked basis vector to the true coefficient.

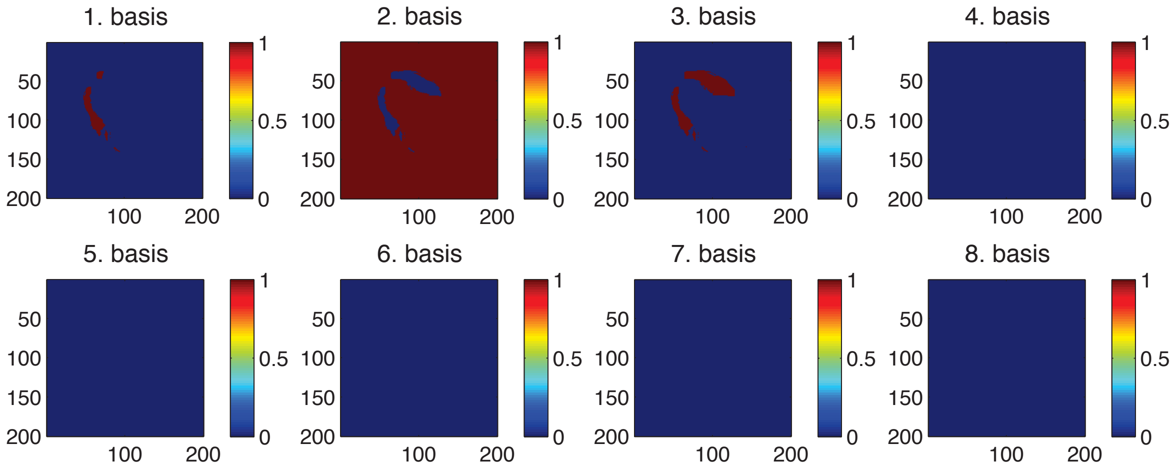
6.3.1. Additional Total Variation on the Image

Let us first discuss the results of Algorithm 2 for problem (5.12). In order to obtain an overview of the regularization parameters and in which combination and order of magnitude they should be chosen, we display the percentage of wrongly reconstructed coefficients for different regularization parameters in Table 6.4. The highlighted cells show the lowest and thus most interesting results. They indicate in which range of parameters we should search for the optimal parameter combination.

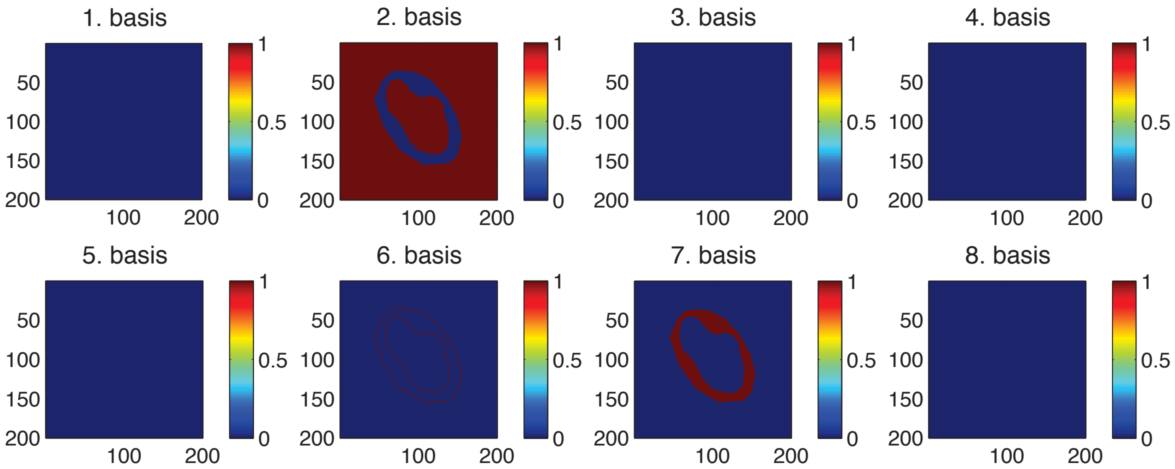
$\tilde{v}$ \ γ	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.8025 %	0.6203 %	0.2116 %	0.1772 %
10^{-3}	6.6056 %	0.3378 %	0.1825 %	0.1772 %
10^{-4}	2.9175 %	1.2641 %	0.1825 %	0.1772 %
10^{-5}	1.8956 %	7.9172 %	0.2266 %	0.2119 %

Table 6.4.: Rough overview about the wrongly reconstructed coefficients for different regularization parameters $\tilde{v}$ and γ. The highlighted cells indicate the smallest errors in this table.

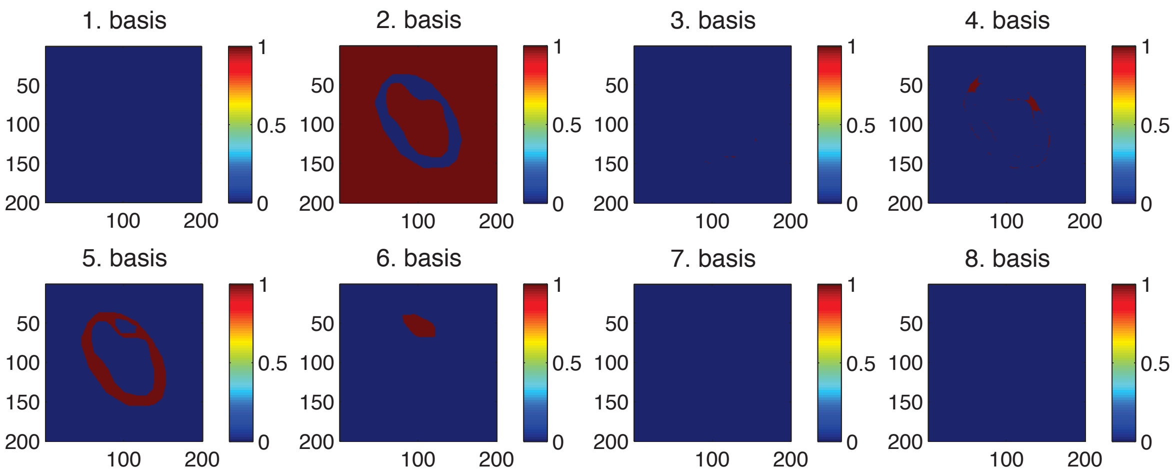
Let us now present some examples for differently strong TV-regularizations.



(a) Reconstructed support with too large TV-regularization; $\tilde{v} = 10^{-3}$ and $\gamma = 10^{-3}$



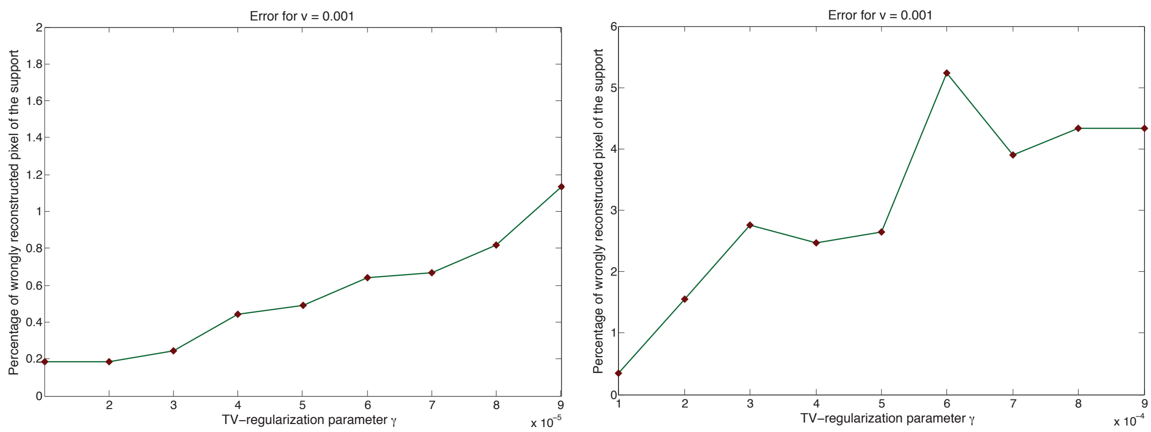
(b) Reconstructed support with too low TV-regularization; $\tilde{v} = 10^{-3}$ and $\gamma = 10^{-6}$



(c) Reconstructed support with medium TV-regularization; $\tilde{v} = 10^{-3}$ and $\gamma = 5 \times 10^{-4}$

Figure 6.9.: Examples of differently strong influences of the total variation

Obviously the results for $\gamma = 10^{-3}$ are not satisfying at all, cf. Subfigure 6.9(a). In this cases the total variation regularization is too strong and most of the nonzero values appear in the coefficient matrices for the former basis vectors. The latter coefficient matrices are completely zero. By choosing $\gamma = 10^{-6}$, cf. Subfigure 6.9(b), we indeed obtain the same percentage of wrongly reconstructed coefficients as without using total variation. However, in this case the TV-regularization has no influence at all on the reconstruction. We especially see this in the reconstructed coefficients for the sixth basis vector. They still do not differ from those obtained without using TV-regularization. Choosing γ in between those values, i.e. $\gamma = 5 \times 10^{-4}$, neither yields satisfying results, cf. Subfigure 6.9(c). It rather deteriorates them by completely ignoring the nonzero values of the coefficient matrices for the 7th basis vector. Instead those nonzero values predominantly appear in the coefficient matrix for the 5th basis vector, which should actually be zero.
Nevertheless, let us analyze some results for $\tilde{v} = 10^{-3}$ for different values of γ, since this seems to be the most promising choice for the $\ell^{1,\infty}$-regularization, cf. Table 6.4. We observe in Figure 6.10 that the error of the support descends as γ becomes smaller. This indicates that we will indeed obtain the smallest error by choosing γ so small that the total variation regularization has no influence on the reconstruction. The same happens if we choose $\tilde{v}$ even smaller than that.



(a) $\tilde{v} = 10^{-3}$ and $\gamma \in [1, 9] \times 10^{-5}$ (b) $\tilde{v} = 10^{-3}$ and $\gamma \in [1, 9] \times 10^{-4}$

Figure 6.10.: Wrongly reconstructed coefficients of the support in percent for $\tilde{v}$ in the order of magnitude 10^{-3} and γ in the order of magnitude 10^{-4} and 10^{-5}

In our attempt to find the reason for this behaviour, we observed that the conjugate gradient method, which we use in every substep to solve a system of linear equations, for some reason does not converge to a desired tolerance. Moreover, the deduced stopping criteria do not abort the algorithm. However, even if we manually abort the algorithm earlier, the results are worse than our results from Section 6.2.

In conclusion, we have to observe that including an additional total variation regularization on the image in every time step yields worse results than without using total variation. Furthermore, having to choose two regularization parameters complicates the process and we receive the impression that in this case both regularizations conflict each other.

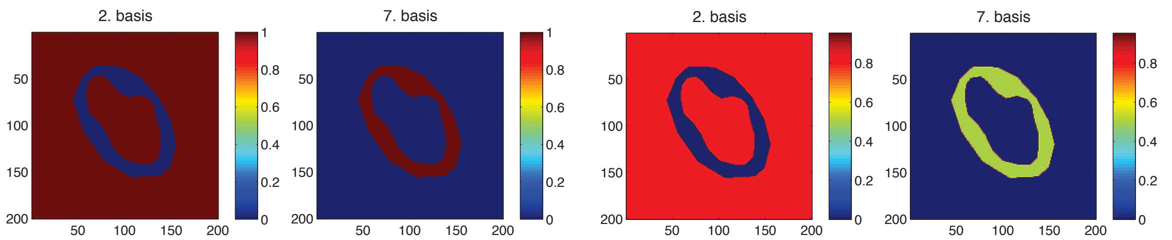
6.3.2. Additional Total Variation on the Coefficient Matrices

Let us now evaluate Algorithm 3 for problem (5.26). In Table 6.5 we compare the results for regularization parameters in different orders of magnitude. The highlighted cells display the lowest percentage of wrongly reconstructed coefficients within this table.

$\tilde{v}$ \ γ	1×10^{-3}	5×10^{-4}	1×10^{-4}	5×10^{-5}	1×10^{-5}
1×10^{-2}	11.7078 %	0.0247 %	0.0025 %	0 %	0.4594 %
5×10^{-3}	9.2963 %	0.0253 %	0.0025 %	0 %	0.0075 %
1×10^{-3}	10.0231 %	10.0231 %	8.7916 %	0.9866 %	0.0075 %

Table 6.5.: Rough overview about the error for different regularization parameters $\tilde{v}$ and γ. The highlighted cells indicate the smallest errors in this table.

The best result without including total variation into our method was 0.1772 % of wrongly reconstructed coefficients. We observe that there are several parameter combinations in Table 6.5, which yield better results than the best results from our first attempt without using total variation, cf. Section 6.2. There even exist two parameter combinations, for which the coefficients are reconstructed exactly, i.e. $\gamma = 5 \times 10^{-5}$ in combination with $\tilde{v} = 10^{-2}$ and $\tilde{v} = 5 \times 10^{-3}$, cf. the highlighted cells in Table 6.5. In Figure 6.11 we present the result for $\tilde{v} = 5 \times 10^{-3}$ and $\gamma = 5 \times 10^{-5}$, which was computed in less iterations than the other exactly reconstructed result, i.e. 314. We show only the coefficient matrices to those basis vectors, which contain nonzero coefficients.



(a) Reconstructed support

(b) Reconstruction on the previously computed support

Figure 6.11.: 2nd and 7th coefficient matrices of the reconstructed support with $\tilde{v} = 5 \times 10^{-3}$ and $\gamma = 5 \times 10^{-5}$ and the result, which was computed on it. The coefficient matrices to every other basis vector are completely zero.

Motivated by these superior results, we tested the algorithm with noisy data as well. Similarly to Section 6.2 we included additive Gaussian noise on the data W with standard deviation $\sigma = 0.01$ and $\sigma = 0.05$.

In Table 6.6(a) we see the results for $\sigma = 0.01$, where $\tilde{v}$ and γ are chosen in different orders of magnitude. Even in the presence of low Gaussian noise our algorithm reconstructs the support exactly. In contrast to Algorithm 1, which does not involve total variation, it is not even necessary to change the regularization parameters.

In case that we have $\sigma = 0.05$, Table 6.6(b) shows that we are not anymore able to reconstruct the support exactly. Nevertheless, the best result is still much better than the best result we could obtain by using Algorithm 1, which reconstructed at best 0.9694 % of the coefficients falsely for $\sigma = 0.05$. By using Algorithm 3, however, we obtain an almost exact reconstruction of the support, i.e. only 0.01 % of the coefficients are reconstructed wrongly.

(a) Standard deviation $\sigma = 0.01$

$\tilde{v}$ \ γ	1×10^{-3}	5×10^{-4}	1×10^{-4}	5×10^{-5}	1×10^{-5}
1×10^{-2}	0.7403 %	0.0247 %	0.0025 %	0 %	0.39 %
5×10^{-3}	9.3662 %	0.0247 %	0.0025 %	0 %	0.2616 %
1×10^{-3}	10.0244 %	10.0231 %	8.3759 %	0.1225 %	0.0075 %

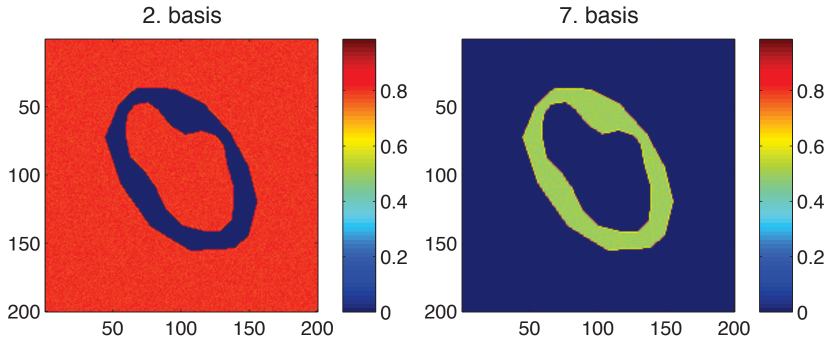
(b) Standard deviation $\sigma = 0.05$

$\tilde{v}$ \ γ	1×10^{-3}	5×10^{-4}	1×10^{-4}	5×10^{-5}	1×10^{-5}
1×10^{-2}	0.8188 %	0.0547 %	0.0172 %	0.01 %	2.4431 %
5×10^{-3}	10.4631 %	0.1119 %	0.0344 %	0.0131 %	0.4834 %
1×10^{-3}	8.5144 %	10.0231 %	2.4806 %	1.0144 %	0.1850 %

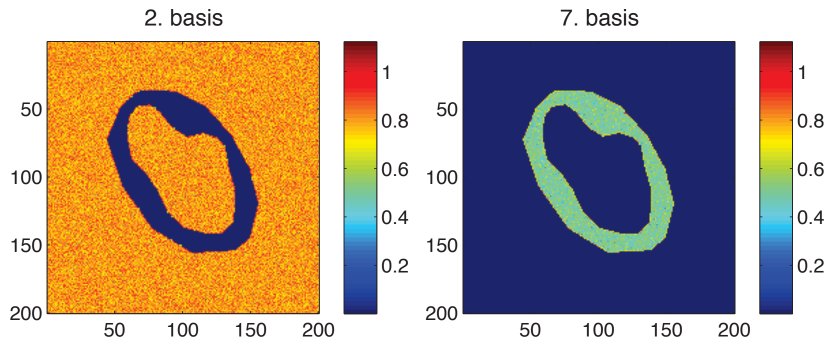
Table 6.6.: Rough overview about the error for different regularization parameters $\tilde{v}$ and γ for noisy data. The highlighted cells indicate the smallest errors in these tables.

For standard deviation $\sigma = 0.01$ we see in Subfigure 6.12(a) the nonzero coefficient matrices of the reconstruction computed on the previously recovered support for $\tilde{v} = 0.01$ and $\gamma = 5 \times 10^{-5}$. This result was reconstructed in 392 iterations. We observe that the algorithm needed slightly more iterations for $\sigma = 0.01$ than in the noise-free case. Subfigure 6.12(b) shows the reconstruction on the previously computed support for $\tilde{v} = 0.01$ and $\gamma = 5 \times 10^{-5}$. This time the data was corrupted by Gaussian noise with standard deviation $\sigma = 0.05$. We observe that the support is still reconstructed almost exactly. For the recovery of the support the algorithm needed the same number of iterations as before, i.e. 392.

Note that, unless stated otherwise, all the figures of the reconstructions in this subsection involve also the second run, which solves the nonnegative least squares problem only on the previously computed support.



(a) Reconstruction including Gaussian noise with standard deviation $\sigma = 0.01$



(b) Reconstruction including Gaussian noise with standard deviation $\sigma = 0.05$

Figure 6.12.: 2nd and 7th reconstructed coefficient matrices, which were reconstructed on the previously recovered support including Gaussian noise. In both cases we used $\tilde{v} = 10^{-2}$ and $\gamma = 5 \times 10^{-5}$. The coefficient matrices to every other basis vector are completely zero.

Since the reconstruction of the support is still almost perfect, we challenge the algorithm even more and add Gaussian noise with higher standard deviation, i.e. $\sigma = 0.1$ and $\sigma = 0.2$, to the data. For the first case we see in Table 6.7(a) the percentage of wrongly reconstructed coefficients. The best result still contains only 0.0891% of falsely reconstructed coefficients, which is less than without using total variation and even without including noise in Algorithm 1.
Table 6.7(b) on the other hand shows the percentage of wrongly recovered coefficients for the reconstructions based on noisy data with standard deviation $\sigma = 0.2$. This time the result is to a greater extent corrupted by the noise. The best result now contains 0.5456% of wrongly reconstructed coefficients. Moreover, we had to increase the influence of the total variation regularization to obtain a reasonable good result.

(a) Standard deviation $\sigma = 0.1$

$\tilde{v}$ \ γ	1×10^{-3}	5×10^{-4}	1×10^{-4}	5×10^{-5}	1×10^{-5}
1×10^{-2}	1.0006 %	0.1681 %	0.0891 %	0.1138 %	8.6659 %
5×10^{-3}	16.2275 %	0.6063 %	0.1691 %	0.1519 %	7.8075 %
1×10^{-3}	9.2931 %	10.0181 %	2.5769 %	1.3253 %	2.8644 %

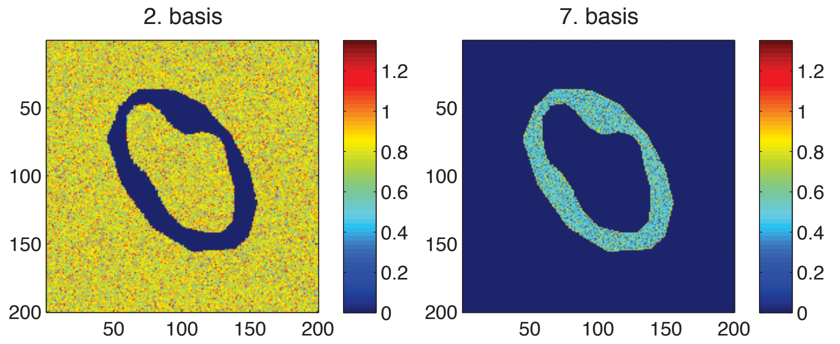
(b) Standard deviation $\sigma = 0.2$

$\tilde{v}$ \ γ	1×10^{-3}	5×10^{-4}	1×10^{-4}	5×10^{-5}	1×10^{-5}
1×10^{-2}	1.3691 %	0.5456 %	3.7209 %	6.9472 %	10.3841 %
5×10^{-3}	13.5228 %	5.1825 %	5.4438 %	6.1172 %	10.0744 %
1×10^{-3}	10.0144 %	8.7988 %	6.8419 %	5.7063 %	8.2309 %

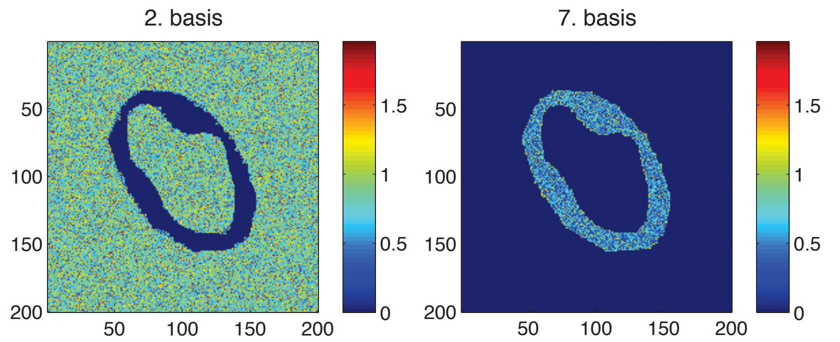
Table 6.7.: Rough overview about the error for different regularization parameters $\tilde{v}$ and γ for noisy data with higher standard deviation. The highlighted cells indicate the smallest errors in these tables.

For Gaussian noise with a standard deviation of $\sigma = 0.1$ we see the result in Subfigure 6.13(a). On the other hand, Subfigure 6.13(b) shows the reconstruction involving Gaussian noise with standard deviation $\sigma = 0.2$. In both cases the coefficient matrices corresponding to those basis vectors, which should be completely zero, are truly zero and thus are not included in the figures. Hence we notice that the support is indeed recovered very well for these two examples.

However, in case we include Gaussian noise with standard deviation $\sigma = 0.2$, we can directly observe that the edges of the nonzero regions are not recovered correctly. We are able to observe the same behaviour for $\sigma = 0.1$ and $\sigma = 0.05$, although not to this extent. In some cases the falsely recovered coefficients, which should be in the coefficient matrix corresponding to the second basis vector, can be found in the coefficient matrix, which belongs to the seventh basis vector, and vice versa. This indicates that the total variation regularization has indeed the desired effect. It keeps the nonzero regions connected and prevents that due to the noise the nonzero coefficients get fragmented over all the other basis vectors. In other cases, however, the falsely recovered coefficients get split up between the second and the seventh basis vector and thus the corresponding pixel consists of a sum of those two weighted basis vectors. For these reasons, the percentage of wrongly reconstructed coefficients can be bigger than zero, even though every coefficient matrix, which should be zero, is indeed completely zero.



(a) Standard deviation $\sigma = 0.1$, regularization parameters $\tilde{v} = 10^{-2}$ and $\gamma = 10^{-4}$



(b) Standard deviation $\sigma = 0.2$, regularization parameters $\tilde{v} = 10^{-2}$ and $\gamma = 5 \times 10^{-4}$

Figure 6.13.: 2nd and 7th reconstructed coefficient matrices, which were reconstructed on the previously recovered support including Gaussian noise. The coefficient matrices to every other basis vector are completely zero.

Note that in all the results there is no further regularization included in the second run except for the previously computed support. Therefore, the nonzero regions are indeed very noisy. This might be improved by including another regularization in the second run. These considerations, however, are not realized in this thesis and left open for future work.

In summary, Algorithm 1 already computes very satisfactory results, however, the reconstructions improve a lot when we combine the $\ell^{1,\infty}$-regularization with an additional total variation regularization on the coefficient matrices for every basis vector. Even in the presence of additive Gaussian noise Algorithm 3 performs very well. Thus we conclude that total variation regularization makes the $\ell^{1,\infty}$-problem more robust to noise.

7 Asymptotics of Spatial Sparsity Priors

In the previous chapters we have seen that sparsity regularization in inverse problems is an important and versatile tool and can even be extended to realize more advanced a-priori information such as local sparsity or joint sparsity. In this chapter we analyze the limits of such discrete approaches to develop a suitable asymptotic theory. First we motivate and introduce this new approach. In the subsequent section we consider spatial sparsity priors under Γ-convergence, which we divide into the convex case, i.e. ℓ^p-regularization for $p \geq 1$, and the special nonconvex case, where $p = 0$ holds. Moreover, we consider the asymptotics of mixed $\ell^{p,q}$-norms.

7.1. Introduction

Regularization functionals, which promote sparse solutions, are usually finite-dimensional. However, in many applications an infinite-dimensional modeling especially in the pre-image space seems to be more reasonable. Nevertheless, due to computational rather than application-related reasons, spatial sparsity is usually imposed on some spatial grid. Therefore, the establishment of an appropriate asymptotic theory behind these discrete approaches is of particular importance. An underlying asymptotic theory moreover permits an analysis independent of discretization and thus yields robustness.

This chapter is devoted to the asymptotics of regularization methods applicable for instance to spatial sparsity regularization. We are interested in recovering continuum limits such as total variation regularization or recent approaches in the space of Radon

measures, cf. for instance Scherzer and Walch (2009), Bredies and Pikkarainen (2013) and Duval and Peyré (2013). Thereby, we will also obtain continuum limits, which were not understood up to now. In order to study the asymptotics of regularization functionals, we utilize Γ-convergence, which is a useful and reasonable framework for this purpose.
We investigate the standard variational regularization of a linear inverse problem

$$\mathcal{K}\mu = w \;, \tag{7.1}$$

where the operator $\mathcal{K} : \mathcal{U} \longrightarrow \mathcal{H}$ maps a Banach space $\mathcal{U}$ into a Hilbert space $\mathcal{H}$. Here w is the noisy version of the exact data $\hat{w} = \mathcal{K}\hat{\mu}$. Forthermore, we consider the Tikhonov-type regularization

$$\min_{\mu \in \mathcal{U}} \; \frac{1}{2} \|\mathcal{K}\mu - w\|_{\mathcal{H}}^2 + \alpha \mathcal{R}(\mu)$$

and its noise-free counterpart

$$\min_{\mu \in \mathcal{U}} \; \mathcal{R}(\mu) \qquad \text{s. t.} \qquad \mathcal{K}\mu = w \;.$$

Since Γ-convergence is stable under continuous perturbations (cf. Subsection 2.1.2) and the data fidelity terms are usually continuous, it is sufficient to only consider the asymptotics of the regularization functional $\mathcal{R}$ in order to asymptotically analyze such variational problems regarding their discretizations. Thus we consider a sequence of regularization functionals $\mathcal{R}^N$, which may also take the value $+\infty$, Γ-converging to its limit $\mathcal{R}$. By this means, we attempt to understand both, the basic convergence and finer properties of the regularized solutions as N tends to infinity.

7.2. Spatial Sparsity under Γ-Convergence

Let us now consider Γ-limits of functionals that promote spatial sparsity. In order to do so, we consider a set $\Omega \subset \mathbb{R}^d$, which shall be partitioned into a union of N disjoint subsets, i.e.

$$\Omega = \bigcup_{j=1}^{N} \Omega_j^N \;.$$

Before going into the details, we want to agree on some notations. By $|\Omega_j^N|$ we denote the volume of an element in the partition. Moreover,

$$h_N := \max_{j\in\{1,\dots,N\}} \operatorname{diam}(\Omega_j^N)$$

denotes the diameter, which tends to zero as N goes to infinity.
Usually spatial sparsity is applied to a vector $u^N \in \mathbb{R}^N$, which is, however, not suitable for examining the asymptotic behaviour of regularization functionals. Due to this reason, we identify a vector $u^N \in \mathbb{R}^N$ with a finite Radon measure $\mu^N \in \mathcal{M}(\Omega)$. For this purpose, μ^N has to be absolutely continuous with respect to the Lebesgue measure λ, i.e. every Borel set $A \in \mathcal{B}(\Omega)$ with $\lambda(A) = 0$ is supposed to be μ^N-null. Furthermore, μ^N has the density function

$$\rho^N = \sum_{j=1}^{N} u_j^N \chi_{\Omega_j^N} \,.$$

Remark 7.1.
We can easily see that

$$\mu^N(\Omega) = \sum_{j=1}^{N} u_j^N |\Omega_j^N| \qquad \text{and} \qquad \mu^N(\Omega_j^N) = u_j^N |\Omega_j^N|$$

holds, since we have

$$\mu^N(\Omega) = \int_\Omega \rho^N \,\mathrm{d}\lambda = \int_\Omega \sum_{j=1}^{N} u_j^N \chi_{\Omega_j^N} \,\mathrm{d}\lambda = \sum_{j=1}^{N} u_j^N \int_{\Omega_j^N} \mathrm{d}\lambda = \sum_{j=1}^{N} u_j^N |\Omega_j^N| \,.$$

In (7.1) we now use $\mathcal{U} = \mathcal{M}(\Omega)$ and hence consider the operator $\mathcal{K}\colon \mathcal{M}(\Omega) \longrightarrow \mathcal{H}$, which shall satisfy $\mathcal{K} = \mathcal{A}^*$ as it was done by Bredies and Pikkarainen (2013). Here $\mathcal{A}\colon \mathcal{H} \longrightarrow C_0(\Omega)$ is a compact operator and thus $\mathcal{K}$ is compact from the weak-*topology to the strong topology of $\mathcal{H}$, cf. also Subsection 2.4.3. Therefore, we redefine all functionals on the subset

$$\mathcal{M}_C := \{\mu \in \mathcal{M}(\Omega) \mid \|\mu\|_{\mathrm{TV}(\Omega)} \le C\}$$

to avoid technical difficulties of the weak-*convergence. However, this is not a crucial restriction, since an upper bound on the total variation of the measure is natural in inverse problems. In doing so, we are able to use the metrizability of the weak-*convergence on bounded sets. Furthermore, we obtain weak-*compactness.

Let us now concentrate on the asymptotics of regularization functionals promoting spatial sparsity.
For $p > 0$ we define the weighted p-norms as

$$\mathcal{R}_p^N(\mu) := \begin{cases} \left\|u^N\right\|_p = \left(\sum_{j=1}^N \omega_j^N |u_j^N|^p\right)^{\frac{1}{p}}, & \text{if } \mu = \sum_{j=1}^N u_j^N |\Omega_j^N|\,, \\ \infty\,, & \text{else,} \end{cases} \tag{7.2}$$

with a positive weight vector $\omega^N \in \mathbb{R}^N$. In an analogous manner we define

$$\mathcal{R}_0^N(\mu) := \begin{cases} \sum_{\substack{j \in \{1,\dots,N| \\ u_j^N \neq 0\}}} \omega_j^N\,, & \text{if } \mu = \sum_{j=1}^N u_j^N |\Omega_j^N|\,, \\ \infty\,, & \text{else.} \end{cases} \tag{7.3}$$

The weight vector $\omega^N \in \mathbb{R}^N$ introduces a scaling of the functional and should depend on $|\Omega_j^N|$ in order to obtain reasonable limits of (7.2) and (7.3).

7.2.1. Convergence for the Convex Case $p \geq 1$

Let us begin the analysis of asymptotic behavior of spatial sparsity by considering the convex case. For appropriate scaling we obtain a rather intuitive limit.

Theorem 7.1.
Let be $p \geq 1$ and $\omega_j^N := |\Omega_j^N|$ for every $j \in \{1, \dots, N\}$. Then $\mathcal{R}_p^N$ Γ-converges with respect to the weak-*topology of $\mathcal{M}(\Omega)$ to

$$\mathcal{R}_1^\infty(\mu) = \|\mu\|_{\mathrm{TV}(\Omega)} \tag{7.4}$$

if we have $p = 1$ and to

$$\mathcal{R}_p^\infty(\mu) = \begin{cases} \|u\|_{L^p(\Omega)}\,, & \text{if } \mu = \lambda(u),\ u \in L^p(\Omega), \\ \infty, & \text{else,} \end{cases} \tag{7.5}$$

in case that $p > 1$ holds.

Proof.

Equi-coercivity:

By construction, equi-coercivity holds in $\mathcal{M}(\Omega)$ as well as in the weak topology of $L^p(\Omega)$ for $p > 1$.

Lower bound inequality:

$\mathcal{R}_p^\infty$ is lower semi-continuous in the weak-*topology, i.e. it holds that

$$\mathcal{R}_p^\infty(\mu) \leq \liminf_{N\to\infty} \mathcal{R}_p^\infty(\mu^N) \qquad \forall\, \mu^N \rightharpoonup^* \mu\ .$$

Let us now show that

$$\liminf_{N\to\infty} \mathcal{R}_p^\infty(\mu^N) \leq \liminf_{N\to\infty} \mathcal{R}_p^N(\mu^N)$$

holds. By definition we have

$$\mathcal{R}_p^N(\mu^N) = \left(\sum_{j=1}^N |\Omega_j^N| |u_j^N|^p \right)^{\frac{1}{p}} \ ,$$

which already equals $\mathcal{R}_p^\infty(\mu^N)$. Thus we obtain

$$\mathcal{R}_p^\infty(\mu) \leq \liminf_{N\to\infty} \mathcal{R}_p^N(\mu^N) \qquad \forall\, \mu^N \rightharpoonup^* \mu\ .$$

Upper bound inequality:

Due to Theorem 2.3 it is sufficient to restrict ourselves to the case of a dense subspace of $\mathcal{M}(\Omega)$ with a stronger topology, where we only consider measures with total variation bounded by C. On account of this, we choose the subspace of measures with continuous densities $u \in L^p(\Omega)$. Hence we are able to construct the following approximation

$$u_j^N \approx \frac{\int_{\Omega_j^N} u \,\mathrm{d}\lambda}{\int_{\Omega_j^N} \mathrm{d}\lambda} = \frac{\tilde{\mu}^N(\Omega_j^N)}{|\Omega_j^N|} \ , \tag{7.6}$$

where $\tilde{\mu}^N$ belongs to the considered dense subspace. Thus we have $\mu^N \approx \tilde{\mu}^N$. Since $u \in L^p(\Omega)$ holds, we have strong L^p-convergence of the piecewise constant approximations. Due to the fact that $\mathcal{R}_p^N(\tilde{\mu}^N) = \mathcal{R}_p^\infty(\tilde{\mu}^N)$ holds, we obtain the assertion by utilizing the strong continuity of $\mathcal{R}_p^\infty$ in $L^p(\Omega)$, i.e.

$$\limsup_{N\to\infty} \mathcal{R}_p^N(\tilde{\mu}^N) = \limsup_{N\to\infty} \mathcal{R}_p^\infty(\tilde{\mu}^N) \leq \mathcal{R}_p^\infty(\tilde{\mu})$$

for all measures $\tilde{\mu} \in \mathcal{M}(\Omega)$ with densities $u \in L^p(\Omega)$ and $\tilde{\mu}^N \rightharpoonup^* \tilde{\mu}$.

□

Let us now consider each case with a faster and slower scaling of the weights. We will see that this already yields trivial limits.

Theorem 7.2.
Let $p \geq 1$ hold.

a) For the weight vector $\omega^N \in \mathbb{R}^N$ the following shall be true:

$$\lim_{N\to\infty} \max_{j\in\{1,\ldots,N\}} \frac{\omega_j^N}{|\Omega_j^N|} = 0 \ . \tag{7.7}$$

Then $\mathcal{R}_p^N$ Γ-converges to

$$\mathcal{R}_p^\infty(\mu) \equiv 0$$

with respect to the weak-*topology of $\mathcal{M}(\Omega)$.

b) Let now for the weight vector $\omega_j^N \in \mathbb{R}^N$ hold that

$$\lim_{N\to\infty} \max_{j\in\{1,\ldots,N\}} \frac{|\Omega_j^N|}{\omega_j^N} = 0 \ . \tag{7.8}$$

Then we obtain Γ-convergence of $\mathcal{R}_p^N$ to

$$\mathcal{R}_p^\infty(\mu) = \begin{cases} 0, & \text{if } \mu = 0, \\ \infty, & \text{else,} \end{cases}$$

with respect to the weak-*topology of $\mathcal{M}(\Omega)$.

Proof.

a) Faster scaling:

Equi-coercivity:

holds by construction.

Lower bound inequality:

Since the weight vector ω^N is always positive by definition, it holds for all converging sequences $\mu^N \rightharpoonup^* \mu$ in $\mathcal{M}(\Omega)$ that

$$0 \equiv \mathcal{R}_p^\infty(\mu) \leq \liminf_{N\to\infty} \mathcal{R}_p^N(\mu^N) \ .$$

Upper bound inequality:

For measures with bounded total variation, we restrict ourselves to a dense subspace of $\mathcal{M}(\Omega)$ with a stronger topology, cf. Theorem 2.3. We choose the subspace of measures with continuous densities $u \in L^p(\Omega)$ and consider

again the approximation (7.6).
Since $\mu^N = \sum_{j=1}^N u_j^N |\Omega_j^N|$ holds, we have

$$\limsup_{N\to\infty} \mathcal{R}_p^N(\mu^N) = \limsup_{N\to\infty} \left(\sum_{j=1}^N \omega_j^N |u_j^N|^p \right)^{\frac{1}{p}} \approx \limsup_{N\to\infty} \left(\sum_{j=1}^N \omega_j^N \left| \frac{\tilde{\mu}^N(\Omega_j^N)}{|\Omega_j^N|} \right|^p \right)^{\frac{1}{p}}$$
$$= \limsup_{N\to\infty} \left(\sum_{j=1}^N \frac{\omega_j^N}{|\Omega_j^N|^p} \left| \int_{\Omega_j^N} u \,\mathrm{d}\lambda \right|^p \right)^{\frac{1}{p}} .$$

By using Hölder's inequality with $\frac{1}{p} + \frac{1}{p_*} = 1$, we obtain

$$\limsup_{N\to\infty} \mathcal{R}_p^N(\mu^N) \lesssim \limsup_{N\to\infty} \left(\sum_{j=1}^N \frac{\omega_j^N}{|\Omega_j^N|^p} \left(\int_{\Omega_j^N} u^p \,\mathrm{d}\lambda \right) \left(\int_{\Omega_j^N} \mathrm{d}\lambda \right)^{\frac{p}{p_*}} \right)^{\frac{1}{p}}$$
$$= \limsup_{N\to\infty} \left(\sum_{j=1}^N \frac{\omega_j^N}{|\Omega_j^N|} \int_{\Omega_j^N} u^p \,\mathrm{d}\lambda \right)^{\frac{1}{p}} ,$$

since we have $\frac{p}{p_*} = p - 1$ and the latter integral is equal to the volume of Ω_j^N. Note that for $p = 1$ Hölder's inequality is indeed not applicable, however, in that case this step is unnecessary. Finally, we have

$$\limsup_{N\to\infty} \mathcal{R}_p^N(\mu^N) \lesssim \limsup_{N\to\infty} \left(\max_{j\in\{1,\dots,N\}} \frac{\omega_j^N}{|\Omega_j^N|} \sum_{j=1}^N \int_{\Omega_j^N} u^p \,\mathrm{d}\lambda \right)^{\frac{1}{p}} = 0 ,$$

due to the faster scaling (7.7) of the weight vectors and since it is $u \in L^p(\Omega)$.

b) Slower scaling:

Equi-coercivity:

holds by construction.

Lower bound inequality:

For $\mu = 0$ the case is trivial. Let thus $\mathcal{R}_p^\infty(\mu) = \infty$ hold. Then we have to show that

$$\liminf_{N\to\infty} \mathcal{R}_p^N(\mu^N) = \infty .$$

Let us assume that there exists a constant $C \in \mathbb{R}$ such that

$$\mathcal{R}_p^N(\mu^N) = \left(\sum_{j=1}^N \omega_j^N |u_j^N|^p \right)^{\frac{1}{p}} \leq C$$

holds. For all $\varphi \in C(\Omega)$ we can consider

$$\left|\int_\Omega \varphi \,\mathrm{d}\mu^N\right| = \left|\sum_{j=1}^N \int_{\Omega_j^N} \varphi u_j^N \,\mathrm{d}\lambda\right| \le \|\varphi\|_\infty \sum_{j=1}^N |u_j^N||\Omega_j^N| \,.$$

For $p = 1$ we obtain

$$\begin{aligned}\left|\int_\Omega \varphi \,\mathrm{d}\mu^N\right| &\le \|\varphi\|_\infty \max_{j\in\{1,\dots,N\}} \frac{|\Omega_j^N|}{\omega_j^N} \sum_{j=1}^N \omega_j^N |u_j^N| \\ &\le \|\varphi\|_\infty \, C \max_{j\in\{1,\dots,N\}} \frac{|\Omega_j^N|}{\omega_j^N}\,,\end{aligned}$$

which, due to (7.8), goes to zero as N goes to infinity. This contradicts the weak-*convergence of $\mu^N \rightharpoonup^* \mu$, since we have considered $\mu \neq 0$.
Let now $p > 1$ hold. Then we have

$$\begin{aligned}\left|\int_\Omega \varphi \,\mathrm{d}\mu^N\right| &\le \|\varphi\|_\infty \sum_{j=1}^N (\omega_j^N)^{\frac{1}{p}} |u_j^N| \frac{|\Omega_j^N|}{(\omega_j^N)^{\frac{1}{p}}} \\ &\le \|\varphi\|_\infty \left(\sum_{j=1}^N \omega_j^N |u_j^N|^p\right)^{\frac{1}{p}} \left(\sum_{j=1}^N \frac{|\Omega_j^N|^{p_*}}{(\omega_j^N)^{\frac{p_*}{p}}}\right)^{\frac{1}{p_*}}\end{aligned}$$

by using Hölder's inequality with $\frac{1}{p} + \frac{1}{p_*} = 1$.
Due to the fact that $\frac{p_*}{p} = p_* - 1$ holds, we obtain

$$\begin{aligned}\left|\int_\Omega \varphi \,\mathrm{d}\mu^N\right| &\le \|\varphi\|_\infty \, C \left(\sum_{j=1}^N \frac{|\Omega_j^N|^{p_*-1}}{(\omega_j^N)^{p_*-1}} |\Omega_j^N|\right)^{\frac{1}{p_*}} \\ &\le \|\varphi\|_\infty \, C \left(\left(\max_{j\in\{1,\dots,N\}} \frac{|\Omega_j^N|}{\omega_j^N}\right)^{p_*-1} \sum_{j=1}^N |\Omega_j^N|\right)^{\frac{1}{p_*}} \\ &\le \|\varphi\|_\infty \, C |\Omega|^{\frac{1}{p_*}} \left(\max_{j\in\{1,\dots,N\}} \frac{|\Omega_j^N|}{\omega_j^N}\right)^{\frac{1}{p}}\,,\end{aligned}$$

which goes to zero as well as N goes to infinity due to (7.8). Thus we obtain again a contradiction to the weak-*convergence of $\mu^N \rightharpoonup^* \mu$. Therefore, $\mathcal{R}_p^N(\mu^N)$ is not bounded, which yields the assertion.

Upper bound inequality:

It trivially holds that for every $\mu \in \mathcal{M}(\Omega)$ there exists a sequence $(\mu^N)_{N\in\mathbb{N}}$ with $\mu^N \rightharpoonup^* \mu$ such that

$$\infty \equiv \mathcal{R}_p^\infty(\mu) \geq \limsup_{N\to\infty} \mathcal{R}_p^N(\mu^N) \, .$$

□

7.2.2. Convergence for the Nonconvex Case $p = 0$

Let us now consider a special case, where the functional is nonconvex, i.e. we assume that $p = 0$ holds. Let us analyze the functional $\mathcal{R}_0^N$ without any weighting, i.e. in case that $\omega^N = \mathbb{1}_N$ holds for the weight vector.

Theorem 7.3.
Let $\omega_j^N = 1$ hold for every $j \in \{1, \dots, N\}$. Then $\mathcal{R}_0^N$ Γ-converges to

$$\mathcal{R}_0^\infty(\mu) = \begin{cases} M, & \text{if } \mu = \sum\limits_{i=1}^{M} c_i \delta_{X_i} \text{ for } c \in \mathbb{R}^M,\ X \in \Omega^M \text{ with pairwise different entries,} \\ \infty, & \text{else,} \end{cases}$$

with respect to the weak-*topology of $\mathcal{M}(\Omega)$, where M is the number of nonzero spikes X_i with height c_i.

Proof.
Equi-coercivity:

Due to the fact that all minimizers of $\mathcal{R}_0^N$ are in $\mathcal{M}_C$, we are able to deduce equi-coercivity in the weak-*topology by applying the Banach-Alaoglu Theorem 2.2.

Lower bound inequality:

- Let us first consider the case, where we have $\mu = \sum\limits_{i=1}^{M} c_i \delta_{X_i}$ with pairwise different entries X_i and $\sum\limits_{i=1}^{M} |c_i| \leq C$ holds, i.e. we have $\mathcal{R}_0^\infty(\mu) = M$. Then we have to show that

  $$\liminf_{N\to\infty} \mathcal{R}_0^N(\mu^N) \geq M \qquad \text{for} \qquad \mu^N \rightharpoonup^* \mu \, .$$

  Let us assume that there exists a sequence in $\mathcal{M}_C$ with $\mu^N \rightharpoonup^* \mu$ and

  $$\liminf_{N\to\infty} \mathcal{R}_0^N(\mu^N) < M \, .$$

Due to the fact that the values of $\mathcal{R}_0^N$ are discrete, there exists an $N_* \in \mathbb{N}$ such that

$$\mathcal{R}_0^N(\mu^N) \leq M - 1 \qquad \text{for} \qquad N > N_* \tag{7.9}$$

holds. Now let ε be less than half the distance between two spikes, i.e.

$$\varepsilon < \frac{1}{2}|X_i - X_j| \ .$$

In addition choose $\varphi_i \in C_0(\Omega)$ with support in $B_\varepsilon(X_i)$ and $\varphi(X_i) = \text{sign}\,(c_i)$, cf. Figure 7.1.

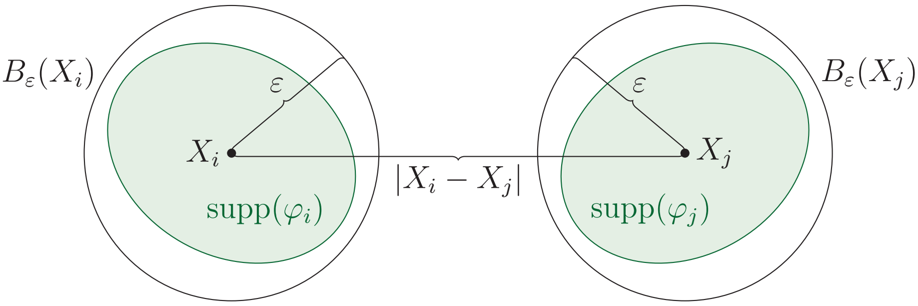


Figure 7.1.: Illustration of a part of the proof for the convergence of the ℓ^0-functional

Since we can choose N sufficiently large, we obtain $h_N < \varepsilon$ and thus each X_i is located in a different cell $\Omega^N_{j(i;N)}$. Due to this argument and the fact that (7.9) holds, we deduce that there exists an $i(N)$ such that we obtain

$$\left\langle \mu^N, \varphi_{i(N)} \right\rangle = 0$$

for N sufficiently large. Moreover, we are able to choose a subsequence N_k such that $I = i(N_k)$ is constant, since we only have a finite number of indices. Hence we obtain

$$\left\langle \mu^{N_k}, \varphi_I \right\rangle = 0 \neq |c_I| = \langle \mu, \varphi_I \rangle \ ,$$

which is a contradiction to the weak-*convergence of μ^N.

- Let us now consider $\mathcal{R}_0^\infty(\mu) = \infty$, i.e. μ shall not be concentrated on a set of finite points. We have to show that

$$\liminf_{N\to\infty} \mathcal{R}_0^N(\mu^N) = \infty$$

holds. Let us assume there exists a sequence $\mu^N \rightharpoonup^* \mu$ such that $\mathcal{R}_0^N(\mu^N)$

has a bounded subsequence, i.e.

$$\mathcal{R}_0^{N_k}(\mu^{N_k}) \leq M \tag{7.10}$$

for $M \in \mathbb{N}$. Due to the fact that the support of μ is not concentrated on a finite number of points, there exist $M+1$ continuous functions $\varphi_i \in C_0(\Omega)$, $i \in \{1, \ldots, M+1\}$, with disjoint support such that

$$\langle \mu, \varphi_i \rangle > 0$$

for all $i \in \{1, \ldots, M+1\}$. Since (7.10) holds, we see that there exists an index $i(N_k) \in \{1, \ldots, M+1\}$ such that

$$\left\langle \mu^{N_k}, \varphi_{i(N^k)} \right\rangle = 0$$

holds for N^k sufficiently large. For this reason there exists an index $i \in \{1, \ldots, M+1\}$ and a subsequence N_k', for which we have

$$\left\langle \mu^{N_k'}, \varphi_i \right\rangle = 0 \ .$$

As before, this is a contradiction to the weak-*convergence of $\mu^{N_k'}$.

Upper bound inequality:

We have to show that for all $\mu \in \mathcal{M}(\Omega)$ there exists a sequence $(\mu^N)_{N\in\mathbb{N}}$ such that

$$\limsup_{N\to\infty} \mathcal{R}_0^N(\mu^N) \leq \mathcal{R}_0^\infty(\mu) \qquad \text{with} \qquad \mu^N \rightharpoonup^* \mu \ .$$

- Let $\mu = \sum\limits_{i=1}^{M} c_i \delta_{X_i}$ hold, i.e. we have $\mathcal{R}_0^\infty(\mu) = M$. Let us assume that there exists a sequence $\mu^N \rightharpoonup^* \mu$ for $N \to \infty$ with

$$\limsup_{N\to\infty} \mathcal{R}_0^N(\mu^N) > M \ .$$

Since the values of $\mathcal{R}_0^N$ are discrete, there exists an $N_* \in \mathbb{N}$ such that

$$\mathcal{R}_0^N(\mu^N) \geq M+1 \qquad \text{for} \qquad N > N_* \ . \tag{7.11}$$

holds. Thus the discrete functional $\mathcal{R}_0^N$ has at least one nonzero element more than the continuous functional $\mathcal{R}_0^\infty$ has nonzero spikes. For N sufficiently large we consider again the previous construction and use the cells $\Omega_{j(i;N)}^N$ as defined above. Additionally we define

$$u_j^N := \begin{cases} \mu(\Omega^N_{j(i;N)}), & \text{if } j = j(i;N), \\ 0, & \text{else,} \end{cases}$$

which yields $\mu^N = \sum_{j=j(i;N)} \mu(\Omega^N_{j(i;N)})$. Let us again consider a function $\varphi \in C_0(\Omega)$ with $X_i \in \text{supp}(\varphi_i) \subset B_\varepsilon(X_i)$ and $\varphi(X_i) = \text{sign}\,(c_i)$. Due to (7.11), we can find an index $i(N)$ such that $X_{i(N)}$ is nonzero in the discrete version, but zero in the continuous setting. Then we obtain

$$\left\langle \mu^N, \varphi_{i(N)} \right\rangle = \mu(\Omega^N_{j(i(N);N)}) = 0 \ ,$$

which contradicts the fact that $X_{i(N)}$ was supposed to be nonzero in the discrete setting.

- For the case that μ is not concentrated on a set of finite points, i.e. $\mathcal{R}_0^\infty(\mu) = \infty$, the construction follows directly from the lower bound inequality.

□

7.3. Asymptotics of Mixed Norms

In the last section we considered different regularization functionals for vectors, namely ℓ^p-norms for $p \geq 1$ and $p = 0$, and analyzed their behaviour for the case that the discretization becomes finer and finer, until we discovered asymptotic limits.
In this section we consider matrices instead of vectors and thus analyze the asymptotics of mixed norms (cf. Chapter 3). For this purpose, we consider two disjoint partitions of sets $\widetilde{\Omega},\ \Omega \subset \mathbb{R}^d$, i.e.

$$\widetilde{\Omega} = \bigcup_{i=1}^{M} \widetilde{\Omega}_i^M \qquad \text{and} \qquad \Omega = \bigcup_{j=1}^{N} \Omega_j^N \ ,$$

and the measurable space $\left(\Omega \times \widetilde{\Omega},\ \mathcal{B}(\Omega) \otimes \mathcal{B}(\widetilde{\Omega})\right)$.
In this setting we can identify every matrix $U^{MN} \in \mathbb{R}^{M \times N}$ with a finite Radon measure $\mu^{MN} \in \mathcal{M}(\Omega \times \widetilde{\Omega})$, which has the density function

$$\rho^{MN} := \sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij}^{MN} \chi_{\Omega_j^N} \chi_{\widetilde{\Omega}_i^M} \ .$$

Remark 7.2.
We can see that the following identification is true:

$$\mu^{MN}(\Omega\times\widetilde{\Omega})=\sum_{i=1}^{M}\sum_{j=1}^{N}u_{ij}^{MN}|\Omega_j^N||\widetilde{\Omega}_i^M| \quad\text{and}\quad \mu^{MN}(\Omega_j^N\times\widetilde{\Omega}_i^M)=u_{ij}|\Omega_j^N||\widetilde{\Omega}_i^M|\,.$$

The first identity is obtained by computing:

$$\begin{aligned}\mu^{MN}(\Omega\times\widetilde{\Omega}) &= \int_{\widetilde{\Omega}}\int_{\Omega}\rho^{MN}\,\mathrm{d}\lambda_\Omega\,\mathrm{d}\lambda_{\widetilde{\Omega}}\\ &= \int_{\widetilde{\Omega}}\int_{\Omega}\sum_{i=1}^{M}\sum_{j=1}^{N}u_{ij}^{MN}\chi_{\Omega_j^N}\chi_{\widetilde{\Omega}_i^M}\,\mathrm{d}\lambda_\Omega\,\mathrm{d}\lambda_{\widetilde{\Omega}}\\ &= \int_{\widetilde{\Omega}}\sum_{i=1}^{M}\sum_{j=1}^{N}u_{ij}^{MN}\chi_{\widetilde{\Omega}_i^M}\int_{\Omega_j^N}\mathrm{d}\lambda_\Omega\,\mathrm{d}\lambda_{\widetilde{\Omega}}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{N}u_{ij}^{MN}|\Omega_j^N|\int_{\widetilde{\Omega}_i^M}\mathrm{d}\lambda_{\widetilde{\Omega}}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{N}u_{ij}^{MN}|\Omega_j^N||\widetilde{\Omega}_i^M|\,.\end{aligned}$$

The second identity can be computed in a similar way.

By using this identification, we are now able to work on the space $\mathcal{M}(\Omega\times\widetilde{\Omega})$. Let us thus consider the inverse problem

$$\mathcal{K}\mu=w\,,$$

where the operator $\mathcal{K}\colon\mathcal{M}(\Omega\times\widetilde{\Omega})\longrightarrow\mathcal{H}$ shall satisfy $\mathcal{K}=\mathcal{A}^*$ and $\mathcal{A}\colon\mathcal{H}\longrightarrow C_0(\Omega\times\widetilde{\Omega})$ is assumed to be a compact operator. As before, we will redefine all functionals on

$$\mathcal{M}'_C:=\left\{\mu\in\mathcal{M}(\Omega\times\widetilde{\Omega})\;\middle|\;\|\mu\|_{\mathrm{TV}(\Omega\times\widetilde{\Omega})}\le C\right\}$$

in order to avoid technical difficulties of the weak-*convergence.

Let us now analyze the following regularization functional:

$$\mathcal{R}_{p,q}^{MN}(\mu):=\begin{cases}\left(\sum\limits_{i=1}^{M}\left(\sum\limits_{j=1}^{N}w_{ij}^{MN}\left|u_{ij}^{MN}\right|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}, & \text{if } \mu=\sum\limits_{i=1}^{M}\sum\limits_{j=1}^{N}u_{ij}^{MN}|\Omega_j^N||\widetilde{\Omega}_i^M|,\\ \infty\,, & \text{else.}\end{cases} \tag{7.12}$$

7.3.1. Convergence for the General Convex Case $p > 1$ and $q > 1$

Let us start by examining the convex case, where p and q are larger than one. Afterwards, we separately consider the special case of local sparsity, i.e. where p is equal to one.

Theorem 7.4.
Let be p, $q > 1$ and define $\omega_{ij}^{MN} := |\Omega_j^N||\widetilde{\Omega}_i^M|^{\frac{p}{q}}$ for all $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, N\}$. Then $\mathcal{R}_{p,q}^{MN}$ Γ-converges with respect to the weak-*topology of $\mathcal{M}(\Omega \times \widetilde{\Omega})$ for $M, N \to \infty$ to

$$\mathcal{R}_{p,q}^{\infty}(\mu) = \begin{cases} \|u\|_{L^{p,q}(\Omega\times\widetilde{\Omega})}, & \text{if } \mu = \lambda_{\Omega\times\widetilde{\Omega}}(u),\ u \in\ L^{p,q}(\Omega \times \widetilde{\Omega}), \\ \infty, & \text{else.} \end{cases} \tag{7.13}$$

The proof works similarly to the proof of Theorem 7.1.

Proof.

Equi-coercivity:
holds by construction.

Lower bound inequality:
Since in the nontrivial case $\mathcal{R}_{p,q}^{\infty}$ is the composition of the L^p- and the L^q-norm, it is lower semi-continuous, i.e. it holds that

$$\mathcal{R}_{p,q}^{\infty}(\mu) \leq \liminf_{M,N\to\infty} \mathcal{R}_{p,q}^{\infty}(\mu^{MN}) \qquad \forall\ \mu^{MN} \rightharpoonup^* \mu\ .$$

We still need to show that

$$\liminf_{M,N\to\infty} \mathcal{R}_{p,q}^{\infty}(\mu^{MN}) \leq \liminf_{M,N\to\infty} \mathcal{R}_{p,q}^{MN}(\mu^{MN})$$

holds. The case, where we have

$$\mu^{MN} \neq \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}^{MN}|\Omega_j^N||\widetilde{\Omega}_i^M|\ ,$$

is trivial, since then $\mathcal{R}_{p,q}^{MN}(\mu^{MN})$ is already infinity. In the other case, however, we have by definition

$$\mathcal{R}_{p,q}^{MN}(\mu^{MN}) = \left(\sum_{i=1}^{M}|\widetilde{\Omega}_i^M|\left(\sum_{j=1}^{N}|\Omega_j^N|\,\left|u_{ij}^{MN}\right|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}\ ,$$

which already equals $\mathcal{R}^{\infty}_{p,q}(\mu^{MN})$. Hence we obtain

$$\mathcal{R}^{\infty}_{p,q}(\mu) \leq \liminf_{M,\,N\to\infty} \mathcal{R}^{MN}_{p,q}(\mu^{MN}) \qquad \forall\, \mu^{MN} \rightharpoonup^{*} \mu \,.$$

Upper bound inequality:

On the basis of Theorem 2.3, we restrict ourselves to the case of a dense subspace of $\mathcal{M}(\Omega\times\widetilde{\Omega})$ with a stronger topology. In this context let us consider the subspace of measures with continuous densities $u \in L^{p,q}(\Omega\times\widetilde{\Omega})$. In so doing, we shall consider the approximation

$$u^{MN}_{ij} \approx \frac{\int_{\widetilde{\Omega}^M_i}\int_{\Omega^N_j} u \,\mathrm{d}\lambda_\Omega \mathrm{d}\lambda_{\widetilde{\Omega}}}{\int_{\widetilde{\Omega}^M_i}\int_{\Omega^N_j} \mathrm{d}\lambda_\Omega \mathrm{d}\lambda_{\widetilde{\Omega}}} = \frac{\tilde{\mu}^{MN}(\widetilde{\Omega}^M_i \times \Omega^N_j)}{|\widetilde{\Omega}^M_i||\Omega^N_j|}\,, \tag{7.14}$$

where $\tilde{\mu}^{MN}$ is an element of the considered dense subspace. Hence we obtain the approximation $\mu^{MN} \approx \tilde{\mu}^{MN}$. Since u belongs to $L^{p,q}(\Omega\times\widetilde{\Omega})$, we observe that the piecewise constant approximations converge strongly with respect to the $L^{p,q}$-norm. By using the fact that we have $\mathcal{R}^{MN}_{p,q}(\tilde{\mu}^{MN}) = \mathcal{R}^{\infty}_{p,q}(\tilde{\mu}^{MN})$ and $\mathcal{R}^{\infty}_{p,q}$ is strongly continuous in $L^{p,q}(\Omega\times\widetilde{\Omega})$, we obtain

$$\limsup_{M,\,N\to\infty} \mathcal{R}^{MN}_{p,q}(\tilde{\mu}^{MN}) = \limsup_{M,\,N\to\infty} \mathcal{R}^{\infty}_{p,q}(\tilde{\mu}^{MN}) \leq \mathcal{R}^{\infty}_{p,q}(\tilde{\mu})$$

for all measures $\tilde{\mu} \in \mathcal{M}(\Omega\times\widetilde{\Omega})$ with densities $u \in L^{p,q}(\Omega\times\widetilde{\Omega})$ and $\tilde{\mu}^{MN} \rightharpoonup^{*} \tilde{\mu}$.

□

7.3.2. Asymptotic Local Sparsity

Let us now consider the case, where p is equal to one. First we consider the Γ-limit in the semi-continuous case.

Theorem 7.5 (Semi-Continuous Local Sparsity)**.**
Let be $p = 1$ and $q > 1$. Let the weights be defined as $\omega^{MN}_{ij} := |\Omega^N_j||\widetilde{\Omega}^M_i|^{\frac{1}{q}}$ for $i \in \{1,\dots,M\}$ and $j \in \{1,\dots,N\}$. Then $\mathcal{R}^{MN}_{1,q}$ Γ-converges with respect to the weak-*topology of $\mathcal{M}^N(\widetilde{\Omega})$ for $M\to\infty$ to

$$\mathcal{R}^{\infty N}_{1,q}(\mu) = \begin{cases} \left\| \sum\limits_{j=1}^{N} |\Omega^N_j||u^N_j| \right\|_{L^q(\widetilde{\Omega})}, & \text{if } \mu = \sum\limits_{j=1}^{N} |\Omega^N_j|\,\lambda_{\widetilde{\Omega}}(u^N_j),\ u^N_j \in L^q(\widetilde{\Omega})\;\forall\, j, \\ \infty, & \text{else.} \end{cases}$$

Proof.
We begin by making the following definition:

$$z_i^{MN} := \sum_{j=1}^{N} |\Omega_j^N||u_{ij}^{MN}|$$

for every $N \in \mathbb{N}$. By inserting this in the definition of $\mathcal{R}_{1,q}^{MN}$, we obtain for every $N \in \mathbb{N}$ that

$$\mathcal{R}_{1,q}^{MN} = \begin{cases} \left(\sum_{i=1}^{M} |\widetilde{\Omega}_i^M||z_i^{MN}|^q\right)^{\frac{1}{q}}, & \text{if } \mu = \sum_{i=1}^{M} z_i^{MN}|\widetilde{\Omega}_i^M| \\ \infty, & \text{else,} \end{cases} \tag{7.15}$$

holds. Now we can apply Theorem 7.1, which states that (7.15) Γ-converges with respect to the weak-*topology of $\mathcal{M}^N(\widetilde{\Omega})$ to

$$\mathcal{R}_{1,q}^{\infty N} = \begin{cases} \left\| z^N \right\|_{L^q(\widetilde{\Omega})}, & \text{if } \mu = \lambda_{\widetilde{\Omega}}(z^N),\ z^N \in L^q(\widetilde{\Omega}), \\ \infty, & \text{else.} \end{cases}$$

By replacing z^N again, we obtain the semi-continuous Γ-limit

$$\mathcal{R}_{1,q}^{\infty N}(\mu) = \begin{cases} \left\| \sum_{j=1}^{N} |\Omega_j^N||u_j^N| \right\|_{L^q(\widetilde{\Omega})}, & \text{if } \mu = \sum_{j=1}^{N} |\Omega_j^N|\, \lambda_{\widetilde{\Omega}}(u_j^N),\ u_j^N \in L^q(\widetilde{\Omega})\ \forall\, j, \\ \infty, & \text{else.} \end{cases}$$

□

Now let us analyze the continuous case.

Theorem 7.6 (Continuous Local Sparsity)**.**
Let be $p = 1$ and $q > 1$. Furthermore, let the weights be defined as $\omega_{ij}^{MN} := |\Omega_j^N||\widetilde{\Omega}_i^M|^{\frac{1}{q}}$ for $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N\}$. Then $\mathcal{R}_{1,q}^{MN}$ Γ-converges with respect to the weak-*topology of $\mathcal{M}(\Omega \times \widetilde{\Omega})$ for $M, N \to \infty$ to

$$\mathcal{R}_{1,q}^{\infty}(\mu) = \begin{cases} \left(\int_{\widetilde{\Omega}} \|\mu(y)\|_{\mathrm{TV}(\Omega)}^q \, \mathrm{d}y\right)^{\frac{1}{q}}, & \text{if } \mu \in L^q(\widetilde{\Omega}, \mathcal{M}(\Omega)), \\ \infty, & \text{else.} \end{cases}$$

The proof works analogously to the proof of Theorem 7.4.

Proof.

Equi-coercivity:

holds by construction.

Lower bound inequality:

Since in the nontrivial case $\mathcal{R}^{\infty}_{1,q}$ is the composition of the TV- and the L^q-norm, $\mathcal{R}^{\infty}_{1,q}$ is lower semi-continuous in the weak-*topology of $\mathcal{M}(\Omega \times \widetilde{\Omega})$, i.e.

$$\mathcal{R}^{\infty}_{1,q}(\mu) \leq \liminf_{M,\,N \to \infty} \mathcal{R}^{\infty}_{1,q}(\mu^{MN}) \qquad \forall\, \mu^{MN} \rightharpoonup^* \mu$$

holds for all sequences $(\mu^{MN})_{M,N\in\mathbb{N}}$ with $\mu^{MN} \in \mathcal{M}(\Omega \times \widetilde{\Omega})$ and $\mu^{MN} \rightharpoonup^* \mu$ for $M, N \to \infty$. It is left to show that

$$\liminf_{M,\,N \to \infty} \mathcal{R}^{\infty}_{1,q}(\mu^{MN}) \leq \liminf_{M,\,N \to \infty} \mathcal{R}^{MN}_{1,q}(\mu^{MN})$$

holds. In the case that

$$\mu^{MN} \neq \sum_{i=1}^{M} \sum_{j=1}^{N} u^{MN}_{ij} |\Omega^N_j| |\widetilde{\Omega}^M_i| \;,$$

holds, the inequality is trivially fulfilled, since then $\mathcal{R}^{MN}_{1,q}(\mu^{MN})$ is equal to infinity. Thus let us consider

$$\mu^{MN} = \sum_{i=1}^{M} \sum_{j=1}^{N} u^{MN}_{ij} |\Omega^N_j| |\widetilde{\Omega}^M_i| \;.$$

Then we have by definition

$$\mathcal{R}^{MN}_{1,q}(\mu^{MN}) = \left(\sum_{i=1}^{M} |\widetilde{\Omega}^M_i| \left(\sum_{j=1}^{N} |\Omega^N_j| \left| u^{MN}_{ij} \right| \right)^q \right)^{\frac{1}{q}} \;,$$

which already equals $\mathcal{R}^{\infty}_{1,q}(\mu^{MN})$. Therefore, we obtain

$$\mathcal{R}^{\infty}_{1,q}(\mu) \leq \liminf_{M,\,N \to \infty} \mathcal{R}^{MN}_{1,q}(\mu^{MN}) \qquad \forall\, \mu^{MN} \rightharpoonup^* \mu \;.$$

Upper bound inequality:

Due to Theorem 2.3 it is sufficient to consider the dense subspace of $\mathcal{M}(\Omega \times \widetilde{\Omega})$, which contains absolutely continuous measures with respect to Lebesgue densities $u \in L^{1,q}(\Omega \times \widetilde{\Omega})$. In this context, let us consider once more the approximation (7.14). Due to the fact that u belongs to $L^{1,q}(\Omega \times \widetilde{\Omega})$, the piecewise constant approximations $\tilde{\mu}^{MN}$ converge strongly with respect to the $L^{1,q}$-norm. Since the equality $\mathcal{R}_{1,q}^{MN}(\tilde{\mu}^{MN}) = \mathcal{R}_{1,q}^{\infty}(\tilde{\mu}^{MN})$ holds and $\mathcal{R}_{1,q}^{\infty}$ is strongly continuous in $L^{1,q}(\Omega \times \widetilde{\Omega})$, we obtain

$$\limsup_{M,\,N\to\infty} \mathcal{R}_{1,q}^{MN}(\tilde{\mu}^{MN}) = \limsup_{M,\,N\to\infty} \mathcal{R}_{1,q}^{\infty}(\tilde{\mu}^{MN}) \leq \mathcal{R}_{1,q}^{\infty}(\tilde{\mu})$$

for all measures $\tilde{\mu} \in \mathcal{M}(\Omega \times \widetilde{\Omega})$ with densities $u \in L^{1,q}(\Omega \times \widetilde{\Omega})$ and $\tilde{\mu}^{MN} \rightharpoonup^* \tilde{\mu}$. Thus the upper bound inequality holds as well.

□

8 Deconvolution of Sparse Spikes

In this chapter we want to verify our analytic results from Chapter 7. In order to do so, we consider the deconvolution of sparse spikes on a discrete grid as the step size of the grid becomes smaller. Predominantly in a continuous setting the deconvolution of sparse δ-spikes has for instance already been considered by Duval and Peyré (2013). In their paper they show that if the so-called *Non Degenerate Source Condition* holds and the signal-to-noise ratio is large enough, total variation regularization recovers the exact number of Dirac spikes. However, in the discretized setting on a grid this is usually not the case. Duval and Peyré show furthermore for the discrete setting that in case the support of the exact measure is contained in the grid and the non degenerate source condition is fulfilled, the support of the recovered solution contains the support of the exact measure and in addition possibly one immediate neighboring grid point if the step size is small enough.

In our considerations we use a different approach and validate our analysis numerically. First we consider the deconvolution of sparse δ-spikes, which are located at the grid points. Analytically as well as numerically we observe that the support can be reconstructed exactly, if the regularization parameter is chosen small enough. Furthermore, we are especially interested in the case, where the spikes are located in between the grid points. Analytically we show that if a spike is close enough to a grid point and the step size is small enough, we are able to recover a solution, which consists of the same number of peaks as the exact signal and the reconstructed peak is located at the grid point, which is closest to the position of the exact δ-spike. However, if the location of a spike is not close enough to a grid point, then it cannot be represented by a single reconstructed peak. Numerically we observe that as the step size of the grid becomes smaller, at most twice as many peaks are reconstructed as there are exact spikes.

8.1. Introduction

Let us consider the continuous inverse problem

$$\mathcal{K}\mu = w \ , \tag{8.1}$$

where $\mathcal{K}$ is a convolution operator with a symmetric kernel $\widetilde{G} \in C^3(\Omega)$. In this context $\widetilde{G}$ shall have its maximum at 0 and $\widetilde{G}(|x|)$ shall be strictly monotonically decreasing for $x \in \Omega$. We are interested in the scenario, where the exact solution $\hat{\mu}$ only consists of a few spikes at the points ξ_i for $i \in \{1, \ldots, S\}$ and is zero otherwise, i.e.

$$\hat{\mu} = \sum_{i=1}^{S} \rho_i \delta_{\xi_i} \ ,$$

with $S > 0$ small, $\rho \in \mathbb{R}^S$ and δ_{ξ_i} being the Dirac delta distribution, which is infinity at the point ξ_i and zero otherwise. For the solution of (8.1), we consider the variational model

$$\min_{\mu^N \in \mathbb{R}^N} \frac{1}{2} \left\| A\mu^N - w \right\|_2^2 + \alpha \left\| \mu^N \right\|_1 \tag{8.2}$$

with $A \in \mathbb{R}^{L \times N}$ being the discrete version of $\mathcal{K}$, $w \in \mathbb{R}^L$ and $N \geq S$, which builds the discrete counterpart to the continuous variational problem

$$\min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \left\| \mathcal{K}\mu - w \right\|_{L^2(\Omega)}^2 + \alpha \left\| \mu \right\|_{\mathrm{TV}(\Omega)} \ .$$

A minimizer of (8.2) can be written as

$$\mu^N = \sum_{i=1}^{N} c_i \delta_{x_i} \ ,$$

with $c \in \mathbb{R}^N$. Note that in the discrete setting δ_{x_i} denotes the Kronecker delta, which is equal to one at the grid point x_i and zero otherwise. Naturally, we would like to reconstruct a solution μ^N, where $c \in \mathbb{R}^N$ is close to being S-sparse. Moreover, we are interested in the distance between the positions of the reconstructed peaks and the positions of the spikes in the exact solution. Therefore, we define for this chapter x_k as the grid point, which is closest to an exact spike at position ξ_i, i.e.

$$x_k = \min_{\substack{x_n \in \mathbb{R} \\ n \in \{1, \ldots, N\}}} |x_n - \xi_i| \ .$$

We would like to understand, under which conditions the reconstructed solution μ^N contains a nonzero peak at the position x_k, while it is zero at the second closest grid point. In order to do so, we consider the easiest case, where the exact solution contains only one nonzero spike and is zero everywhere else, i.e.

$$\hat{\mu} = \delta_\xi \; .$$

8.2. Deconvolution of a Single δ-Spike

In this section we analyze the different positions of a continuous δ-spike at and between two grid points of a certain discretization, i.e. $\xi \in [x_k,\, x_{k+1}]$ with $k \in \{1, \ldots, N-1\}$. Before going into the details, we state some prior computations and definitions. First let us compute the optimality condition of (8.2):

$$\begin{aligned} 0 &= \left(A^* \left(A\mu^N - w\right)\right)_j + \alpha p_j \\ \Leftrightarrow \quad \alpha p_j &= (A^* w)_j - (A^* A \mu^N)_j \; , \end{aligned} \tag{8.3}$$

where $p \in \mathbb{R}^N$ is contained in the subdifferential of $\left\|\mu^N\right\|_1$, which can be computed componentwise, i.e. we have $p_j \in \partial|\mu_j^N|$ for every $j \in \{1, \ldots, N\}$. Since $\widetilde{G}$ is symmetric and the convolution is associative, we know that

$$A^* A \mu = \widetilde{G} * \left(\widetilde{G} * \mu\right) = \left(\widetilde{G} * \widetilde{G}\right) * \mu$$

holds. For the sake of simplicity, we define

$$G := \widetilde{G} * \widetilde{G}$$

in the whole section. Moreover, we have the equality

$$w_j = (A\hat{\mu})_j = \left(\widetilde{G} * \delta_\xi\right)_j = \widetilde{G}\left(x_j - \xi\right)$$

for every $j \in \{1, \ldots, N\}$, which yields that

$$(A^* w)_j = G(x_j - \xi)$$

holds for every $j \in \{1, \ldots, N\}$.

In order to prevent the reconstruction of the trivial solution $\mu^N \equiv 0$, let us first cite a result from Burger et al. (2013), which we state for our context.

Theorem 8.1.
Let $\frac{1}{\alpha}\|A^*w\|_\infty \leq 1$ hold. Then $\mu^N \equiv 0$ is a solution of (8.2).

Proof.
Let us consider the optimality condition (8.3). By inserting $\mu^N \equiv 0$ we obtain

$$p = \frac{1}{\alpha}A^*w \ ,$$

which is due to our assumptions less or equal to 1. Thus we obtain that $p \in \partial\|0\|_1$ holds and the optimality condition is fulfilled. Hence $\mu^N \equiv 0$ is a solution of (8.2). □

From this theorem we learn that we have to choose $\alpha < \|A^*w\|_\infty$ in order to reconstruct nontrivial solutions. In our setting we have

$$\|A^*w\|_\infty = \|\widetilde{G} * w\|_\infty = \|G * \hat{\mu}\|_\infty = \max_{j\in\{1,\dots,N\}} G(x_j - \xi) \ .$$

Thus we always have to consider regularisation parameters, for which

$$\alpha < \max_{j\in\{1,\dots,N\}} G(x_j - \xi)$$

holds.
Let us now analyze the case, where the position of the exact spike ξ coincides with a grid point.

Theorem 8.2.
Let be $\xi = x_k$ for one $k \in \{1,\dots,N\}$ and let $\alpha < G(0)$ hold in (8.2). Then there exists a 1-sparse solution μ^N of (8.2), which is nonzero at x_k and can be written as $\mu^N = c\delta_{x_k}$ with $c = \frac{G(0)-\alpha}{G(0)} \in (0,1)$.

Proof.
In order to proof the assertion, we have to check whether the optimality condition of (8.2) holds under the assumptions mentioned above. Due to our prior computations, the optimality condition (8.3) reduces to

$$\begin{aligned} \alpha p_j &= G(x_j - \xi) - cG(x_j - x_k) \\ &= (1-c)G(x_j - x_k) \ . \end{aligned} \tag{8.4}$$

We have to differentiate between the cases, where we have $j = k$ and $j \neq k$.
Thus let $j = k$ hold. In this case the optimality condition (8.4) reads as follows:

$$p_k = (1-c)\frac{G(0)}{\alpha} .$$

Due to the definition of c, we obtain that $p_k = 1$ holds and the optimality condition is fulfilled.
Let now be $j \neq k$. Then we have

$$p_j = \frac{1-c}{\alpha} G(x_j - x_k) < (1-c)\frac{G(0)}{\alpha} ,$$

due to the fact that $G(0)$ is the maximum of G. By inserting c, we obtain that $|p_j| < 1$ has to be true.
Hence in both cases the optimality condition is fulfilled and we obtain the assertion. □

From this theorem we learn that we are indeed able to reconstruct the support of a delta spike exactly if the position of the spike coincides with a grid point of our discretization and the regularization parameter is small enough. However, this is rarely the case. Thus let us consider the more frequent case, where ξ is located in between two grid points, i.e. $\xi \in (x_k, x_{k+1})$ holds. For the following consideration, the interval length h will be defined as

$$h := |x_{k+1} - x_k| .$$

Theorem 8.3.
Let be $\xi \in (x_k, x_k + \frac{h}{2})$ for $k \in \{1, \ldots, N-1\}$ and let $\alpha < G(x_k - \xi)$ hold. Furthermore, h shall be sufficiently small.
In case that we have $\xi \in (x_k,\, x_k + \frac{\alpha h}{2G(0)})$, there exists a 1-sparse solution of (8.2), which can be written as $\mu^N = c\delta_{x_k}$ with $c = \frac{G(x_k-\xi)-\alpha}{G(0)} \in (0,1)$.
Moreover, if we have $\xi \in (x_k + \frac{\alpha h}{2G(0)},\, x_k + \frac{h}{2})$, then $\mu^N = c\delta_{x_k}$ is *not* a solution of (8.2) for any $c \in \mathbb{R}$.

Figure 8.1 illustrates the assertion of Theorem 8.3. Note that due to the symmetry of G, the same claim holds for the other half of the interval.

Proof.
Once more we have to check the optimality condition (8.4).
Let us first consider the case, where $j = k$ holds. Since $p_k \in \partial|\mu_k^N| = \partial|c|$ and $c > 0$ holds, we obtain $p_k = 1$. Then the optimality condition (8.4) reduces to

$$\begin{aligned} \alpha &= G(x_k - \xi) - cG(0) \\ \Leftrightarrow \quad c &= \frac{G(x_k - \xi) - \alpha}{G(0)} . \end{aligned} \tag{8.5}$$

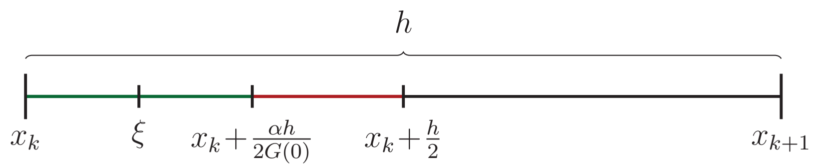


Figure 8.1.: If ξ is in the green interval, the reconstructed solution μ^N consists of only one peak. In the case that ξ is located in the red interval, then one peak is not sufficient.

Thus the optimality condition is fulfilled.
Now let us consider the case, where $j = k + 1$ holds. Then the optimality condition reads as follows:

$$\alpha p_{k+1} = G(x_{k+1} - \xi) - cG(x_{k+1} - x_k) \, . \tag{8.6}$$

Inserting (8.5) into (8.6) yields

$$\alpha p_{k+1} = G(x_{k+1} - \xi) - \frac{G(x_k - \xi) - \alpha}{G(0)} G(h) \, .$$

For this equation we consider the second order Taylor expansion of G around zero. Note that $G'(0) = 0$ holds, due to the maximum of G in zero. In so doing, we obtain that

$$\begin{aligned} \alpha p_{k+1} = {} & G(0) + \frac{1}{2} G''(0)(x_{k+1} - \xi)^2 \\ & - \left(1 + \frac{G''(0)}{2G(0)}(x_k - \xi)^2\right)\left(G(0) + \frac{1}{2}G''(0)h^2\right) \\ & + \alpha + \frac{\alpha}{2G(0)} G''(0) h^2 + \mathcal{O}(h^3) \, , \end{aligned}$$

holds, which reduces to

$$p_{k+1} = 1 - \frac{1}{2\alpha} G''(0) \left(\underbrace{(x_k - \xi)^2 - (x_{k+1} - \xi)^2 + h^2\left(1 - \frac{\alpha}{G(0)}\right)}_{J} \right) + \mathcal{O}(h^3) \, .$$

Note that $G''(0) < 0$ holds, due to the maximum of G in zero. In order to obtain the inequality $|p_{k+1}| < 1$, which would yield the assertion, J has to be negative. This is true if and only if we have

$$x_k^2 - x_{k+1}^2 + 2\xi h < h^2 \left(\frac{\alpha}{G(0)} - 1 \right) \, .$$

This is equivalent to

$$\begin{aligned} & 2\xi h < h^2\left(\frac{\alpha}{G(0)}-1\right)+(x_{k+1}+x_k)h \\ \Leftrightarrow \quad & \xi < \frac{h}{2}\left(\frac{\alpha}{G(0)}-1\right)+\frac{1}{2}x_{k+1}+\frac{1}{2}x_k \\ \Leftrightarrow \quad & \xi - x_k < \frac{h}{2}\left(\frac{\alpha}{G(0)}-1\right)+\frac{h}{2} \,. \end{aligned}$$

Thus we obtain

$$\xi - x_k < \frac{\alpha h}{2G(0)} < \frac{1}{2}h \,,$$

since we have $\alpha < G(x_k-\xi)$ and $G(x_k-\xi) < G(0)$. This yields the assertion. □

8.3. Algorithm for Sparse Spikes Deconvolution

We want to validate our results numerically and solve the variational problem (8.2) for different grid sizes. Therefore, we need an algorithm for the reconstruction of μ^N, which we present in this section.

Similarly to Chapter 5, we base this algorithm on Boyd et al. (2010). However, since we elaborated on the deduction of similar algorithms in Chapter 5, we keep this section as concise as possible.

After splitting in the following way:

$$\min_{\mu^N,\, z^N\in\mathbb{R}^N} \frac{1}{2}\left\|A\mu^N - w\right\|_2^2 + \alpha\left\|z^N\right\|_1 \qquad \text{s. t.} \qquad z^N = \mu^N \,,$$

we consider the Lagrange functional

$$\mathcal{L}(\mu^N, z^N; \tilde{p}^N) = \frac{1}{2}\left\|A\mu^N - w\right\|_2^2 + \alpha\left\|z^N\right\|_1 + \left\langle \tilde{p}^N, \mu^N - z^N\right\rangle$$

and by defining $p^N := \frac{\tilde{p}^N}{\lambda}$ we obtain the scaled augmented Lagrangian

$$\mathcal{L}_{\text{sc}}^{\lambda}(\mu^N, z^N; p^N) = \frac{1}{2}\left\|A\mu^N - w\right\|_2^2 + \alpha\left\|z^N\right\|_1 + \frac{\lambda}{2}\left\|\mu^N - z^N + p^N\right\|_2^2 \,.$$

We have the following optimality conditions:

1. primal feasibility condition

$$0 = \partial_{p^N}\mathcal{L} = \mu^N - z^N \,,$$

2. dual feasibility conditions

$$0 = \partial_{\mu^N} \mathcal{L} = A^T(A\mu^N - w) + \lambda p^N \ , \tag{8.7}$$

$$0 \in \partial_{z^N} \mathcal{L} = \alpha \partial \left\| z^N \right\|_1 - \lambda p^N \ . \tag{8.8}$$

By definition μ_{k+1}^N minimizes $\mathcal{L}_{\text{sc}}^{\lambda}$, i.e.

$$0 \in \partial_{\mu^N} \mathcal{L}_{\text{sc}}^{\lambda} = A^T(A\mu_{k+1}^N - w) + \lambda(\mu_{k+1}^N - z_k^N + p_k^N) \ ,$$

which is equivalent to

$$\lambda(z_k^N - z_{k+1}^N) \in A^T(A\mu_{k+1}^N - w) + \lambda p_{k+1}^N \ .$$

Since this fulfills the first dual feasibility condition (8.7), we can consider the dual residual

$$s_{k+1}^N = \lambda(z_k^N - z_{k+1}^N) \ .$$

An analogous derivation for z_{k+1}^N yields that p_{k+1}^N always satisfies the second dual feasibility condition (8.8). Furthermore, we define the primal residual as

$$r_{k+1}^N := z_{k+1}^N - \mu_{k+1}^N \ .$$

By using a bound for the objective suboptimality of the current point κ^*, we obtain

$$\frac{1}{2} \left\| A\mu^N - w \right\|_2^2 + \alpha \left\| z^N \right\|_1 - \kappa^* \leq \left\| p_k^N \right\|_2 \left\| r_k^N \right\|_2 + \nu \left\| s_k^N \right\|_2$$

by estimating $\left\| \mu_k^N - \hat{\mu} \right\|_2 \leq \nu$, where $\hat{\mu}$ is the unknown exact solution. Thus the primal and dual residuals must be small, i.e.

$$\left\| r_k^N \right\|_2 \leq \varepsilon^{\text{pri}} \qquad \text{and} \qquad \left\| s_k^N \right\|_2 \leq \varepsilon^{\text{dual}} \ .$$

Here ε^{pri} and $\varepsilon^{\text{dual}}$ can be chosen via an absolute and relative criterion:

$$\varepsilon^{\text{pri}} = \sqrt{N}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\left\| \mu_k^N \right\|_2, \left\| z_k^N \right\|_2, 0\} \ ,$$

$$\varepsilon^{\text{dual}} = \sqrt{N}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \left\| \lambda p_k^N \right\|_2 \ .$$

Moreover, we use the following adaptive parameter choice:

$$\lambda^{k+1} = \begin{cases} \tau^{\text{incr}}\lambda^k, & \text{if } \left\|r_k^N\right\|_2 > \eta \left\|s_k^N\right\|_2, \\ \frac{\lambda^k}{\tau^{\text{decr}}}, & \text{if } \left\|s_k^N\right\|_2 > \eta \left\|r_k^N\right\|_2, \\ \lambda^k, & \text{else,} \end{cases} \qquad p_{k+1}^N = \begin{cases} \frac{p_k^N}{\tau^{\text{incr}}}, & \text{if } \left\|r_k^N\right\|_2 > \eta \left\|s_k^N\right\|_2, \\ \tau^{\text{decr}} p_k^N, & \text{if } \left\|s_k^N\right\|_2 > \eta \left\|r_k^N\right\|_2, \\ p_k^N, & \text{else.} \end{cases}$$

Finally the algorithm reads as follows:

Algorithm 4 ℓ^1-regularized problem via ADMM

Parameters: $\alpha > 0,\ \lambda > 0,\ A \in \mathbb{R}^{L\times N},\ w \in \mathbb{R}^L,\ \eta > 1,\ \tau^{\text{incr}} > 1,\ \tau^{\text{decr}} > 1,$
$\varepsilon^{\text{rel}} = 10^{-3}$ or 10^{-4}, $\varepsilon^{\text{abs}} > 0$
Initialization: $\mu^N, z^N, r^N, s^N, p^N = 0,\ \varepsilon^{\text{pri}} = \sqrt{N}\,\varepsilon^{\text{abs}}, \varepsilon^{\text{dual}} = \sqrt{N}\,\varepsilon^{\text{abs}}$
while $\left\|r^N\right\|_2 > \varepsilon^{\text{pri}}$ **and** $\left\|s^N\right\|_2 > \varepsilon^{\text{dual}}$ **do**

  $z_{\text{old}}^N = z^N;$
  ▷ Solve Subproblems
  $\mu^N = \left(A^TA + \lambda I\right)^{-1}\left(A^Tw + \lambda\left(z^N - p^N\right)\right);$
  $z^N = \operatorname{sign}\left(\mu^N + p^N\right)\max\left(\left|\mu^N + p^N\right| - \frac{\alpha}{\lambda}, 0\right);$

  $r^N = z^N - \mu^N;$ ▷ Update Primal Residual
  $s^N = \lambda\left(z_{\text{old}}^N - z^N\right);$ ▷ Update Dual Residual

  $p^N = p^N - \left(z^N - \mu^N\right);$ ▷ Lagrange Updates

  ▷ Varying Penalty/Lagrange Parameters
  if $\left\|r^N\right\|_2 > \eta\left\|s^N\right\|_2$ **then**
    $\lambda = \lambda\tau^{\text{incr}};$
    $p^N = \frac{p^N}{\tau^{\text{incr}}};$
  else if $\left\|s^N\right\|_2 > \eta\left\|r^N\right\|_2$ **then**
    $\lambda = \frac{\lambda}{\tau^{\text{decr}}};$
    $p^N = p^N\tau^{\text{decr}};$
  end if
  ▷ Stopping Criteria
  $\varepsilon^{\text{pri}} = \sqrt{N}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\max\left\{\left\|\mu^N\right\|_2, \left\|z^N\right\|_2, 0\right\};$
  $\varepsilon^{\text{dual}} = \sqrt{N}\,\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\left\|\lambda p^N\right\|_2;$

end while
return μ^N ▷ Solution of (8.2)

8.4. Numerical Deconvolution of Three δ-Spikes as the Grid Becomes Finer

In this section we present some results for the solution of the variational problem (8.2) for different discretizations. We consider the deconvolution of three δ-spikes with weights ρ_1, ρ_2 and ρ_3 at the positions ξ_1, ξ_2 and ξ_3. Thus the exact continuous solution reads as follows:

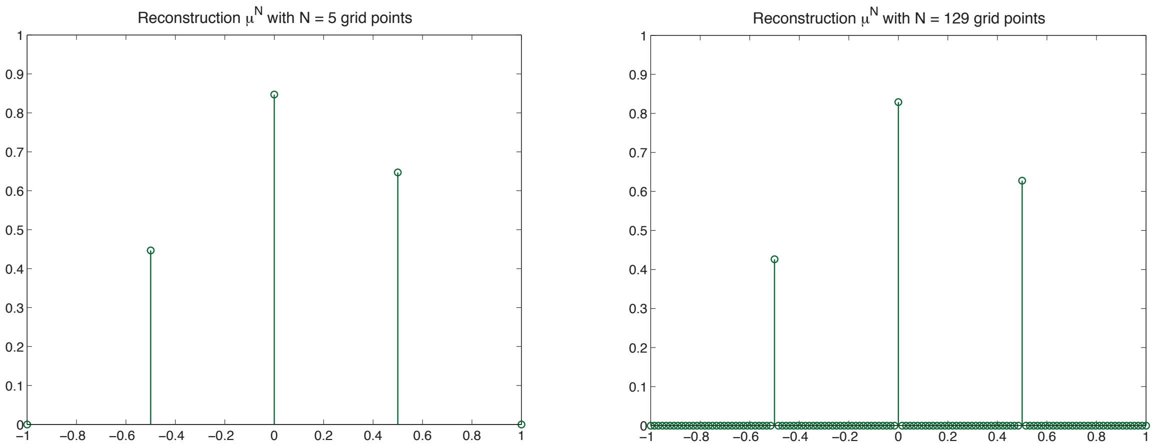
$$\hat{\mu} = \rho_1 \delta_{\xi_1} + \rho_2 \delta_{\xi_2} + \rho_3 \delta_{\xi_3} \ .$$

Moreover, we choose a Gaussian convolution kernel $\widetilde{G}$ with standard deviation $\sigma = 0.05$. The continuous convolved data can be computed analytically as

$$w(y) = \mathcal{K}\hat{\mu}(y) = \rho_1 \widetilde{G}(y - \xi_1) + \rho_2 \widetilde{G}(y - \xi_2) + \rho_3 \widetilde{G}(y - \xi_3)$$

and shall be discretized for different grid sizes.

In accordance with Theorem 8.2, let us first consider discretizations, which include ξ_1, ξ_2 and ξ_3 as supporting points. As an exact solution $\hat{\mu}$ we choose spikes at the positions $\xi_1 = -0.5$, $\xi_2 = 0$ and $\xi_3 = 0.5$ and weights $\rho_1 = 0.5$, $\rho_2 = 0.9$ and $\rho_3 = 0.7$. In these cases we should be able to recover the exact positions of the spikes.
As an example, we choose discretizations with 5 and 129 grid points. In both cases every point ξ_1, ξ_2 and ξ_3 coincides with a grid point. In Figure 8.2 we observe that we are indeed able to exactly recover the positions of the spikes. This meets our expectations. Other discretizations, which include the point ξ_1, ξ_2 and ξ_3, lead to the same results.



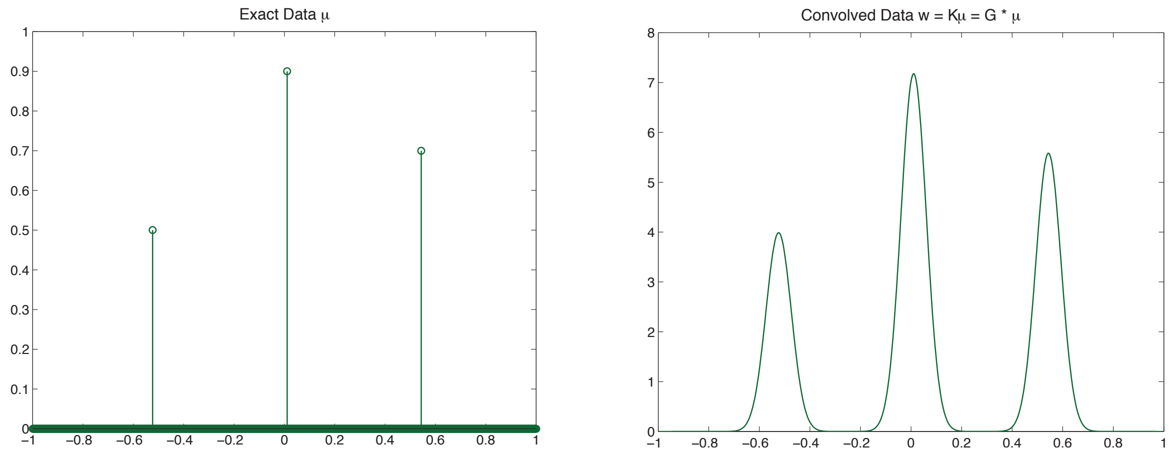
(a) Reconstruction μ^N involving $N = 5$ grid points

(b) Reconstruction μ^N involving $N = 129$ grid points

Figure 8.2.: If the positions of the spikes are part of the grid, we are able to exactly reconstruct the support

Let us now consider the case of Theorem 8.3. As an example, we consider an exact solution $\hat{\mu}$, which again shall consist of three delta spikes. However, this time the spikes shall be located in between the grid points. In order to guarantee that the exact positions are indeed always located between the grid points, we choose irrational locations, i.e. $\xi_1 = -\frac{\pi}{6}$, $\xi_2 = \frac{\pi}{300}$ and $\xi_3 = \frac{e}{5}$ with the same weights ρ_1, ρ_2 and ρ_3 as before.

Subfigure 8.3(a) illustrates the exact solution $\hat{\mu}$. Subfigure 8.3(b) shows the exact convolved data $w = \mathcal{K}\hat{\mu}$ plotted by using a very fine discretization.



(a) Three spikes at the positions $\xi_1 = -\frac{\pi}{6}$, $\xi_2 = \frac{\pi}{300}$ and $\xi_3 = \frac{e}{5}$. Note that this plot is for demonstration only. Clearly, the heights of the spikes are actually $+\infty$.

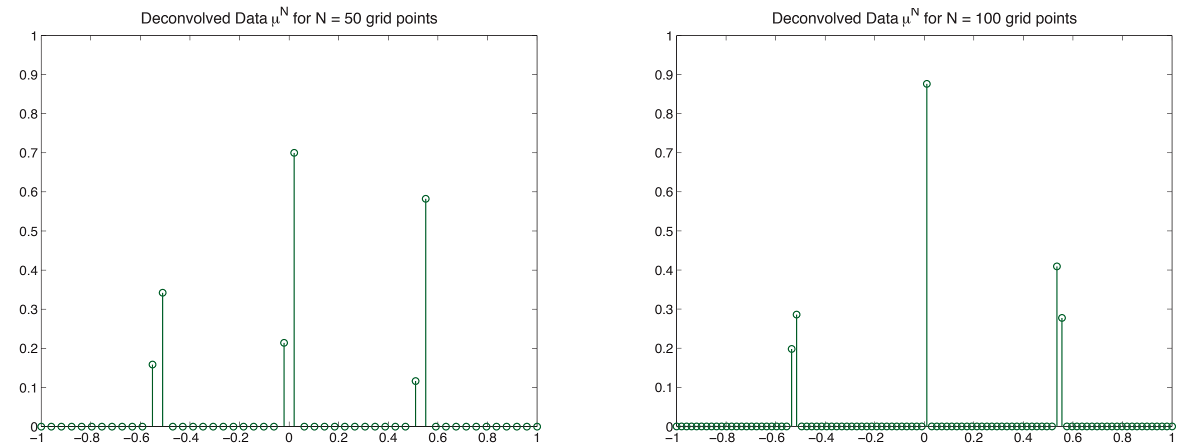
(b) Continuous convolution $w = \mathcal{K}\hat{\mu}$ of the exact data $\hat{\mu}$ with kernel $\widetilde{G}$

Figure 8.3.: Three spikes at irrational positions and their continuous convolution

As an example for the reconstructions via Algorithm 4, Figure 8.4 visualizes the reconstruction for 50 and for 100 grid points with regularization parameter $\alpha = 0.1$. We observe that in most of the cases we reconstruct two peaks in a small region around the location of the exact spike. In some cases on the other hand, we obtain only one peak located very closely to the position of the exact spike.

In order to verify our prior analysis, let us now consider the results as the step size of the grid becomes smaller. Therefore, we let Algorithm 4 run for different discretizations, where we start from $N = 10$ and increment N by one until we reach $N = 100$.

In Figure 8.5 we can see the results for these different discretizations. Subfigure 8.5(a) indicates that the reconstructed peaks draw nearer to the positions of the exact spikes. Furthermore, Subfigure 8.5(b) illustrates the number of reconstructed peaks. We especially see that for discretizations with at least 41 nodes, the number of reconstructed peaks does not exceed 6, which is twice the number of exact spikes. This yields the supposition that, in case $\xi \in (x_k + \frac{\alpha h}{2G(0)}, x_k + \frac{h}{2})$ holds in the formulation of Theorem

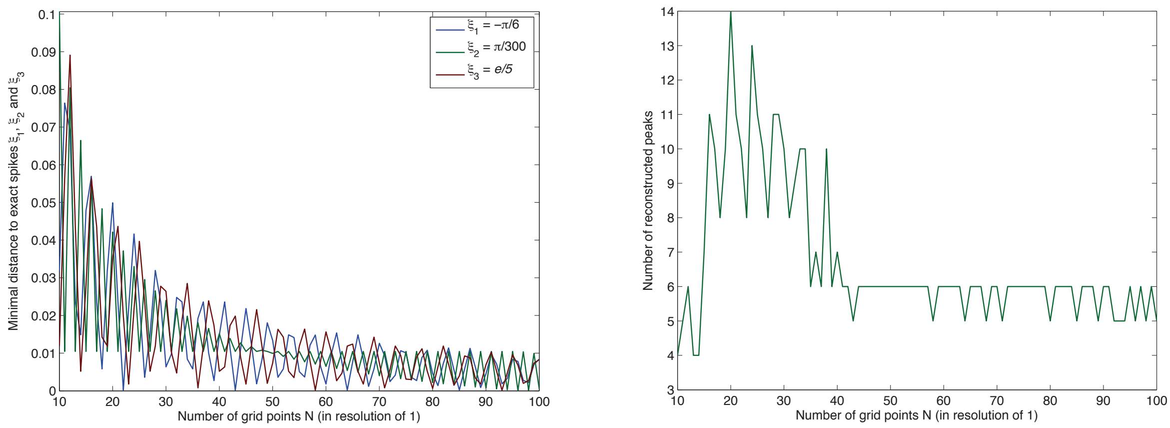


(a) Deconvolved data μ^N for $N = 50$ and $\alpha = 0.1$

(b) Deconvolved data μ^N for $N = 100$ and $\alpha = 0.1$

Figure 8.4.: Deconvolution of three spikes with locations in between the grid points

8.3, the solution of (8.2) consists of two peaks, which are located at the two grid points closest to ξ, i.e. x_k and x_{k+1}. These results are consistent with the observations made by Duval and Peyré.



(a) Minimal distances of the reconstructed spikes to the exact spikes

(b) Number of reconstructed peaks

Figure 8.5.: Results for different discretizations; $\alpha = 0.1$

9 Conclusions and Outlook

Throughout this thesis we have not only engaged in the analysis and implementation of a novel, local sparsity promoting regularization type, namely $\ell^{1,\infty}$-regularization, but also in the analysis of the asymptotics of certain sparsity priors. In order to conclude, we shall divide this chapter into two sections accordingly.

9.1. Local Sparsity

For the solution of inverse problems, where the unknown is considered to be a matrix, mixed $\ell^{p,q}$-norms can be used as regularization functionals in order to promote certain structures in the reconstructed matrix. Joint sparsity via $\ell^{p,1}$-regularization for instance has already been examined in the literature, c.f. Teschke and Ramlau (2007) and Fornasier and Rauhut (2008).
Motivated by dynamic positron emission tomography for myocardial perfusion, we proposed a novel variational model for a dictionary based matrix completion problem incorporating local sparsity via $\ell^{1,\infty}$-regularization as an alternative to the more commonly considered joint sparsity model. We not only analyzed the existence and potential uniqueness of a solution, but also investigated the subdifferential of the $\ell^{1,\infty}$-functional and a source condition. One of the main results of this thesis consists of the deduction of an equivalent formulation, which not only simplifies the analysis of the problem, but also facilitates its numerical implementation. Moreover, we discussed exact recovery for locally 1-sparse solutions by analyzing the noise-free case, in which we considered the minimization of the nonnegative $\ell^{1,\infty}$-functional with an equality constraint in the data fidelity term. As a result of this analysis, we discovered that the dictionary matrix has to be normalized in a certain way in order to exactly reconstruct locally 1-sparse data under simplified conditions.

In our prior research (cf. Heins 2011), the numerical implementation of the $\ell^{1,\infty}$-regularized variational problem relied on the usage of the Kronecker product. Due to this, however, the matrix of a linear system, which had to been solved, became extremely large and we could not reconstruct reasonably large data. Thus this approach did not yield satisfactory results. In this work, a novel implementation of the problem was developed that relies on a double splitting via the alternating direction method of multipliers (ADMM). The algorithm yields superior results, in particular an almost exact recovery of the true support of the solution. Nevertheless, one drawback of the reformulation of the problem we introduced is that the results are not very close to the true solution. However, having a good estimate of the support of the solution allows us to refine our first result by solving the inverse problem restricted to the previously recovered support with no further regularization. This second result shows promising features, even in the presence of Gaussian noise.

However, for some coefficients at the boundary of the exact nonzero region the algorithm still picked the wrong basis vector. In order to overcome this problem and to further improve the results, we added a total variation term to the variational regularization scheme. This could either be done by incorporating total variation on the images in every time step or on the coefficient matrices for every basis vector. We proposed two supplemental algorithms, which compute a solution to exactly those problems. Once again we used ADMM, however, this time we had to split thrice due to the additional total variation minimization. Since we now have two regularizations, the choice of the right combination of regularization parameters becomes challenging.

We first tested the algorithm, which includes total variation on the image in every time step. In the case that total variation regularization had a strong influence during the reconstruction, the results were unsatisfactory. On the other hand, when we chose the TV-regularization parameter low, total variation had no effect during the recovery process and the algorithm computed the same results as without using total variation. Unfortunately, choosing the TV-regularization parameter somewhere in between neither led to better results than without using total variation. The reason for this unsatisfactory results might be the fact that the conjugate gradient method, which we used to solve a linear system during ADMM, did not always converge. However, we did not further engage in the improvement of this approach, since the original method already yielded satisfactory results and by including total variation on the coefficient matrices instead, the reconstructions improved and even better results could be recovered.

Afterwards, we examined the algorithm, which includes total variation on the coefficient matrices for every basis vector. Once again choosing a good combination of regularization parameters was challenging, however, the results were promising. We found combinations of regularization parameters, which led to a smaller percentage of wrongly reconstructed coefficients than we could obtain without incorporating total variation regularization. The coefficients at the boundary of the nonzero region were now reconstructed correctly and for a few parameter combinations we were even able to reconstruct the support exactly. Even in the presence of Gaussian noise, the algorithm performed outstandingly. Indeed, some of the results, which were computed on noisy data, yielded even better results than our first algorithm was able to compute on noiseless data. However, the challenge of finding the right parameter combination remains and at the moment this can only be done by visual inspection. Therefore, it would be advantageous to develop a rule for the parameter choice.

In summary, the results obtained by our first approach without using total variation were very satisfactory and could be even improved by incorporating an additional total variation regularization on the coefficient matrices. For some parameter combinations, we even recovered the support exactly. Moreover, by including an additional total variation regularization, the algorithm became more robust to noise. All in all, these results motivate to investigate the interplay between the parameters to develop parameter choice rules, which eventually turn the approach including total variation into an effective reconstruction scheme for practical applications.

9.2. Sparsity Asymptotics

In many applications of inverse problems, sparsity-enforcing regularizations have become indispensable tools. Unfortunately, most of these methods crucially rely on the way they are discretized and are thus rather difficult to analyze in a more general way. Moreover, an infinite rather than a finite-dimensional modeling would be advantageous in many applications. For these reasons, we analyzed the Γ-limits of such discrete approaches in order to develop a suitable asymptotic theory.

By utilizing Γ-convergence, we analyzed the asymptotic behaviour of the ℓ^p-norm for differently scaled weights. Afterwards, we deduced a Γ-limit for the non-weighted ℓ^0-"norm", which states that in case the solution consists of a sum of δ-spikes, the ℓ^0-"norm" Γ-converges to the exact number of these spikes. This result stimulated the numerical experiments presented in the subsequent chapter. Moreover, we examined the asymptotics of mixed norms. The results were consistent with the expectations we had based on our previous analysis. For the general case and a reasonable scaling, we were able to deduce, that the continuous $L^{p,q}$-norm is indeed the Γ-limit of the $\ell^{p,q}$-norm. Besides this, we investigated the local sparsity promoting $\ell^{1,\infty}$-norm, which built the bridge to the first part of this thesis.

As a numerical scenario to validate our asymptotic results, we examined the deconvolution of a sparse spike pattern. We provided a continuous exact solution, which could be represented as a linear combination of a few δ-spikes. By analytically convolving these spikes with a Gaussian kernel and discretizing it for different grid sizes, we obtained the data, which we used for the reconstruction. In order to observe the asymptotic behaviour of the ℓ^1-norm for such a sparse spike pattern, we computed ℓ^1-regularized deconvolutions as the step size of the grid became smaller. For the solution of this variational problem we used again ADMM. As the grid became finer, we indeed observed, that at most twice as many discrete peaks are reconstructed as there are continuous δ-spikes. In our corresponding theoretical considerations we drew the conclusion that the support of one δ-spike can be reconstructed exactly if the spike is at the same position as a grid point. More interestingly, we furthermore discovered, that one discrete delta peak can only be the solution for this problem if the exact δ-spike is close enough to a supporting point. Our numerical results, however, give reason to the assumption that as the grid becomes finer no more than two peaks for every exact δ-spike will be reconstructed and the reconstructed peaks are located at the two grid points, which are closest to the exact δ-spike. The theoretical verification of this hypothesis is still open and subject to future work.

A

Solving the Positive $\ell^{1,\infty} - \ell^{1,1}$-Projection-Problem

Let us now solve the following problem:

$$\min_{D \in G} \frac{\lambda}{2} \|D - U + P\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} d_{ij} \quad \text{s.t.} \quad \sum_{j=1}^{N} d_{ij} \leq \tilde{v} \; . \tag{A.1}$$

In order to do so, we reformulate the first part of the problem, i.e.

$$\begin{aligned} &\frac{\lambda}{2} \|D - U + P\|_F^2 + \beta \sum_{i=1}^{M} \sum_{j=1}^{N} d_{ij} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\frac{\lambda}{2} (d_{ij} - u_{ij} + p_{ij})^2 + \beta d_{ij} \right) \\ &= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{\lambda}{2} \left(d_{ij}^2 - 2 d_{ij} (u_{ij} + p_{ij}) + (u_{ij} + p_{ij})^2 + \frac{2\beta}{\lambda} d_{ij} \right) \\ &= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{\lambda}{2} \left(d_{ij}^2 - 2 d_{ij} \left(u_{ij} + p_{ij} - \frac{\beta}{\lambda} \right) + (u_{ij} + p_{ij})^2 \right) \\ &= \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{\lambda}{2} \left(d_{ij} - \left(u_{ij} + p_{ij} - \frac{\beta}{\lambda} \right) \right)^2 - \frac{\lambda}{2} \left(\left(u_{ij} + p_{ij} - \frac{\beta}{\lambda} \right)^2 + (u_{ij} + p_{ij})^2 \right) \; . \end{aligned}$$

Since the last part of the sum is independent of d_{ij}, we can consider

$$\min_{D \in G} \frac{\lambda}{2} \left\| D - U + P - \frac{\beta}{\lambda} \mathbb{1}_{M \times N} \right\|_F^2 \quad \text{s.t.} \quad \sum_{j=1}^{N} d_{ij} \leq \tilde{v}$$

instead. We can minimize this expression with respect to every row independently, i.e.

$$\min_{D \in \mathbb{R}^{M \times N}} \frac{\lambda}{2} \left\| d_{(i)} - u_{(i)} + p_{(i)} - \frac{\beta}{\lambda} \mathbb{1}_N \right\|_2^2 , \tag{A.2}$$

with the constraints

$$\left(d_{(i)}\right)_j \geq 0 \qquad \forall\, j \in \{1, \dots, N\} , \tag{Constr1}$$

$$\sum_{j=1}^{N} \left(d_{(i)}\right)_j \leq \tilde{v} , \tag{Constr2}$$

where $d_{(i)}$ denotes the ith *transposed* row of D, respectively $u_{(i)}$, $p_{(i)}$.
In order to minimize this problem, we first consider (A.2) under (Constr1) only. In this case the solution is given by

$$\widetilde{d}_{(i)} = \max \left\{ u_{(i)} + p_{(i)} - \frac{\beta}{\lambda} \mathbb{1}_N, 0 \right\} . \tag{A.3}$$

In order to include (Constr2), we have to do a case-by-case-analysis:

<u>Case a:</u>

Let (A.3) satisfy (Constr2). In this case the solution of (A.2) under (Constr1) and (Constr2) is given by

$$d_{(i)} = \widetilde{d}_{(i)} .$$

<u>Case b:</u>

Let (A.3) *not* satisfy (Constr2), i.e. $\sum_{j=1}^{N} \left(\widetilde{d}_{(i)}\right)_j > \tilde{v}$ holds. Then the solution of (A.2) under (Constr1) and (Constr2) has to fulfill

$$\sum_{j=1}^{N} \left(d_{(i)}\right)_j = \tilde{v} . \tag{Constr3}$$

Thus we have to solve (A.2) under (Constr1) and (Constr3). For this purpose we propose the corresponding Lagrange functional as

$$\begin{aligned}\mathcal{L}^\lambda(d_{(i)},\mu_{(i)},\vartheta) = \min_{D\in\mathbb{R}^{M\times N}} \frac{\lambda}{2}\left\|d_{(i)} - u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N\right\|_2^2 \\ + \vartheta\left(\sum_{j=1}^{N}\left(d_{(i)}\right)_j - \tilde{v}\right) - \sum_{j=1}^{N}\left(d_{(i)}\right)_j\left(\mu_{(i)}\right)_j \ .\end{aligned} \tag{A.4}$$

Once we know ϑ, we can compute the optimal $d_{(i)}$ as

$$\begin{aligned}d_{(i)} &= \operatorname{shrink}^+\left(u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N,\ \frac{\vartheta}{\lambda}\mathbb{1}_N\right) \\ &:= \max\left\{u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N - \frac{\vartheta}{\lambda}\mathbb{1}_N,\ 0\right\} \ .\end{aligned} \tag{A.5}$$

We can see this by computing the optimality condition of (A.4)

$$\begin{aligned}0 &= \partial_{d_{(i)}}\mathcal{L}^\lambda(d_{(i)},\mu_{(i)},\vartheta) \\ &= \lambda\left(d_{(i)} - u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N\right) + \vartheta\mathbb{1}_N - \mu_{(i)} \ ,\end{aligned} \tag{A.6}$$

with the complementary conditions

$$\left(\mu_{(i)}\right)_j \geq 0 \qquad \text{and} \qquad \left(\mu_{(i)}\right)_j\left(d_{(i)}\right)_j = 0 \quad \forall\, j \in \{1,\dots,N\}.$$

If $\left(d_{(i)}\right)_j \neq 0$ holds, then we have $\left(\mu_{(i)}\right)_j = 0$ and thus we obtain from (A.6) that

$$\begin{aligned}0 &= \lambda\left(d_{(i)} - u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N\right) + \vartheta\mathbb{1}_N \\ \Leftrightarrow\ d_{(i)} &= u_{(i)} - p_{(i)} + \frac{\beta-\vartheta}{\lambda}\mathbb{1}_N\end{aligned}$$

holds. On the other hand if we have $\left(d_{(i)}\right)_j = 0$, then $\left(\mu_{(i)}\right)_j \geq 0$ has to hold. Hence we obtain

$$\begin{aligned}\left(\mu_{(i)}\right)_j &= \lambda\left(p_{(i)} - u_{(i)}\right)_j - \beta + \vartheta \geq 0 \\ \Leftrightarrow \qquad & \left(u_{(i)} - p_{(i)}\right)_j + \frac{\beta-\vartheta}{\lambda} \leq 0 \ .\end{aligned}$$

by using (A.6). The Lagrange parameter ϑ should be chosen such that (Constr3) holds. Therefore, we investigate

$$\sum_{j\in I}\left(\left(u_{(i)} - p_{(i)}\right)_j + \frac{\beta-\vartheta}{\lambda}\right) = v \ ,$$

where the set I contains all indices, for which

$$\left(u_{(i)} - p_{(i)}\right)_j + \frac{\beta - \vartheta}{\lambda} \geq 0 \tag{A.7}$$

holds. This is reasonable, since for all other indices $j \notin I$ the term $\left(u_{(i)} - p_{(i)}\right)_j + \frac{\beta-\vartheta}{\lambda}$ is projected to zero, which is true due to (A.5).
Hence we obtain

$$\vartheta = \frac{\lambda}{|I|}\left(\sum_{j\in I}\left(u_{(i)} - p_{(i)}\right)_j + \frac{\beta}{\lambda} - \tilde{v}\right) .$$

Now we have to compute I. Since we are able to sort the vectors according to value, it is sufficient to find $|I|$. Then we obtain

$$\vartheta = \frac{\lambda}{|I|}\left(\sum_{r=1}^{|I|}\left(\widehat{u_{(i)} - p_{(i)}}\right)_r + \frac{\beta}{\lambda} - \tilde{v}\right) ,$$

where $\widehat{\cdot}$ denotes the respective vector sorted according to value.
In order to obtain $|I|$, we use the following result:

Theorem A.1 (Duchi et al. (2008, p. 3)).
Let $\widehat{u_{(i)} - p_{(i)}}$ denote the vector obtained by sorting $u_{(i)} - p_{(i)}$ in a descending order. Then the number of indices, for which (A.7) holds, is

$$|I| = \max\left\{ j \in \{1,\dots,N\} \;\middle|\; \lambda\left(\widehat{u_{(i)} - p_{(i)}}\right)_j + \beta - \frac{\lambda}{j}\left(\sum_{r=1}^{j}\left(\widehat{u_{(i)} - p_{(i)}}\right)_r + \frac{\beta}{\lambda} - \tilde{v}\right) > 0 \right\} .$$

Now we are able to propose the algorithm for the solution of (A.1).

Algorithm 5 Positive $\ell^{1,\infty}$-$\ell^{1,1}$-projection

Parameters: $U \in \mathbb{R}^{M \times N},\ P \in \mathbb{R}^{M \times N}, \tilde{v} > 0,\ \beta > 0,\ \lambda > 0,\ M, N \in \mathbb{N}$
Initialization: $D = 0,\ |I| = 0,\ \vartheta = 0$
for all $i \in \{1, ..., M\}$ **do**

$\widetilde{d}_{(i)} = \max\left\{u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N, 0\right\};$

if $\sum_{j=1}^{N} \widetilde{d}_{ij} \leq \tilde{v}$ **then** ▷ Solve with (Constr1) and (Constr2)

$d_{(i)} = \widetilde{d_{(i)}};$

else ▷ Solve with (Constr1) and (Constr3)

$$|I| = \max\left\{ j \in \{1, \dots, N\} \,\middle|\, \lambda\left(\widehat{u_{(i)} - p_{(i)}}\right)_j + \beta - \frac{\lambda}{j}\left(\sum_{r=1}^{j}\left(\widehat{u_{(i)} - p_{(i)}}\right)_r + \frac{\beta}{\lambda} - \tilde{v}\right) > 0 \right\};$$

$$\vartheta = \frac{\lambda}{|I|}\left(\sum_{r=1}^{|I|}\left(\widehat{u_{(i)} - p_{(i)}}\right)_r + \frac{\beta}{\lambda} - \tilde{v}\right);$$

$d_{(i)} = \operatorname{shrink}^{+}\left(u_{(i)} + p_{(i)} - \frac{\beta}{\lambda}\mathbb{1}_N,\ \frac{\vartheta}{\lambda}\mathbb{1}_N\right);$

end if

end for
return D ▷ Solution of (A.1)

B

Inequality for Stopping Criteria

In order to proof the inequality

$$\begin{aligned} &\frac{1}{2}\left\|AZ - W\right\|_F^2 + \beta \sum_{i=1}^{M}\sum_{j=1}^{N} d_{ij} + J(D) - \kappa^* \\ &\leq \left\langle P^k, R_1^k \right\rangle + \left\langle Q^k, R_2^k \right\rangle + \left\langle U^k - U^*, S^k \right\rangle \ , \end{aligned} \tag{5.11}$$

which is needed in Subsection 5.1.2, we consider the unscaled augmented Lagrangian (5.2).

By definition U^{k+1} minimizes

$$\mathcal{L}_{un}^{\lambda,\mu}\left(U, D^k, Z^k; \widetilde{P}^k, \widetilde{Q}^k\right) \ ,$$

D^{k+1} minimizes

$$\mathcal{L}_{un}^{\lambda,\mu}\left(U^{k+1}, D, Z^k; \widetilde{P}^k, \widetilde{Q}^k\right)$$

and Z^{k+1} minimizes

$$\mathcal{L}_{un}^{\lambda,\mu}\left(U^{k+1}, D^{k+1}, Z; \widetilde{P}^k, \widetilde{Q}^k\right) \ .$$

We now have to examine the optimality conditions.

<u>OPT1</u>:

Starting with

$$\begin{aligned} 0 &\in \partial_U \mathcal{L}_{un}^{\lambda,\mu}\left(U^{k+1}, D^k, Z^k; \widetilde{P}^k, \widetilde{Q}^k\right) \\ &= \widetilde{P}^k + \lambda(U^{k+1} - D^k) + \widetilde{Q}^k B + \mu(U^{k+1}B^T - Z^k)B \end{aligned}$$

we insert the Lagrange updates

$$\widetilde{P}^k = \widetilde{P}^{k+1} + \lambda(D^{k+1} - U^{k+1}), \qquad \widetilde{Q}^k = \widetilde{Q}^{k+1} + \mu(Z^{k+1} - U^{k+1}B^T) \tag{B.1}$$

and obtain

$$0 \in \widetilde{P}^{k+1} + \widetilde{Q}^{k+1}B + \lambda(D^{k+1} - D^k) + \mu(Z^{k+1} - Z^k)B \; .$$

Thus we see that U^{k+1} minimizes

$$\left\langle \widetilde{P}^{k+1} + \widetilde{Q}^{k+1}B, U \right\rangle + \lambda \left\langle D^{k+1} - D^k, U \right\rangle + \mu \left\langle Z^{k+1} - Z^k, UB^T \right\rangle \; .$$

<u>OPT2</u>:

Now we have

$$\begin{aligned} 0 &\in \partial_D \mathcal{L}_{un}^{\lambda,\mu}\left(U^{k+1}, D^{k+1}, Z^k; \widetilde{P}^k, \widetilde{Q}^k\right) \\ &= \beta \mathbb{1}_{M\times N} - \widetilde{P}^k - \lambda(U^{k+1} - D^{k+1}) + \partial J(D^{k+1}) \; , \end{aligned}$$

with J as defined in (5.1). Inserting $\widetilde{P}^k$ from (B.1) yields

$$0 \in \beta \mathbb{1}_{M\times N} - \widetilde{P}^{k+1} + \partial J(D^{k+1}) \; .$$

Hence we see that D^{k+1} minimizes

$$\beta \sum_{i=1}^{M} \sum_{j=1}^{N} d_{ij} - \left\langle \widetilde{P}^{k+1}, D \right\rangle \quad \text{s.t.} \quad \sum_{j=1}^{N} d_{ij} \leq \tilde{v}, \; d_{ij} \geq 0 \; .$$

<u>OPT3</u>:

In this case the optimality condition reads as follows:

$$\begin{aligned} 0 &\in \partial_Z \mathcal{L}_{un}^{\lambda,\mu}\left(U^{k+1}, D^{k+1}, Z^{k+1}; \widetilde{P}^k, \widetilde{Q}^k\right) \\ &= A^T(AZ^{k+1} - W) - \widetilde{Q}^k - \mu(U^{k+1}B^T - Z^{k+1}) \; . \end{aligned}$$

Inserting $\widetilde{Q}^k$ from (B.1) yields

$$0 \in A^T(AZ^{k+1} - W) - \widetilde{Q}^{k+1} \; .$$

Therefore, Z^{k+1} minimizes

$$\frac{1}{2}\|AZ-W\|_F^2-\left\langle \widetilde{Q}^{k+1},Z\right\rangle \ .$$

All in all, we obtain that the following inequalities have to hold:

$$\begin{aligned}&\left\langle \widetilde{P}^{k+1}+\widetilde{Q}^{k+1}B,U^{k+1}\right\rangle+\lambda\left\langle D^{k+1}-D^k,U^{k+1}\right\rangle+\mu\left\langle Z^{k+1}-Z^k,U^{k+1}B^T\right\rangle\\ \leq\ &\left\langle \widetilde{P}^{k+1}+\widetilde{Q}^{k+1}B,U^{*}\right\rangle+\lambda\left\langle D^{k+1}-D^k,U^{*}\right\rangle+\mu\left\langle Z^{k+1}-Z^k,U^{*}B^T\right\rangle\ ,\end{aligned}\tag{B.2}$$

$$\begin{aligned}&\beta\sum_{i=1}^{M}\sum_{j=1}^{N}d_{ij}^{k+1}-\left\langle \widetilde{P}^{k+1},D^{k+1}\right\rangle+J(D^{k+1})\\ \leq\ &\beta\sum_{i=1}^{M}\sum_{j=1}^{N}d_{ij}^{*}-\left\langle \widetilde{P}^{k+1},D^{*}\right\rangle+J(D^{*})\end{aligned}\tag{B.3}$$

and

$$\frac{1}{2}\left\|AZ^{k+1}-W\right\|_F^2-\left\langle \widetilde{Q}^{k+1},Z^{k+1}\right\rangle\ \leq\ \frac{1}{2}\|AZ^{*}-W\|_F^2-\left\langle \widetilde{Q}^{k+1},Z^{*}\right\rangle\ .\tag{B.4}$$

Adding equations (B.2), (B.3) and (B.4) together yields

$$\begin{aligned}&\frac{1}{2}\left\|AZ^{k+1}-W\right\|_F^2+\beta\sum_{i=1}^{M}\sum_{j=1}^{N}d_{ij}^{k+1}+J(D^{k+1})\\ &-\frac{1}{2}\|AZ^{*}-W\|_F^2-\beta\sum_{i=1}^{M}\sum_{j=1}^{N}d_{ij}^{*}-J(D^{*})\\ \leq\ &\left\langle \widetilde{P}^{k+1},D^{k+1}-U^{k+1}\right\rangle+\left\langle \widetilde{Q}^{k+1},Z^{k+1}-U^{k+1}B^T\right\rangle+\lambda\left\langle D^{k+1}-D^k,U^{*}-U^{k+1}\right\rangle\\ &+\mu\left\langle Z^{k+1}-Z^k,(U^{*}-U^{k+1})B^T\right\rangle+\left\langle \widetilde{P}^{k+1},U^{*}-D^{*}\right\rangle+\left\langle \widetilde{Q}^{k+1},U^{*}B^T-Z^{*}\right\rangle\ .\end{aligned}$$

By using the definitions of $R_{1,2}^{k+1}$ and S^{k+1} (see for instance (5.9),(5.10) and (5.8)) and the fact that we have $U^{*}=D^{*}$ and $U^{*}B^T=Z^{*}$, we finally obtain

$$\begin{aligned}&\frac{1}{2}\left\|AZ^{k+1}-W\right\|_F^2+\beta\sum_{i=1}^{M}\sum_{j=1}^{N}d_{ij}^{k+1}+J(D^{k+1})-\kappa^{*}\\ \leq\ &\left\langle P,R_1^k\right\rangle+\left\langle Q,R_2^k\right\rangle+\left\langle U^k-U^{*},S^k\right\rangle\ .\end{aligned}\tag{5.11}$$

C

Computation Of The Kinetic Modeling Matrix

In order to compute a solution to the inverse problem (6.6) in Chapter 6, we need the matrix B, which contains the kinetic modeling basis vectors as proposed in Section 6.1. For this reason we want to outline the computation of the kinetic modeling matrix B, which we used in the numerical experiments.

Due to the linearization in (6.3), we want to compute

$$b_j(t) = \int_0^t C_A(\tau) e^{-\tilde{b}_j(t-\tau)} \, \mathrm{d}\tau$$

for every $j \in \{1, ..., N\}$. For this purpose we consider the corresponding linear ordinary differential equation

$$b_j'(t) = C_A(t) - \tilde{b}_j b_j(t) \; .$$

In order to solve this differential equation via *Euler method* (cf. Hairer et al. 1993, p. 204), the coefficients $\tilde{b}_j$, which describe the blood flow, are necessary. However, as we have mentioned in Section 6.1, these are indeed known. We have seen that they equal the perfusion F divided by the ratio $\lambda = \frac{C_T(t)}{C_V(t)}$ between the tracer concentration in tissue C_T and the venous tracer concentration C_V, i.e. $\frac{F(x)}{\lambda}$. Note that due to the fact that we have set radioactive water ($H_2^{15}O$) as the default tracer, we can use its property as being highly diffusible and thus the concentrations C_T and C_V reach an equilibrium very quickly. For this reason we are able to assume the ratio $\lambda = \frac{C_T}{C_V}$ to be constant.

Furthermore we require a vector containing typical measure times for dynamic PET and the input curve $C_A(t)$, which consists of the tracer concentration in blood dependent on time. Now in order to obtain the basis vectors b_j, Euler method is applied to every known $\tilde{b}_j$. Due to the results in Section 4.2, we afterwards normalize the basis vectors. Thus, we obtain the following algorithm:

Algorithm 6 Euler method for kinetic modeling

Parameters: $\tilde{b} \in \mathbb{R}^N$, $time \in \mathbb{R}^T$, $C_A \in \mathbb{R}^T$, $h \in \mathbb{R}$
Initialization: $b_{t,j} = 0 \; \forall \, t, j$

Interpolate C_A with step size h ▷ initially C_A depends on *time*

for $j = \{1, ..., N\}$ **do**
  for $t = \left\{1, ..., \widetilde{T}\right\}$ **do**
    $b_{t+1,j} = b_{t,j} + h\left(C_{At} - \tilde{b}_j b_{t,j}\right)$;
  end for
end for

Interpolate B to make it dependent on *time* again

for $j = \{1, ..., N\}$ **do** ▷ normalize B
  $b_{:,j} = \frac{b_{:,j}}{\|b_{:,j}\|_2}$;
end for

return B

Note that in order to obtain better accuracy, we had to interpolate the input curve.

D

Exact ℓ^1-Reconstruction of 1-Sparse Signals in 1D

In order to gain some understanding into suitable and necessary scaling conditions for the recovery of locally 1-sparse solutions, we first consider the simplest case, namely $M = 1$, when the problem reduces to standard ℓ^1-minimization.

Theorem D.1 (Exact Reconstruction of a 1-Sparse Signal in 1D)**.**
Let the vector $w := e_j^T B^T = b_j^T$ be the jth basis vector and $c = 1 - (\alpha + \beta)$ for $(\alpha + \beta) \in (0, 1)$.

If $\hat{u} = ce_j^T$ is the solution of (4.2), then the matrix B has to meet the scaling condition

$$\|b_n\|_{\ell^2} = 1 \qquad \text{and} \qquad |\langle b_n, b_m \rangle| \leq 1 \quad \text{for} \quad n \neq m \, . \tag{4.21}$$

Proof.
We first calculate the optimality condition of (4.2) as

$$0 = \left(\left(uB^T - w \right) B \right)_n + (\alpha + \beta)\, p_n \qquad \text{with} \qquad p_n \in \partial |u_n| \, .$$

Then it follows that

$$p_n = \frac{1}{\alpha + \beta} \left(\left(w - uB^T \right) B \right)_n$$

holds. Subsequently we insert $\hat{u} = ce_j^T$ and $w = e_j^T B^T$ to obtain

$$\begin{aligned} p_n &= \frac{1}{\alpha+\beta} \left(\left(e_j^T B^T - c e_j^T B^T \right) B \right)_n \\ &= \frac{1-c}{\alpha+\beta} \left(e_j^T B^T B \right)_n \\ &= \frac{1-c}{\alpha+\beta} \left(b_j^T B \right)_n \\ &= \frac{1-c}{\alpha+\beta} \sum_{t=1}^{T} b_{tj} b_{tn} \\ &= \frac{1-c}{\alpha+\beta} \langle b_j, b_n \rangle \ . \end{aligned}$$

Since $p_n \in \partial |\hat{u}_n|$ has to be satisfied, we need to ensure that

$$p_j = 1 \quad \text{and} \quad p_i \in [-1, 1] \quad \text{for} \quad i \neq j$$

holds, which is true under the assumptions mentioned above. □

For this reason we know that we have to normalize our basis vectors with respect to the ℓ^2-norm in order to reconstruct at least a δ-peak exactly in one dimension. Note that in the one-dimensional case the $\ell^{1,\infty}$-regularization and the $\ell^{1,1}$-regularization reduce to a single ℓ^1-regularization with regularization parameter $\alpha + \beta$.

Bibliography

AMBROSIO, F.L., FUSCO, N. AND PALLARA, D., *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, Clarendon Press (2000). 43, 46

AUBERT, G. AND KORNPROBST, P., *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer (2002). 33

BAILEY, D.L., TOWNSEND, D.W., VALK, P.E. AND MAISEY, M.N. (editors) *Positron Emission Tomography: Basic Sciences*, Springer, London (2005). 128

BARON, D., WAKIN, M.B., DUARTE, M.F., SARVOTHAM, S. AND BARANIUK, R.G., *Distributed compressed sensing*, Technical report, Department of Electrical and Computer Engineering Rice University Houston, TX 77005, USA (2005). 69

BENEDEK, A. AND PANZONE, R., *The spaces lp, with mixed norm*, Duke Mathematical Journal (1961), 28(3): pp. 301–324. 65

BENNING, M., *Singular Regularization of Inverse Problems*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2011). 25, 57

BOREL, E., *Leçons sur la théorie des fonctions*, Paris: Gauthier-Villars (1950). 38

BOURBAKI, N., *Elements of Mathematics: Integration I*, Springer-Verlag (2004). 49

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. AND ECKSTEIN, J., *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Machine Learning (2010), 3(1): pp. 1–122. 105, 107, 108, 109, 110, 114, 116, 117, 121, 122, 123, 167

BOYD, S. AND VANDENBERGHE, L., *Convex Optimization*, Cambridge University Press (2004). 51, 55

BRAIDES, A., *Γ-convergence for beginners*, Oxford University Press (2002). 35, 36, 37

Bredies, K. and Pikkarainen, H.K., *Inverse problems in spaces of measures*, ESAIM: Control, Optimization and Calculus of Variations (2013), 19: pp. 190–218. 27, 61, 62, 144, 145

Brune, C., *4D Imaging in Tomography and Optical Nanoscopy*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2010). 25

Burger, M., Mennucci, A.C.G., Osher, S. and Rumpf, M., *Level Set and PDE Based Reconstruction Methods in Imaging*, Springer (2013). 43, 164

Burger, M. and Osher, S., *Convergence rates of convex variational regularization*, Inverse problems (2004), 20(5): p. 1411. 85

Candès, E.J., Li, X., Ma, Y. and Wrigh, J., *Robust principal component analysis?*, Journal of the ACM (2011), 58(3). 85

Candès, E.J. and Plan, Y., *A probabilistic and ripless theory of compressed sensing*, IEEE Transactions on Information Theory (2010), 57(11): pp. 7235–7254. 85

Candès, E.J. and Tao, T., *Decoding by linear programming*, IEEE Transactions on Information Theory (2005), 51(12): pp. 4203 – 4215. 60

Chen, G. and Teboulle, M., *A proximal-based decomposition method for convex minimization problems*, Mathematical Programming (1994), 64: pp. 81–101. 105

Cicalese, M., DeSimone, A. and Zeppieri, C.I., *Discrete-to-continuum limits for strain-alignment-coupled systems: Magnetostrictive solids, ferroelectric crystals and nematic elastomers*, Networks and Heterogeneous Media (2009), 4(4): pp. 667–708. 36

Cohen, A., Dahmen, W. and Devore, R., *Compressed sensing and best k-term approximation*, Journal of the American Mathematical Society (2009), 22: pp. 211–231. 60

Dacorogna, B., *Introduction to the Calculus of Variations*, Imperial College Press (2004). 31

Davis, G., Mallat, S. and Avellaneda, M., *Adaptive greedy approximations*, Constructive Approximation (1997), 13: pp. 57–98. 59

Dirac, P., *Principles of quantum mechanics*, Oxford at the Clarendon Press, 4th edition (1958). 39

Donoho, D.L., *Compressed sensing*, IEEE Transactions on Information Theory (2006), 52(4): pp. 1289–1306. 58

Donoho, D.L. and Elad, M., *Optimally sparse representation in general (non-orthogonal) dictionaries via l1 minimization*, PNAS (2003), 100(5): pp. 2197–2202. 25

Donoho, D.L., Elad, M. and Temlyakov, V.N., *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Transactions on Information Theory (2006), 52: pp. 6–18. 60

Donoho, D.L. and Huo, X., *Uncertainty principles and ideal atomic decomposition*, IEEE Transactions on Information Theory (2001), 47(7): pp. 2845–2862. 60

Duchi, J., Shalev-Shwartz, S., Singer, Y. and Chandra, T., *Efficient projections onto the l1-ball for learning in high dimensions*, Proceedings of the 25th International Conference on Machine Learning (2008). 180

Duval, V. and Peyré, G., *Exact support recovery for sparse spikes deconvolution*, Technical report, CNRS and Université Paris-Dauphine (2013). 27, 144, 161, 172

Eckstein, J. and Fukushima, M., *Some reformulations and applications of the alternating direction method of multipliers*, Large Scale Optimization: State of the Art (1993), pp. 119–138. 105

Ekeland, I. and Témam, R., *Convex Analysis and Variational Problems*, SIAM, corrected reprint edition (1999). 31, 51, 53, 57, 60

Elstrodt, J., *Maß- und Integrationstheorie*, Springer-Verlag (2004). 38, 46, 50

Engl, H.W., Hanke, M. and Neubauer, A., *Regularization of Inverse Problems*, Kluwer Academic Publishers (1996). 23, 85

Finegold, J.A., Asaria, P. and Francis, D.P., *Mortality from ischaemic heart disease by country, region, and age: Statistics from world health organisation and united nations*, International Journal of Cardiology (2013), 168(2): pp. 934–945. 127

Fornasier, M., *Theoretical foundations and numerical methods for sparse recovery*, Radon Series on Computational and Applied Mathematics, De Gruyter (2010). 60

Fornasier, M. and Rauhut, H., *Recovery algorithms for vector valued data with joint sparsity*, SIAM Journal on Numerical Analysis (2008), 46(2): pp. 577–613. 26, 69, 173

Fortin, M. and Glowinski, R., *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, North-Holland: Amsterdam (1983a). 105

Fortin, M. and Glowinski, R., *On decomposition-coordination methods using an augmented lagrangian*, in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, chapter 3, pp. 97–146, North-Holland: Amsterdam (1983b). 105

Fuchs, J.J., *On sparse representations in arbitrary redundant bases*, IEEE Transactions on Information Theory (2004), 50(6): pp. 1341–1344. 97

Fukushima, M., *Application of the alternating direction method of multipliers to separable convex programming problems*, Computational Optimization and Applications (1992), 1: pp. 93–111. 105

Gabay, D., *Applications of the method of multipliers to variational inequalities*, in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, chapter 9, pp. 299–332, North-Holland: Amsterdam (1983). 105

Gabay, D. and Mercier, B., *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Computers and Mathematics with Applications (1976), 2: pp. 17–40. 105

Gilbert, A.C., Strauss, M.J. and Tropp, J.A., *Algorithms for simultaneous sparse approximation part i: Greedy pursuit*, Signal Processing (2006), 86: pp. 572–588. 69

Gilbert, A.C. and Tropp, J.A., *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Transactions on Information Theory (2007), 53: pp. 4655–4666. 60

Giusti, E., *Minimal Surfaces and Functions of Bounded Variation*, volume 80, Monographs in Mathematics, Birkhäuser Boston, Inc. (1984). 45

Glowinski, R. and Marrocco, A., *Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problems de dirichlet non lineares*, Revue Française d'Automatique, Informatique, et Recherche Opérationelle (1975), 9: pp. 41–76. 105

Glowinski, R. and Tallec, P.L., *Augmented lagrangian methods for the solution of variational problems*, Technical report, University of Wisconsin-Madison (1987). 105

Gothika, *Allegory of the cave (plato)*, `https://upload.wikimedia.org/wikipedia/commons/9/97/Allegory_of_the_Cave_blank.png` (2008). 22

HADAMARD, J., *Sur les problèmes aux dérivées partielles et leur signification physique*, Princeton University Bulletin (1902), 13: pp. 49–52. 23

HAHN, H., *Theorie der reellen Funktionen, Band I*, Springer-Verlag (1921). 47

HAIRER, E., NØRSETT, S.P. AND WANNER, G., *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics*, Springer, second revised edition (1993). 187

HE, B.S., YANG, H. AND WANG, S.L., *Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities*, Journal of Optimization Theory and Applications (2000), 106(2): pp. 337–356. 109

HEINS, P., *Sparse Model-Based Reconstruction In Dynamic Positron Emission Tomography*, Master's thesis, Westfälische Wilhelms-Universität Münster (2011). 25, 174

HEINS, P., MÖLLER, M. AND BURGER, M., *Locally sparse reconstruction using $\ell^{1,\infty}$-norms*, Technical report, Westfälische Wilhelms-Universität Münster (2014). 73, 105

HIRIART-URRUTY, J.B. AND LEMARÉCHAL, C., *Convex Analysis and Minimization Algorithms I, Grundlehren der mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), A Series of Comprehensive Studies in Mathematics*, Springer (1993). 51, 98

HSU, D., KAKADE, S.M. AND ZHANG, T., *Robust matrix decomposition with sparse corruptions*, IEEE Transactions on Information Theory (2011), 57: pp. 7221–7234. 85

IOFFE, A.D. AND TIHOMIROV, V.M., *Theory of Extremal Problems*, North-Holland Publishing Company (1979). 56

JUDITSKY, A. AND NEMIROVSKI, A., *On verifiable sufficient conditions for sparse signal recovery via l1-minimization*, Mathematical Programming (2011), 127(1): pp. 57–88. 60, 97

KAKUTANI, S., *Concrete representation of abstract (m)-spaces. (a characterization of the space of continuous functions.)*, Annals of Mathematics (1941), 42(2): pp. 994–1024. 49

KOWALSKI, M., *Sparse regression using mixed norms*, Applied and Computational Harmonic Analysis (2009), 27(3): pp. 303–324. 65, 71

Kowalski, M. and Torrésani, B., *Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients*, Signal, Image and Video Processing (2009), 3(3): pp. 251–264. 65, 71

Kunis, S. and Rauhut, H., *Random sampling of sparse trigonometric polynomials ii – orthogonal matching pursuit versus basis pursuit*, Foundations of Computational Mathematics (2008), 8(6): pp. 737–763. 60

Langner, J., *Development of a Parallel Computing Optimized Head Movement Correction Method in Positron Emission Tomography*, Master's thesis, University of Applied Sciences Dresden and Research Center Dresden-Rossendorf (2003). 128

Lebesgue, H., *Intégrale, longueur, aire*, Ph.D. thesis, Université de Paris (1902). 39

Mallat, S., *A Wavelet Tour of Signal Processing*, Elsevier Academic Press, 3rd edition (2008). 58

Markov, A., *On mean values and exterior densities*, Recueil Mathématique Moscou (1938), 4: pp. 165–190. 49

Maso, G.D., *An introduction to* Γ*-convergence*, Birkhäuser, Basel (1993). 35

Moeller, M., Brinkmann, E.M., Burger, M. and Seybold, T., *Color bregman tv*, Technical report, Technische Universität München and Westfälische Wilhelms-Universität Münster (2014). 70

Möller, M., *Multiscale Methods for (Generalized) Sparse Recovery and Applications in High Dimensional Imaging*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2012). 57

Natarajan, B.K., *Sparse approximate solutions to linear systems*, SIAM Journal on Computing (1995), 24(2): pp. 227–234. 59

Needell, D. and Vershynin, R., *Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit*, Foundations of Computational Mathematics (2009), 9: pp. 317–334. 60

Nikodym, O., *Sur une généralisation des intégrales de M. J. Radon*, Fundamenta Mathematicae (1930), 15: pp. 131–179. 48

Plato, *The Allegory of the Cave*, P & L Publications (2010). 21

Rakbud, J. and Ong, S.C., *Sequence spaces of operators in l2*, Journal of the Korean Mathematical Society (2011), 48(6): pp. 1125–1142. 64

RAUHUT, H., *On the impossibility of uniform sparse reconstruction using greedy methods*, Sampling Theory in Signal and Image Processing (2008), 7: pp. 197–215. 60

READER, A.J., *Fully 4d image reconstruction by estimation of an input function and spectral coefficients*, IEEE Nuclear Science Symposium Conference Record (2007), M17-1: pp. 3260–3267. 130

RIESZ, F., *Sur les opérations fonctionnelles linéaires*, Comptes Rendus de l'Académie des Sciences Paris (1909), 149: pp. 974–977. 49

ROCKAFELLAR, R.T., *Convex Analysis (Princeton Mathematical Series)*, Princeton University Press (1970). 51, 55, 56

ROCKAFELLAR, R.T., *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization (1976), 14(5). 5, 105, 107, 109

ROCKAFELLAR, R.T. AND WETS, R.J.B., *Variational Analysis*, volume 317, Springer (1997). 55

RUDIN, L.I., OSHER, S. AND FATEMI, E., *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena (1992), 60(1-4): pp. 259–268. 25

RUDIN, W., *Functional Analysis*, McGraw-Hill Book Company (1973). 35

RUDIN, W., *Real and Complex Analysis*, McGraw-Hill Book Company, 3rd edition (1987). 40, 46

SAMARAH, S., OBEIDAT, S. AND SALMAN, R., *A schur test for weighted mixed-norm spaces*, Analysis Mathematica (2005), 31: pp. 277–289. 65, 66

SCHERZER, O. AND WALCH, B., *Sparsity regularization for radon measures*, in *Scale Space and Variational Methods in Computer Vision*, Springer Berlin Heidelberg (2009) pp. 452–463. 27, 144

SCHUSTER, T., KALTENBACHER, B., HOFMANN, B. AND KAZIMIERSKI, K.S., *Regularization methods in Banach spaces*, volume 10, Walter de Gruyter (2012). 85

SCHWARTZ, L., *Radon measures on arbitrary topological spaces and cylindrical measures*, Oxford University Press London (1973). 46, 50

TESCHKE, G. AND RAMLAU, R., *An iterative algorithm for nonlinear inverse problems with joint sparsity constraints in vector-valued regimes and an application to color image inpainting*, Inverse Problems (2007), 23: pp. 1851–1870. 26, 69, 173

TIBSHIRANI, R., *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B (Methodological) (1996), 58(1): pp. 267–288. 26

TIBSHIRANI, R., *Regression shrinkage and selection via the lasso: a retrospective*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) (2011), 73(3): pp. 273–282. 26

TIKHONOV, A., *On the stability of inverse problems*, Doklady Akademii nauk SSSR (1943), 39(5): pp. 195–198. 25

TROPP, J.A., *Greed is good: Algorithmic results for sparse approximation*, IEEE Transactions on Information Theory (2004), 50: pp. 2231–2242. 60

TROPP, J.A., *Algorithms for simultaneous sparse approximation part ii: Convex relaxation*, Signal Processing (2006a), 86(3): pp. 589 – 602. 65, 69, 83

TROPP, J.A., *Just relax: Convex programming methods for identifying sparse signals in noise*, IEEE Transactions on Information Theory (2006b), 52(3): pp. 1030–1051. 97

TSENG, P., *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM Journal on Control and Optimization (1991), 29: pp. 119–138. 105

U. S. NATIONAL CANCER INSTITUTE, *Physiology of circulation - roles of capillaries*, `http://training.seer.cancer.gov/anatomy/cardiovascular/blood/physiology.html` (2006). 129

U. S. NATIONAL HEART, LUNG AND BLOOD INSTITUTE, *What is coronary heart disease?*, `http://www.nhlbi.nih.gov/health/health-topics/topics/cad/printall-index.html` (2012). 128

VARDI, Y., SHEPP, L.A. AND KAUFMAN, L., *A statistical model for positron emission tomography*, Journal of the American Statistical Association (1985), 80(389): pp. 8–20. 128

WANG, S.L. AND LIAO, L.Z., *Decomposition method with a variable parameter for a class of monotone variational inequality problems*, Journal of Optimization Theory and Applications (2001), 109(2): pp. 415–429. 109

WERNICK, M.N. AND AARSVOLD, J.N., *Emission Tomography: The Fundamentals of PET and SPECT*, Elsevier Academic Press (2004). 129, 132

YUAN, M. AND LIN, Y., *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) (2006), 68(1): pp. 49–67. 26

ZEIDLER, E., *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*, Springer, New York (1985). 31, 33

ZHANG, C.H. AND ZHANG, T., *A general framework of dual certificate analysis for structured sparse recovery problems*, Technical report, Rutgers University, NJ (2012). 85