Andreas Habring, BSc

A Generative Variational Model for Inverse Problems in Imaging

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Mathematics

submitted to

Graz University of Technology

Supervisor

Ass.-Prof. Mag. Dr.rer.nat. Martin Holler

Institut für Mathematik und Wissenschaftliches Rechnen

Karl-Franzens-Universität Graz

Graz, August 2020

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date, Signature

Abstract

This thesis is concerned with the development and analysis of a novel regularization method for mathematical image processing. The proposed method is a combination of total variation-type regularization with deep learning methods. Total variation regularization is mathematically well understood, but unfortunately performs rather badly on images containing texture. On the contrary, deep learning methods are empirically known to be well suited also for texture in images, but often lack an underlying mathematical foundation. The proposed method aims to combine the strengths of both types of methods. The thesis contains a chapter providing a mathematical analysis in a continuous setting, including existence and stability results, as well as a chapter with an application of the method in a discrete setting on different test images.

Kurzfassung

Diese Masterarbeit ist der Entwicklung und mathematischen Analyse einer neuartigen Regularisierungsmethode im Bereich der mathematischen Bildverarbeitung gewidmet. Die vorgeschlagene Regularisierung kombiniert klassische variationelle Methoden, wie etwa Regularisierung mittels totaler Variation, mit aktuellen machine-learning Techniken. Regularisierung mit totaler Variation ist ein mathematisch gut verstandenes Verfahren und eignet sich hervorragend für stückweise glatte Bilder, allerdings sind die experimentellen Resultate noch deutlich verbesserungswürdig bei Bildern mit feineren Mustern/Textur. Im Gegensatz dazu erzielen viele machine-learning Verfahren vielversprechende Ergebnisse auch mit texturlastigen Bildern, es fehlt allerdings häufig eine zugrunde liegende mathematische Theorie. Mit der vorgeschlagenen Methode versuchen wir, die Stärken beider erwähnten Typen von Verfahren zu kombinieren. Diese Arbeit beinhaltet sowohl die theoretische Behandlung der vorgeschlagenen Methode inklusive Resultaten, die die Existenz von Lösungen und Stabilität der Methode sicherstellen, als auch Beispiele zur praktischen Anwendung unseres Verfahrens mit verschiedenen Testbildern.

Acknowledgements

Firstly, I would like to thank my parents for enabling me to live a life mostly free of worries, which played a crucial role also in enjoying my university education. Secondly, I would like to thank Dr. Martin Holler, who introduced me to the topic of this thesis and supported my work, whenever I needed help. Lastly, I would like to thank my girlfriend Paola, who has been on my side for many years now. You have always been an important source of support and balance in my life and altogether my favorite human being.

Contents

List of Figures

List of Tables

List of Acronyms and Symbols

d A natural number, denoting the space dimension as in $\mathbb{R}^d$

Ω If not stated differently, $\Omega \subseteq \mathbb{R}^d$ is a bounded, open domain with compact Lipschitz boundary.

$\overline{\mathbb{R}}$ The extended real line, $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$.

$\mathcal{P}(A)$ The power set, that is, $\mathcal{P}(A)$ denotes the set of all subsets of A.

$\Subset$ For $A, B \subset \mathbb{R}^d$, we write $A \Subset B$, if $\overline{A}$ is compact and $\overline{A} \subset B$, i.e., the closure of A is a compact subset of B.

$\mathcal{L}^d$ The d dimensional Lebesgue measure.

a.e. a.e. is short for "almost everywhere" or "almost every". A property is true a.e. in Ω, if there exists a Lebesgue measurable set $N \subseteq \Omega$ with $\mathcal{L}^d(N) = 0$, such that said property is true for all $x \in \Omega \setminus N$.

$|\,.\,|_p$ For $z \in \mathbb{R}^d$, if $1 \leq p < \infty$, $|z|_p := (\sum\limits_{i=1}^{n} |z_i|^p)^{\frac{1}{p}}$ and in the case $p = \infty$, $|z|_\infty := \max\limits_{i=1,2,\dots d} |z_i|$.

$|\,.\,|$ $|\,.\,|_2$

$\|.\|_p$ For $1 \leq p < \infty$ the L^p-Norm, i.e., for $f : \Omega \to \mathbb{R}^n$ Lebesgue measurable, $\|f\|_p := (\int\limits_\Omega |f(x)|_2^p dx)^{1/p}$.

$\|.\|_\infty$ The L^∞-Norm, i.e., for $f : \Omega \to \mathbb{R}^n$ Lebesgue measurable, $\|f\|_\infty := \inf\limits_{\substack{N \subseteq \Omega \\ \mathcal{L}^d(N)=0}} \sup\limits_{x \in \Omega \setminus N} |f(x)|_2$.

$K^n(\Omega)$ The space of constant functions, $K^n(\Omega) = \{f : \Omega \to \mathbb{R}^n \mid \exists w \in \mathbb{R}^n : f(x) = w \text{ for a.e. } x \text{ in } \Omega\}$

$\mathcal{L}^p(\Omega, \mathbb{R}^n)$ For $1 \leq p \leq \infty$ the set $\{f : \Omega \to \mathbb{R}^n | f \text{ measurable}, \|f\|_p < \infty\}$.

$L^p(\Omega,\mathbb{R}^n)$ The set of all equivalence classes of functions in $\mathcal{L}^p(\Omega,\mathbb{R}^n)$, where an equivalence class consists of all functions, which are equal on Ω up to a set of Lebesgue measure zero.

$L^p(\Omega)$ $L^p(\Omega,\mathbb{R})$

$C(\Omega,\mathbb{R}^n)$ The set of all continuous functions $f:\Omega\to\mathbb{R}^n$. This definition is also valid for more general Ω.

$C(\Omega)$ $C(\Omega,\mathbb{R})$

$\partial^\alpha f$ For $f:\Omega\to\mathbb{R}$ and $\alpha\in\mathbb{N}^d$, $\partial^\alpha f := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}}\dots\frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}f$. $|\alpha| := \alpha_1+\alpha_2+\dots+\alpha_d$ is the order of the derivative.

$C^k(\Omega,\mathbb{R}^n)$ The set of all functions $f:\Omega\to\mathbb{R}^n$, that are k times continuously differentiable, i.e., $\forall\alpha\in\mathbb{N}^d$, such that $|\alpha|\le k$ it holds $\partial^\alpha f\in C(\Omega,\mathbb{R}^n)$.

$C^k(\overline{\Omega},\mathbb{R}^n)$ The set of all functions $f\in C^k(\Omega,\mathbb{R}^n)$, such that $\forall\alpha\in\mathbb{N}^d$ with $|\alpha|\le k$, $\partial^\alpha f$ can be extended continuously to $\overline{\Omega}$.

$C^k(S)$ $C^k(S,\mathbb{R}^n)$ for $S\in\left\{\Omega,\overline{\Omega}\right\}$.

$C^\infty(S,\mathbb{R}^n)$ The set of all functions, that are infinitely many times differentiable, i.e., $C^\infty(S,\mathbb{R}^n) = \bigcap_{k\in\mathbb{N}} C^k(S,\mathbb{R}^n)$ for $S\in\left\{\Omega,\overline{\Omega}\right\}$.

$C^\infty(S)$ $C^\infty(S,\mathbb{R})$ for $S\in\left\{\Omega,\overline{\Omega}\right\}$.

$C_c^k(\Omega,\mathbb{R}^n)$ The set of all k-times continuously differentiable functions compactly supported in Ω, $C_c^k(\Omega,\mathbb{R}^n) := \{f\in C^k(\Omega,\mathbb{R}^n)\big|\operatorname{supp}(f)\subseteq\Omega\}$, where $\operatorname{supp}(f) := \overline{\{x\in\Omega\big|f(x)\neq 0\}}$.

$C_c^k(\Omega)$ $C_c^k(\Omega,\mathbb{R})$

$C_c^\infty(\Omega,\mathbb{R}^n)$ The set of all smooth functions compactly supported in Ω, $C_c^\infty(\Omega,\mathbb{R}^n) := \{f\in C^\infty(\Omega,\mathbb{R}^n)\big|\operatorname{supp}(f)\subseteq\Omega\}$

$C_c^\infty(\Omega)$ $C_c^\infty(\Omega,\mathbb{R})$

$C_0^k(\Omega,\mathbb{R}^n)$ For $k\in\mathbb{N}\cup\infty$, $\overline{C_0^k(\Omega,\mathbb{R}^n)}$ where the closure is taken with respect to $\|.\|_\infty$.

$C_0^k(\Omega)$ $C_0^k(\Omega,\mathbb{R})$

$\mathcal{L}(X,Y)$ For X,Y Banach spaces, $\mathcal{L}(X,Y) := \{A: X\to Y \mid \text{A is a linear and bounded operator}\}$.

$\operatorname{rg}(T)$ For $T: X\to Y$ linear, $\operatorname{rg}(T) := \{Tx \mid x\in X\}\subseteq Y$ is the so-called range of the linear operator.

$\ker(T)$ For $T : X \to Y$ linear, $\ker(T) \coloneqq \{x \in X \mid Tx = 0\} \subseteq X$ is the so-called kernel of the linear operator.

$\tilde{g}$ For $g : \Omega \to \mathbb{R}^n$, $\tilde{g}$ denotes the zero extension of g, i.e., $\tilde{g} : \mathbb{R}^d \to \mathbb{R}^n$,

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \in \Omega \\ 0 & \text{else.} \end{cases}$$

$\mathcal{I}_M$ The indicator function on the set M, i.e., $\mathcal{I}_M(x) = 0$ if $x \in M$ and $\mathcal{I}_M(x) = \infty$ else.

$\text{size}(\mu)$ For a multidimensional vector μ, $\text{size}(\mu)$ denotes the tuple of lengths of μ in each direction, e.g., for $\mu \in \mathbb{R}^{n \times m \times d}$, $\text{size}(\mu) = (n, m, d)$

Chapter 1

Introduction

In the present thesis, we introduce and investigate a novel method for regularizing linear inverse problems in image processing. We present the method in a continuous setting and prove well-posedness before applying it to images in a discrete setting.

In the following, images are meant to be grayscale images. In the discrete setting, images are described by matrices, where each entry represents a pixel and the value of the entry represents a shade of gray. In the continuous setting, an image is described by a function defined on the image domain and the function value at a point represents the shade of gray of the corresponding point in the image.

1.1 Linear Inverse Problems

In the field of linear inverse problems one usually deals with problems of the form

$$\begin{gathered}\text{given } y \in Y,\\ \text{find } x \in X : Tx = y,\end{gathered} \tag{1.1}$$

where X, Y are normed vector spaces (often Banach spaces) and $T : X \to Y$ is a bounded linear operator. Note, that without further assumptions (1.1) may not even have a solution at all, since T might not be bijective, which already shows the need of some kind of reformulation of the problem. In applications, y stems from a measurement and is therefore distorted by some random noise, which additionally complicates the task described in (1.1). In other words, instead of the idealistic problem (1.1), we rather aim to solve

$$\begin{gathered}\text{given } y^\delta \in Y, \text{ such that } \left\|y - y^\delta\right\| \leq \delta,\\ \text{find } x \in X : Tx = y,\end{gathered} \tag{1.2}$$

where $y \in Y$ is the so-called ground truth data, $\delta > 0$ the noise-level, both unknown, and y^δ is the known, noisy data. Again, finding a solution to (1.2) is not possible in general.

Moreover, since y^δ contains noise, it would be futile to search for x, satisfying $Tx = y^\delta$, since Tx is only known to be in a δ neighborhood of y^δ. A very common approach, often referred to as Tikhonov (type) regularization, is to consider the problem of finding x^δ, such that

$$x^\delta \in \operatorname*{argmin}_{x\in X} \left\|Tx - y^\delta\right\| + \lambda\mathcal{R}(x) \tag{1.3}$$

with $\lambda > 0$. $\mathcal{R} : X \to [0, \infty]$ is a so-called regularization functional, which ensures, that the minimizer of (1.3) satisfies specific properties, i.e., $\mathcal{R}(x)$ has a small value, if x is desirable as a solution and a large value else. By solving (1.3), we balance fitting the data, by keeping $\left\|Tx - y^\delta\right\|$ small, with enforcing our solution to be 'nice', by keeping $\mathcal{R}$ small. In image processing a popular choice of $\mathcal{R}$ is the so-called total variation functional, which is a generalization of the 1-norm of the gradient, $\mathcal{R}(x) = \|\nabla x\|_1$ for an image x [5]. As the name suggests, the total variation measures, how much variation/oscillation is contained in an image. By the very nature of the total variation functional, using it as a regularization in (1.3) enforces the resulting image to be piecewise smooth or even piecewise constant. The reason for choosing such a regularization is mainly due to the fact, that random noise is naturally of high variation. Hence, in order to minimize noise, we enforce low variation on the image. Unfortunately, when we aim to process images, which contain a lot of texture (e.g., an image of a person with a plaid shirt), regularization with total variation turns out to be counterproductive as shown in Figure 1.1. Note, that after the reconstruction in Figure 1.1 in the second example, the textures on the trousers of the pictured woman and on the tablecloth are partly removed. For problems of this kind, machine/deep learning algorithms using convolutional neural networks have shown to be rather successful.

1.2 Convolutional Neural Networks in Image Processing - Deep Image Prior

Although, there are a lot of subtle details, that can be added, a basic convolutional neural network (CNN) consists of multiple layers, where each layer is usually composed of a convolution-type operation, the adding of a constant, called bias, and the application of a non-linear activation function. Formally, in imaging a layer in a CNN processes input images $(z_i)_{i=1}^m = z$ as follows

$$z \mapsto \phi(Wz + b), \tag{1.4}$$

where $(Wz)_j = \sum_{i=1}^{m} k_{i,j} * z_i$ for $j = 1, 2, ...k$, i.e., for each j, $(Wz)_j$ is a sum of convolutions of the input images with convolution kernels $k_{i,j}$, b is the bias and ϕ the activation function, which is applied pointwise to its input images. A common example for ϕ is the ReLu (rectified linear unit), $\phi(t) = \max\{0, t\}$ for $t \in \mathbb{R}$. Note, that such a layer of a neural network has multiple images as input and output in general. The convolution kernels $k_{i,j}$ as well as the bias b are parameters of this convolutional layer and the dimensions of input

Figure 1.1: Denoising using total variation regularization. In both cases, we used additive, Gaussian random noise with zero mean and variance 0.05. The first column pictures the noisy images, the second the results after total variation regularization.

and output are part of the architecture of the layer. A CNN is a composition of multiple such layers, i.e., a basic CNN with n layers acting on an input z can be written as

$$z \mapsto \phi\left(W_n\left(...\phi\left(W_2\phi\left(W_1 z + b_1\right) + b_2\right)...\right) + b_n\right), \tag{1.5}$$

where W_i and b_i are the convolution operator and bias of layer i. In Figure 1.2, there is a simplified sketch of a neural network with two layers. The green nodes are the input images, the yellow node is the output image and the lines connecting the nodes denote operations of the type of (1.4). The blue nodes are the intermediate result of the first layer of this network. Note, that the number of images in each layer is variable and can in principle be chosen arbitrarily, but the output layer shall be a single image in our case. In the following, let us denote a CNN as a function f_θ, where θ are all parameters of the network. In [15], as a method for image reconstruction, authors tackle (1.2) by solving

$$\min_\theta \left\| T f_\theta(z) - y^\delta \right\|, \tag{1.6}$$

where z is the input image of the neural network, which is fixed (in [15], z is e.g. initialized as random noise). The reconstructed image is afterwards obtained as $f_\theta(z)$. This means,

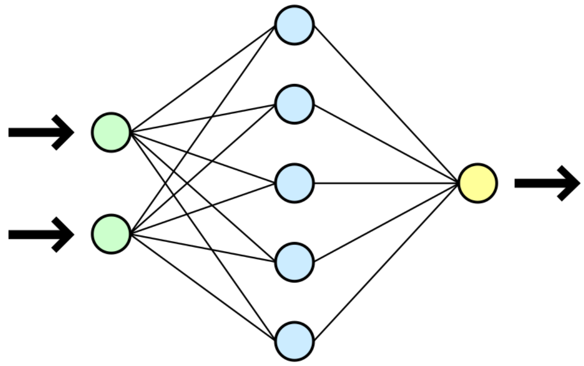


Figure 1.2: Sketch of a neural network [11]

that instead of adding a functional $\mathcal{R}$, the regularization happens implicitly, by only allowing for images, which are the output of a given CNN. This has become popular under the name deep image prior (DIP). Formulating (1.6) as

$$\min_x \left\| Tx - y^\delta \right\| + \mathcal{I}_M(x), \tag{1.7}$$

where $M = \{f_\theta(z) \mid \theta\}$ and $\mathcal{I}_M(x)$ evaluates to zero, if there exist parameters θ, such that $f_\theta(z) = x$ and to infinity otherwise, shows, that (1.7) fits into the framework of (1.3) with $\mathcal{R}$, ensuring, that the image is the output of the given CNN with some parameters. However, $\mathcal{I}_M$ does not satisfy the necessary conditions of a regularization functional in order to ensure existence of solutions and stability. In the present work, we modify the set M in such a way, that it becomes closed and bounded, which makes $\mathcal{I}_M$ a suitable regularization functional. Despite the experimental results in [15] being very successful, there are also downsides to this and similar methods. Usually, deep learning algorithms as in [15] lack a mathematical foundation. Mostly, there is no underlying continuous model and no guaranteed well-posedness of the method. Moreover, in order to obtain a visually satisfying result, methods similar to (1.6) often have to be combined with counter-intuitive stopping rules [15, 23], i.e., the algorithm, used to solve the minimization problem, has to be stopped prematurely, which means, that the true minimizer of the considered objective functional is in general not the desired result, hence, the applied model is still flawed/incomplete. The reason for this issue is, that a CNN of sufficient size may be able to produce almost any given image, even random noise [23], which implies, that effectively there is no regularization in (1.7). The CNN solely changes the way, the image space is searched for a solution with a specific algorithm and therefore a stopping rule is needed, in order to avoid overfitting of the noisy data.

1.3 The Proposed Method

The aim of the method proposed in the present work, is to combine the strengths of recent machine learning methods like [15] with the mathematical foundation of variational methods like total variation based regularization. The idea is to implicitly perform a decomposition of the image into a cartoon part (piecewise smooth) and a texture part, which is the output of a CNN. The resulting problem is of the form

$$\min_{u,v} \mathcal{D}(Tu, y^\delta) + \lambda \mathcal{R}(u - v) + \mu \mathcal{G}(v), \tag{1.8}$$

where y^δ is the given data, $\mathcal{D}$ is a functional measuring the data fidelity of u, v is the texture part of u and $u - v$ the cartoon part, $\mathcal{R}$ is a total variation-like functional ensuring that the cartoon part is piecewise smooth and $\mathcal{G}$ is a further regularization which, among other things, ensures, that v is generated from a convolutional network. To be a little more precise, given a network architecture, $\mathcal{G}$ ensures, that there exist parameters θ and input z, such that $f_\theta(z) = v$, while penalizing the norms of θ and z. $\lambda, \mu > 0$ are parameters to balance the data fidelity and the regularization. The main difference of this thesis compared to existing publications, which focus on similar problems [15, 23, 19, 22, 26] is the introduction and investigation of the problem in a continuous setting including a rigorous proof of existence of a solution and continuous dependence of the solution on the data.

1.4 Related Works

The already mention deep image prior introduced in [15] has been a great incentive for research in the direction the present work is headed. The deep image prior performs rather well in experiments, but lacks a mathematical investigation. Although, most research about CNNs in image processing is rather experimental, there is also some work directed more towards a mathematical point of view. We will list some references in the following.

In [9], a single layer version of the model, provided in the present work, containing only one convolutional layer is introduced and theoretically investigated as well as tested on images, showing promising results for images comprised of texture and piecewise smooth parts. The minimization problem considered in [9] is of the form

$$\min_{u,v} \mathcal{D}(Tu, y^\delta) + \lambda \mathcal{R}(u - v) + \mu \mathcal{G}(v), \tag{1.9}$$

where $\mathcal{R}$ is a total variation like functional used as a cartoon prior. The definition of the functional $\mathcal{G}$ is somewhat subtle. The idea is, that $\mathcal{G}$ shall force v to be a sum of convolutions, i.e., $v = \sum_{i=1}^n c_i * p_i$ for some n and c_i, p_i. Moreover, $\mathcal{G}$ penalizes the norms of c_i, p_i and large n. The authors further use relaxation and lifting strategies, to obtain well-posedness. To be more precise, they introduce a tensor space, enabling a linear lifting of the convolution. The main difference to the present work is, that we modify the representation of v by employing a deeper convolutional network (instead of one layer). Further, we do not use similar lifting techniques in order to show well-posedness of the problem.

In [6, 18], a neural network combined with additional regularization is used for solving inverse problems. In [6], authors consider a problem of the form

$$\min_{\theta} \mathcal{D}(T f_\theta(z), y^\delta) + \lambda \mathcal{R}(f_\theta(z)), \tag{1.10}$$

where $\mathcal{R}$ is a regularizing functional, f_θ is a continuous function with the structure of a neural network and z is a fixed input for f_θ. The problem is introduced in general Banach spaces and well-posedness and convergence under some assumptions on the used functions and spaces are ensured. A main difference to our method, is that we penalize θ and z instead of $f_\theta(z)$. The problem formulation in [18] is of the form

$$\min_{z} \mathcal{D}(T f_\theta(z), y^\delta) + \lambda \|z\| . \tag{1.11}$$

Also in [18], authors provide proofs for well-posedness and convergence of the method. A further difference to both [6, 18] is the implicit cartoon-texture decomposition we employ, which is motivated by the particular application in image processing.

In [23], authors combine a deep image prior with total variation regularization leading to

$$\min_{\theta} \left\| T f_\theta(z) - y^\delta \right\|^2 + \lambda TV(f_\theta(z)) + \mu \mathcal{R}(\theta), \tag{1.12}$$

where TV is the total variation, f_θ is a neural network, z a fixed input and $\mathcal{R}$ is a learnt regularization functional (i.e., it uses some parameters, that have to be learnt from data before). Contrary to the present work, the total variation regularization is applied to the output of the neural network, which in our model shall be the texture of the image. Moreover, the problem in [23] is only considered in a discrete setting.

In [12], a different point of view is introduced, showing that under some conditions, a deep image prior can be interpreted as learning an optimal Tikhonov functional instead of training a neural network.

A field, that is also related to the problem of concern in this work is sparse coding [22, 25, 26, 2, 14, 19, 21, 1]. The sparse coding problem is usually formulated as follows. Given a dictionary $D \in \mathbb{R}^{m \times n}$ and a vector $y \in \mathbb{R}^m$, we ought to find a sparse vector $x \in \mathbb{R}^n$ (i.e., x only has few non-zero entries), such that $y = Dx$. In many cases, the sparsity of x is relaxed to enforcing small 1-norm. Convolutional sparse coding denotes the special case, where the matrix D is assumed to be the matrix representation of a convolution operator, i.e., $Dx = \sum_{i=1}^{k} d_i * x$ for some vectors d_i. The dictionary D is often supposed to be known already. Typically, the dictionary is learned from problem specific data, i.e., assume we are given n training samples $(y_i)_{i=1}^n$, then D is determined by solving

$$\min_{(x_i)_i, D} \sum_{i=1}^{n} \|y_i - Dx_i\|_2^2 + \lambda \mathcal{R}((x_i)_i) + \mu \mathcal{G}(D) \tag{1.13}$$

where, $\mathcal{R}$ and $\mathcal{G}$ are regularizing functionals. Often, $\mathcal{R} = \| \, . \, \|_1$ to enforce before mentioned sparsity of the x_i and $\mathcal{G}$ is either a norm or some constraint on the dictionary D [25, 2]. Note, that if in (1.13) we use as a single training sample the given noisy data, we obtain a problem very similar to the approach of this work. In [22], even a multilayer convolutional sparse coding method is discussed, i.e., instead of the single dictionary D, a composition of convolutional dictionaries $D_N D_{N-1} ... D_1$ is used. Although the model is again only analysed in a discrete setting and differs in the cartoon texture decomposition we utilize, it is similar in the sense, that one aims to represent a given image as the outcome of a composition of convolutions. Worth mentioning is also the solution algorithm for the dictionary learning (1.13) in [22]. Omitting a lot of details, authors propose an algorithm, which iteratively for each training sample y_i first finds a sparse representation x_i, such that $y_i \approx D_N D_{N-1} ... D_1 x_i$ for fixed dictionaries $(D_k)_{k=1}^N$. Afterwards, the dictionaries are updated by performing several steps of projected gradient descent of the problem (1.13) with respect to the D_k. The algorithm we use to solve our discrete problem, on the other hand, minimizes with respect to all variables simultaneously.

The works [26, 19] present methods for cartoon texture decomposition via convolutional sparse coding utilizing the total variation in a discrete setting. In [26], the dictionary is assumed to be known, which can be compared to assuming, that the CNN in our approach was pretrained. In [19], this results in a minimization problem similar to a single layer version of the one we consider, namely

$$\min_{x_C, \Gamma_T, D_T} \|y - x_C - D_T \Gamma_T\|_2^2 + \lambda TV(x_C) + \mu \|\Gamma_T\|_1 , \tag{1.14}$$

where y is the given data, x_C the cartoon part of the image, D_T a texture dictionary, Γ_T a (sparse) vector, such that $D_T \Gamma_T$ is the texture part of the image and $TV(x_C)$ is the total variation of x_C.

Chapter 2

Theoretical Results

2.1 Functional Analytic Background

In order to formulate and analyze the proposed method for image processing, we will need some tools from functional analysis, measure theory and variational calculus. In this section, we will provide the most important definitions and results for later use.

2.1.1 Miscellaneous Results and Definitions

Lemma 2.1. *[Uniformly continuous extension, [3, U2.18 Stetige Fortsetzung]] Let X be a metric space and Y a complete metric space. Further, let $A \subset X$ be dense in X. Then for every uniformly continuous $f : A \to Y$, there exists a unique uniformly continuous extension $\hat{f} : X \to Y$.*

Definition 2.2. *[Continuous, linear functions, [24, Satz II.1.4]] Let X, Y be normed vector spaces. Then we denote the space of all continuous, linear mappings $X \to Y$ as $\mathcal{L}(X, Y)$. For $T \in \mathcal{L}(X, Y)$, we define the norm*

$$\|T\| := \sup_{\substack{x \in X \\ \|x\| \leq 1}} \|Tx\|,$$

which makes $\mathcal{L}(X, Y)$ a normed space. If Y is a Banach space, then $\mathcal{L}(X, Y)$ is a Banach space as well.

Theorem 2.3. *[Uniform boundedness principle, [24, Theorem IV.2.1]] Let X be a Banach space, Y a normed space, I an index set and $T_i \in \mathcal{L}(X, Y)$ for all $i \in I$. Then*

$$\sup_{i \in I} \|T_i x\| < \infty \quad \forall x \in X$$

implies, that

$$\sup_{i \in I} \|T\| < \infty.$$

Definition 2.4. *[Dual space] Let X be a normed vector space over the field $\mathbb{K}$. Then the dual space of X, denoted X^*, is defined as the space of all continuous, linear functionals on X, i.e., $X^* \coloneqq \mathcal{L}(X, \mathbb{K})$. Moreover, we define on X^* the dual norm*

$$\|x^*\| \coloneqq \sup_{\substack{x \in X \\ \|x\| \leq 1}} |x^*(x)|.$$

Further, we will denote the dual space of X^* as X^{**} and call it the bidual (space) of X.

Lemma 2.5. *[24, Korollar II.2.2] The dual space of a normed vector space equipped with the dual norm is a Banach space.*

Definition 2.6. *[Adjoint operator, [24, Definition III.4.1]] Let X and Y be normed spaces and $A \in \mathcal{L}(X, Y)$. We define the adjoint (operator) of A, $A^* \in \mathcal{L}(Y^*, X^*)$, via*

$$(A^* y^*)(x) = y^*(Ax), \quad \text{for } x \in X.$$

Definition 2.7. *[Weak/weak* convergence] Let X be a normed vector space.*

- *We say, that $(x_n)_n \subset X$ converges weakly to $x \in X$ and write $x_n \rightharpoonup x$ (as $n \to \infty$), if for every $x^* \in X^*$, we have $x^*(x_n) \to x^*(x)$ as $n \to \infty$.*
- *We say, that $(x_n^*)_n \subset X^*$ converges to $x^* \in X^*$ in the weak* topology and write $x_n^* \overset{*}{\rightharpoonup} x^*$ (as $n \to \infty$), if for every $x \in X$, we have $x_n^*(x) \to x^*(x)$ as $n \to \infty$.*

Remark 2.8. *Weak* convergence can also be regarded as pointwise convergence of a sequence of functionals.*

Lemma 2.9. *Let X and Y be Banach spaces and $T \in \mathcal{L}(X, Y)$. Then T is weak-to-weak continuous, i.e., if $x_n \rightharpoonup x$ in X, then $Tx_n \rightharpoonup Tx$ in Y.*

Proof. Let $x_n \rightharpoonup x$ in X. Then, denoting $T^* \in \mathcal{L}(Y^*, X^*)$ the adjoint operator of T (see Definition 2.6), we find for arbitrary $y^* \in Y^*$,

$$y^*(Tx_n) \underbrace{=}_{\text{defin. of } T^*} (T^* y^*)(x_n) \underbrace{\xrightarrow{n \to \infty}}_{(T^* y^*) \text{ continuous}} (T^* y^*)(x) = y^*(Tx).$$

Hence, for any $y^* \in Y^*$, $y^*(Tx_n) \to y^*(Tx)$, that is $Tx_n \rightharpoonup Tx$ in Y. □

Lemma 2.10. *In the dual space of a Banach space X, every weak* convergent sequence is bounded.*

Proof. Let $(x_n^*)_n \subset X^* = \mathcal{L}(X, \mathbb{K})$ be weak* convergent to $x^* \in X^*$. Then

$$\sup_{n \in \mathbb{N}} |x_n^*(x)| < \infty \quad \forall x \in X,$$

due to convergence. Therefore, the uniform boundedness principle, Theorem 2.3, implies, that

$$\sup_{n\in\mathbb{N}} \|x_n^*\| < \infty,$$

concluding the proof. □

Definition 2.11. *Let X be a Banach space. We define the the canonical embedding i_X as*

$$\begin{aligned} i_X : X &\to X^{**} \\ x &\mapsto i_X(x). \end{aligned}$$

with

$$\begin{aligned} i_X(x) : X^* &\to \mathbb{K} \\ x^* &\mapsto x^*(x). \end{aligned}$$

Lemma 2.12. *[[24, Satz III.3.1]] Let X be a Banach space. Then the mapping i_X from Definition 2.11 is a linear isometry.*

Definition 2.13. *[Reflexive space] A Banach space X is called reflexive, if i_X is surjective.*

Theorem 2.14. *[[24, Theorem III.3.7]] In a reflexive Banach space, every bounded sequence admits a weakly convergent subsequence.*

Definition 2.15. *[Separable space] A metric space is called separable, if it admits a countable, dense subset.*

Theorem 2.16. *[Banach-Alaoglu theorem, [3, 6.5 Satz]] Let X be a separable Banach space. Then the closed unit ball in X^* is sequentially compact with respect to the weak* topology, i.e., every bounded sequence in X^* admits a weak* convergent subsequence.*

Definition 2.17. *[Precompact sets, [3, 2.5 Kompaktheit]] Let (X, d) be a metric space. $A \subset X$ is called precompact, if for every $\epsilon > 0$, there exists $n \in \mathbb{N}$ and $x_1, x_2, ..., x_n \in A$ such that*

$$A \subset \bigcup_{i=1}^{n} B_\epsilon(x_i),$$

where $B_\epsilon(x_i) \coloneqq \{z \in X \mid d(x_i, z) < \epsilon\}$.

Lemma 2.18. *[[3, 2.6 Bemerkungen]] Let (X, d) be a complete metric space. Then $A \subset X$ is precompact if and only if $\overline{A}$ is compact.*

Corollary 2.19. *Let (X, d) be a complete metric space. Then $A \subset X$ is precompact if and only if every sequence in A admits a convergent subsequence in X.*

Proof." $\Rightarrow$ ": Assume A is precompact and let $(x_n)_n$ be a sequence in A. Then $(x_n)_n$ is also a sequence in $\overline{A}$, which is compact by Lemma 2.18. Hence, there is a convergent subsequence in $\overline{A} \subset X$.

"$\Leftarrow$": Assume, every sequence in A admits a convergent subsequence in X. By Lemma 2.18, we need to show, that $\overline{A}$ is compact. So let $(x_n)_n$ be a sequence in $\overline{A}$. For every n, pick $y_n \in A$, such that $d(y_n, x_n) < \frac{1}{n}$. By assumption, there exists a convergent subsequence of $(y_n)_n$ in X, i.e., $y_{n_k} \to y$ as $k \to \infty$ for some $y \in X$. For this y we find,

$$d(x_{n_k}, y) \leq d(x_{n_k}, y_{n_k}) + d(y_{n_k}, y) \leq \frac{1}{n_k} + d(y_{n_k}, y) \xrightarrow{k\to\infty} 0,$$

i.e., $x_{n_k} \to y$ as $k \to \infty$. Moreover, since $\overline{A}$ is closed and $(x_{n_k})_k \subset \overline{A}$, also $y \in \overline{A}$. Hence, every sequence in $\overline{A}$ admits a convergent subsequence, whose limit is in $\overline{A}$, that is, $\overline{A}$ is compact.

□

Theorem 2.20. *[Arzelà-Ascoli theorem, [3, Satz 2.11]] Let $K \subset \mathbb{R}^d$ be compact and $A \subset C(K)$, which is equipped with the uniform norm. Then A is precompact if and only if*

1. $\sup\limits_{f\in A} \|f\|_\infty < \infty$ *(A is bounded) and*
2. $\sup\limits_{\substack{x,y\in K\\ |x-y|<\delta}} \sup\limits_{f\in A} |f(x) - f(y)| \to 0$ *as* $\delta \to 0$ *(A is uniformly equicontinuous).*

Corollary 2.21. *Let $K \subset \mathbb{R}^d$ be compact and $(f_n)_n \subset C(K)$. Assume*

1. $\sup\limits_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$ *and*
2. $\sup\limits_{\substack{x,y\in K\\ |x-y|<\delta}} \sup\limits_{n\in\mathbb{N}} |f_n(x) - f_n(y)| \to 0$ *as* $\delta \to 0$.

Then there exists a uniformly convergent subsequence of $(f_n)_n$, i.e. there exist $(f_{n_k})_k$ and $f \in C(K)$, such that

$$\|f_{n_k} - f\|_\infty \to 0 \quad \textit{as } n \to 0.$$

Proof. The set $\{f_n \mid n \in \mathbb{N}\}$ satisfies the conditions of Theorem 2.20, therefore, it is precompact. Since $C(K)$ is a Banach space with respect to the uniform norm, using Corollary 2.19, we see, that $(f_n)_n$ admits a convergent subsequence in $C(K)$, concluding the proof. □

2.1.2 Measure Theory

Definition 2.22. *[Measures] Let T be a set and $\mathcal{A}$ a σ-algebra on T.*

- *A function $\mu : \mathcal{A} \to \mathbb{R}$ (or $\mathbb{C}$) is called a signed (or complex) measure, if μ is σ-additive, i.e. for any sequence of pairwise disjoint sets $A_n \in \mathcal{A}$*

$$\mu(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu(A_n).$$

- *A function $\mu : \mathcal{A} \to [0, \infty]$ is called a positive measure, if*
  - $\mu(\emptyset) = 0$,
  - *μ is σ-additive.*

  We say, that μ is finite if $\mu(T) < \infty$ and that μ is σ-finite, if there exists a sequence of sets $(B_n)_n \subset \mathcal{A}$, such that $B_n \subset B_{n+1}$ and $\mu(B_n) < \infty$ for all n and $T = \bigcup\limits_{n=0}^{\infty} B_n$.

- *Let μ be a signed or complex measure. Define for $A \in \mathcal{A}$,*

$$|\mu|(A) := \sup \left\{ \sum_{k=1}^{n} |\mu(E_k)| \;\middle|\; n \in \mathbb{N},\ E_k \text{ pairwise disjoint},\ A = \bigcup_{k=1}^{n} E_k \right\}.$$

  The map $A \mapsto |\mu|(A)$ is called the variation of μ.

Proposition 2.23. *[24, Satz A.4.3] The variation $|\mu|$ of a signed or complex measure μ on the σ-algebra $\mathcal{A}$ is a positive, finite measure on $\mathcal{A}$.*

Definition and Proposition 2.24. *[24, Section I.1] Let T be a set and $\mathcal{A}$ a σ-algebra on T. We denote by $\mathcal{M}(T, \mathcal{A}, \mathbb{R})$ and $\mathcal{M}(T, \mathcal{A}, \mathbb{C})$ the set of all signed and complex measures, respectively, on $\mathcal{A}$. $\mathcal{M}(T, \mathcal{A}, \mathbb{R})$ and $\mathcal{M}(T, \mathcal{A}, \mathbb{C})$ are vector spaces and $\|\mu\|_{\mathcal{M}} := |\mu|(T)$ defines a norm on $\mathcal{M}(T, \mathcal{A}, \mathbb{R})$ and $\mathcal{M}(T, \mathcal{A}, \mathbb{C})$, respectively. To simplify notation, we will write $\mathcal{M}(T, \mathcal{A})$ for $\mathcal{M}(T, \mathcal{A}, \mathbb{R})$.*

Definition and Proposition 2.25. *[5, Remark 1.17 and Section 1.2] Let T be a topological space, $\mathcal{A}$ a σ-algebra on T and μ a positive measure on $\mathcal{A}$. Define for μ-measurable $f : T \to \mathbb{R}$ and $1 \leq p < \infty$*

$$\|f\|_{L^p(T,\mu)} := \left(\int_T |f|^p \, d\mu \right)^{\frac{1}{p}}$$

and

$$\|f\|_{L^\infty(T,\mu)} := \inf_{\substack{N \subset T \\ \mu(N)=0}} \sup_{x \in T \setminus N} |f(x)|.$$

Further, define for $1 \leq p \leq \infty$

$$\mathcal{L}^p(T, \mu) := \left\{ f : T \to \mathbb{R} \;\middle|\; f \ \mu\text{-measurable}, \|f\|_{L^p(T,\mu)} < \infty \right\}$$

and $L^p(T, \mu)$ as the space of all equivalence classes of functions in $\mathcal{L}^p(T, \mu)$, where one equivalence class consists of all functions in $\mathcal{L}^p(T, \mu)$, which are equal everywhere up to a set of measure zero. Then, for all $1 \leq p \leq \infty$, $(L^p(T, \mu), \|.\|_{L^p(T,\mu)})$ is a Banach space.

Definition 2.26. *[Absolute continuity and singularity, [5, Definition 1.24]]*

- *Let T be a topological space and $\mathcal{A}$ a σ-algebra on T. Let further μ be a positive measure and ν a signed measure on $\mathcal{A}$. We say, that ν is absolutely continuous with respect to μ and write $\nu \ll \mu$ if for every $B \in \mathcal{A}$*

$$\mu(B) = 0 \Rightarrow |\nu|(B) = 0.$$

- *If μ and ν are positive measures, we say, that they are mutually singular and write $\mu \perp \nu$, if there exists $E \in \mathcal{A}$, such that*

$$\mu(E) = 0 \text{ and } \nu(T \setminus E) = 0.$$

  If μ or ν are signed measures, we say, that they are mutually singular if $|\mu|$ and $|\nu|$ are.

Theorem 2.27. *[Radon-Nikodým, [5, Theorem 1.28]] Let T be a topological space and $\mathcal{A}$ a σ-algebra on T. Let further μ be a positive, σ-finite measure and ν a signed measure on $\mathcal{A}$. Then, there exists a unique pair of signed measures ν^a and ν^s, such that $\nu^a \ll \mu$, $\nu^s \perp \mu$ and $\nu = \nu^a + \nu^s$. Moreover, there is a unique function $f \in L^1(T, \mu)$, such that $\nu^a = f\mu$, i.e., for all $B \in \mathcal{A}$,*

$$\nu^a(B) = \int_B f(x)\, d\mu(x).$$

In this case, f is called the density or Radon-Nikodým derivative of ν with respect to μ and denoted $\frac{\nu}{\mu}$.

Definition 2.28. *[Borel measures, $\mathcal{M}(T)$] Let T be a topological space and $\mathcal{A}$ the Borel σ-algebra on T, i.e., the σ-algebra generated by the open sets in T. A (signed, complex or positive) measure μ on $\mathcal{A}$ is called (signed, complex or positive) Borel measure. A positive Borel measure is called regular if*

- *$\mu(K) < \infty$ for any compact $K \in \mathcal{A}$ and*
- *for every $A \in \mathcal{A}$*

$$\begin{aligned} \mu(A) &= \sup \{\mu(K) \mid K \subset A,\ K \text{ compact}\} = \\ &= \inf \{\mu(O) \mid A \subset O,\ O \text{ open}\}. \end{aligned}$$

A signed or complex Borel measure is called regular if its variation $|\mu|$ is regular. We denote the set of all signed and complex regular Borel measures on $\mathcal{A}$ as $\mathcal{M}(T, \mathbb{R})$ and $\mathcal{M}(T, \mathbb{C})$, respectively. For the sake of brevity, we will write $\mathcal{M}(T)$ for $\mathcal{M}(T, \mathbb{R})$.

Theorem 2.29. *[24, Satz I.2.14] If $T \subset \mathbb{R}^d$ is compact or open, any finite, positive Borel measure on T is regular.*

Theorem 2.30. *[Riesz' representation theorem, [24, Theorem II.2.5]], [5, Theorem 1.54 (Riesz)]]*

- *Let* $K \subset \mathbb{R}^d$ *be compact, then* $\mathcal{M}(K) \cong (C(K))^*$ *via the isometric isomorphism*

$$T : \mathcal{M}(K) \to (C(K))^*$$
$$(T\mu)(f) = \int\limits_K f \, d\mu, \quad \forall f \in C(K).$$

In particular, $\|\mu\|_{\mathcal{M}} = |\mu|(K) = \sup\limits_{\substack{f \in C(K) \\ \|f\|_\infty \le 1}} \left| \int\limits_K f \, d\mu \right|.$

- *Let* $\Omega \subset \mathbb{R}^d$ *be open, then* $\mathcal{M}(\Omega) \cong (C_0(\Omega))^*$ *via the isometric isomorphism*

$$T : \mathcal{M}(\Omega) \to (C_0(\Omega))^*$$
$$(T\mu)(f) = \int\limits_\Omega f \, d\mu, \quad \forall f \in C_0(\Omega).$$

In particular, $\|\mu\|_{\mathcal{M}} = |\mu|(\Omega) = \sup\limits_{\substack{f \in C_0(\Omega) \\ \|f\|_\infty \le 1}} \left| \int\limits_\Omega f \, d\mu \right|.$

Remark 2.31. *Based on Theorem 2.30, we will identify Borel measures on a compact set* K *and an open set* Ω *with elements of* $(C(K))^*$ *and* $(C_0(\Omega))^*$*, respectively. That is, we will write, e.g., for* $\mu \in \mathcal{M}(K)$ *and* $f \in C(K)$

$$\mu(f) = (T\mu)(f) = \int\limits_K f \, d\mu,$$

for the application of the measure μ *regarded as an element of* $(C(K))^*$*.*

Remark 2.32. *Since* $C(K)$ *is separable if* K *is compact [24, Section I.2], we can apply the theorem of Banach-Alaoglu, Theorem 2.16, and obtain, that every bounded sequence in* $\mathcal{M}(K) \cong (C(K))^*$ *admits a weak* convergent subsequence. Analogously, every bounded sequence in* $\mathcal{M}(\Omega) \cong (C_0(\Omega))^*$ *admits a weak* convergent subsequence [5, Theorem 1.59 (Weak* compactness].*

Remark 2.33. *Let* $\Omega \subset \mathbb{R}^d$ *open and bounded and* $u \in L^1(\Omega)$*. Then,*

$$C(\overline{\Omega}) \to \mathbb{R} \qquad\qquad C_0(\Omega) \to \mathbb{R}$$
$$f \mapsto \int\limits_\Omega u(x) f(x) \, dx \qquad \text{and} \qquad f \mapsto \int\limits_\Omega u(x) f(x) \, dx$$

define elements of $\mathcal{M}(\overline{\Omega})$ *and* $\mathcal{M}(\Omega)$*, respectively.*

2.1.3 Function Spaces and Related Results

Definition and Proposition 2.34. *[Sobolev spaces, [3, 1.25 Sobolev-Räume]] Let $\Omega \subset \mathbb{R}^d$ open. For $k \in \mathbb{N}$ and $p \in [1, \infty]$ we define the Sobolev space $W^{k,p}(\Omega)$ as*

$$W^{k,p}(\Omega) = \left\{ f \in L^p(\Omega) \;\middle|\; \forall \alpha \in \mathbb{N}^d, |\alpha| \leq k : \; \partial^\alpha f \in L^p(\Omega) \right\},$$

where $|\alpha| = \sum_{i=1}^d \alpha_i$, $\partial^\alpha f = \partial_{x_1}^{\alpha_1} ... \partial_{x_d}^{\alpha_d} f$ and the derivatives are understood in the weak sense. Equipped with the norm

$$\|f\|_{k,p} := \sum_{\substack{\alpha \in \mathbb{N}^d \\ |\alpha| \leq k}} \|\partial^\alpha f\|_p \,,$$

$W^{k,p}(\Omega)$ is a Banach space.

Theorem 2.35. *[Approximation of Sobolev functions, [3, 1.26 Satz]] For $k \in \mathbb{N}$ and $p \in [1, \infty)$, $C^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$, i.e., for every $f \in W^{k,p}(\Omega)$, there exists a sequence $(f_n)_n \subset C^\infty(\Omega) \cap W^{k,p}(\Omega)$, such that*

$$\|f - f_n\|_{k,p} \to 0$$

as $n \to \infty$.

Definition and Proposition 2.36. *[Total variation, functions of bounded variation, [5, Chapter 3]] For $\Omega \subset \mathbb{R}^d$ open and $u \in L^1_{loc}(\Omega)$ we define the total variation of u as*

$$\mathrm{TV}(u) := \sup \left\{ \int_\Omega u \; \mathrm{div}(\phi) \; dx \;\middle|\; \phi \in C_c^1(\Omega, \mathbb{R}^d), \|\phi\|_\infty \leq 1 \right\}.$$

Moreover, the space of functions of bounded variation is defined as

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

Equipped with the norm

$$\|u\|_{\mathrm{BV}} := \|u\|_1 + \mathrm{TV}(u)$$

$\mathrm{BV}(\Omega)$ is a Banach space.

Lemma 2.37. *[5, Proposition 3.6] Let $\Omega \subset \mathbb{R}^d$ open and $u \in L^1(\Omega)$. Then $u \in \mathrm{BV}(\Omega)$ if and only if the distributional derivative of u can be represented by a regular Borel measure, i.e., there exists $\mu = (\mu_i)_{i=1}^d \in [\mathcal{M}(\Omega)]^d$, such that for all $\phi \in C_c^\infty(\Omega)$*

$$\int\limits_\Omega u \frac{\partial}{\partial x_i} \phi \; dx = - \int\limits_\Omega \phi \; d\mu_i$$

for all $i = 1, 2, ...d$. In this case, we denote $\mu = Du$ for the distributional derivative of u. Moreover, for any $u \in \mathrm{BV}(\Omega)$

$$\mathrm{TV}(u) = |Du|(\Omega) = \|Du\|_{\mathcal{M}},$$

where for $\mu \in [\mathcal{M}(\Omega)]^d$ the variation and norm are defined as

$$|\mu|(A) := \sup\left\{ \sum_{k=1}^{n} |\mu(E_k)| \;\middle|\; n \in \mathbb{N},\ E_k \text{ pairwise disjoint},\ A = \bigcup_{k=1}^{n} E_k \right\}, \quad \text{for } A \text{ measurable and}$$

$$\|\mu\|_{\mathcal{M}} := \sup_{\substack{f \in [C_0(\Omega)]^d \\ |f| \le 1}} \left| \sum_{i=1}^{d} \int_{\Omega} f_i \, d\mu_i \right|,$$

respectively.

Definition 2.38. *[Approximate jump points, [5, Definition 3.67]] Let $u \in L^1(\Omega)$ and $x \in \Omega$. We say, that x is an approximate jump point of u if there exist $a, b \in \mathbb{R}$ and $\nu \in \mathbb{R}^d$ with $|\nu| = 1$, such that $a > b$ and*

$$\lim_{\rho \downarrow 0} \int_{B_\rho^+(x,\nu)} |u(y) - a| \, dy = 0,$$

$$\lim_{\rho \downarrow 0} \int_{B_\rho^-(x,\nu)} |u(y) - b| \, dy = 0,$$

where $B_\rho^\pm = \{y \in \mathbb{R} \mid |y - x| < \rho,\ \pm\langle y - x, \nu\rangle > 0\}$. The triplet (a, b, ν) is uniquely determined and denoted by $(u^+(x), u^-(x), \nu_u(x))$. We denote the set of all approximate jump points of u as J_u.

In the next statement, we use the notion of the restriction of a measure. For a measure μ on the σ-algebra $\mathcal{A}$ and a set $A \in \mathcal{A}$, we define the restriction of μ on A as $\mu|_A(B) := \mu(A \cap B)$ for $B \in \mathcal{A}$.

Corollary 2.39. *[5, Definition 3.91] Let $u \in \mathrm{BV}(\Omega)$. Then, according to Theorem 2.27, we can decompose the distributional derivative of u into an absolutely continuous part $D^a u$ and a singular part $D^s u$ with respect to the Lebesgue measure $\mathcal{L}^d$. Denoting $\nabla u := \frac{Du}{\mathcal{L}^d} \in L^1(\Omega)$ we obtain*

$$Du = \nabla u \mathcal{L}^d + D^s u.$$

Moreover, denoting the restrictions $D^j u = D^s u|_{J_u}$ (the so-called jump part of Du) and $D^c u = D^s u|_{(\Omega \setminus J_u)}$ (the so-called Cantor part of Du) we may write

$$Du = \nabla u \mathcal{L}^d + D^j u + D^c u.$$

Proposition 2.40. *[Approximation of BV functions, [5, Theorem 3.9]] Let $\Omega \subset \mathbb{R}^d$ open and $u \in L^1(\Omega)$. Then $u \in \mathrm{BV}(\Omega)$ if and only if there exists a sequence $(u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$ converging to u in $L^1(\Omega)$, such that*

$$\lim_{n\to\infty} \int_\Omega |\nabla u_n| \, dx < \infty.$$

Moreover,

$$\mathrm{TV}(u) = \inf \left\{ \lim_{n\to\infty} \int_\Omega |\nabla u_n| \, dx \;\middle|\; (u_n)_n \subset C^\infty(\Omega), u_n \to u \text{ in } L^1(\Omega), \lim_{n\to\infty} \int_\Omega |\nabla u_n| \, dx \text{ exists} \right\}$$

Let us introduce a convenient notation before stating the next result. For $1 \le p \le d$, we denote $p^* = \frac{dp}{d-p}$, where we use the convention $\frac{q}{0} = \infty$ for $q > 0$.

Lemma 2.41. *[Embedding theorem, [5, Remark 3.49]] Let $\Omega \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain with compact boundary. Then $\mathrm{BV}(\Omega)$ is continuously embedded in $L^{1^*}(\Omega)$.*

Lemma 2.42. *[Poincaré inequality, [5, Remark 3.50]] Let $\Omega \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain with compact boundary. Then there exists a constant $C > 0$, such that for all p with $1 \le p \le 1^*$*

$$\|u - u_\Omega\|_p \le C\,\mathrm{TV}(u), \quad \forall u \in \mathrm{BV}(\Omega),$$

where $u_\Omega = \frac{1}{|\Omega|} \int_\Omega u \, dx$.

2.1.4 Lower Semi-Continuity

Definition 2.43. *[Lower semi-continuity] Let X be a normed space and $F : X \to \overline{\mathbb{R}}$. We say, that F is lower semi-continuous (lsc), if for every $x \in X$ and every sequence $(x_n)_n \subset X$ converging to x, it holds true, that*

$$F(x) \le \liminf_{n\to\infty} F(x_n). \tag{2.1}$$

Similarly, we call F weakly/weak lower semi-continuous if (2.1) holds true, for every weakly/weak* convergent sequence.*

Lemma 2.44. *Let X be a normed space and $F, G : X \to \overline{\mathbb{R}}$ be lsc. Then also the sum $F + G$ is lsc.*

Proof. Let $x_n \to x$ in X. We compute

$$F(x) + G(x) \underbrace{\le}_{F,G \text{ lsc}} \liminf_{n\to\infty} F(x_n) + \liminf_{n\to\infty} G(x_n) = \lim_{n\to\infty} \inf_{k\ge n} F(x_k) + \lim_{n\to\infty} \inf_{k \ge n} G(x_k) =$$

$$\lim_{n\to\infty} \left(\inf_{k\ge n} F(x_k) + \inf_{k\ge n} G(x_k) \right) \le \lim_{n\to\infty} \inf_{k\ge n} (F(x_k) + G(x_k)) = \liminf_{n\to\infty} (F(x_n) + G(x_n)).$$

□

Lemma 2.45. *[Weak* lower semi-continuity of the dual norm] Let X be a normed space. Then the dual norm on X^* is lsc with respect to weak*-convergence.*

Proof. Let $x_n^* \overset{*}{\rightharpoonup} x^*$. We compute,

$$\|x^*\| = \sup_{\|x\|\leq 1} |x^*(x)| = \sup_{\|x\|\leq 1} \lim_{n\to\infty} \underbrace{|x_n^*(x)|}_{\leq \|x_n^*\| \underbrace{\|x\|}_{\leq 1}} \leq \sup_{\|x\|\leq 1} \liminf_{n\to\infty} \|x_n^*\| = \liminf_{n\to\infty} \|x_n^*\| .$$

□

Lemma 2.46. *[[13, Corollary 2.2]] Let X be a normed space and $F : X \to \overline{\mathbb{R}}$ be convex. Then F is lsc, if and only F is weakly lsc.*

Lemma 2.47. *Let X be a normed space, I an index set and $F_i : X \to \overline{\mathbb{R}}$ for all $i \in I$. Define $F : X \to \overline{\mathbb{R}}$, $F(x) = \sup_{i\in I} F_i(x)$. Then,*

1. *if for all $i \in I$, F_i is convex, then also F is convex and*
2. *if for all $i \in I$, F_i is lsc, then also F is lsc.*

Proof. 1. Assume F_i is convex for all $i \in I$. Let $x, y \in X$ and $\lambda \in (0, 1)$. We compute,

$$\begin{aligned} F(\lambda x + (1-\lambda)y) &= \sup_{i\in I} [F_i(\lambda x + (1-\lambda)y)] \leq \\ &\underbrace{\leq}_{F_i \text{ convex}} \sup_{i\in I} [\lambda F_i(x) + (1-\lambda)F_i(y)] \leq \\ &\underbrace{\leq}_{F_i \leq F} \sup_{i\in I} [\lambda F(x) + (1-\lambda)F(y)] = \\ &= \lambda F(x) + (1-\lambda)F(y). \end{aligned}$$

2. Assume F_i is lsc for all $i \in I$. Let $x_n \to x$ in X.

$$\begin{aligned} F(x) &= \sup_{i\in I} F_i(x) \underbrace{\leq}_{F_i \text{ lsc}} \sup_{i\in I} \liminf_{n\to\infty} F_i(x_n) \leq \\ &\underbrace{\leq}_{F_i\leq F} \sup_{i\in I} \liminf_{n\to\infty} F(x_n) = \liminf_{n\to\infty} F(x_n). \end{aligned}$$

□

Definition 2.48. *[lsc-regularization, [13, Section 2.2]] Let X be a normed vector space and $F : X \to \overline{\mathbb{R}}$. We define the lsc regularization of F as the largest lsc function everywhere less than F and denote it as $\overline{F}$. It exists as the pointwise supremum of all lsc functions everywhere less than F.*

Lemma 2.49. *[13, Corollary 2.1] Let X be a normed vector space and $F : X \to \overline{\mathbb{R}}$. Then*

$$\operatorname{epi}(\overline{F}) = \overline{\operatorname{epi}(F)},$$

where for $G : X \to \overline{\mathbb{R}}$, the epigraph of G is defined as $\operatorname{epi}(G) := \{(x,t) \in X \times \mathbb{R} \mid G(x) \leq t\}$.

Lemma 2.50. *Let X be a normed vector space and $F : X \to \overline{\mathbb{R}}$. Then*

$$\overline{F}(x) = \inf\left\{\liminf_{n\to\infty} F(x_n) \;\middle|\; x_n \to x \text{ as } n \to \infty\right\}.$$

Proof. By definition of $\overline{F}$, $\overline{F} \leq F$ and $\overline{F}$ is lsc. Therefore, for any sequence $(x_n)_n$ converging to x, we have

$$\overline{F}(x) \leq \liminf_{n\to\infty} \overline{F}(x_n) \leq \liminf_{n\to\infty} F(x_n).$$

Therefore,

$$\overline{F}(x) \leq \inf\left\{\liminf_{n\to\infty} F(x_n) \mid x_n \to x \text{ as } n \to \infty\right\}.$$

Since $(x, \overline{F}(x)) \in \operatorname{epi}(\overline{F})$, by Lemma 2.49, there exists a sequence $(x_n, t_n)_n \subset \operatorname{epi}(F)$ converging to $(x, \overline{F}(x))$, i.e., $x_n \to x$, $F(x_n) \leq t_n$ and $t_n \to \overline{F}(x)$. Hence,

$$\liminf_{n\to\infty} F(x_n) \leq \liminf_{n\to\infty} t_n = \overline{F}(x)$$

and consequently also

$$\overline{F}(x) \geq \inf\left\{\liminf_{n\to\infty} F(x_n) \mid x_n \to x \text{ as } n \to \infty\right\}$$

concluding the proof. □

Lemma 2.51. *Let $F : X \to \overline{\mathbb{R}}$. Then for every $x \in X$ there exists a sequence $(x_n)_n$ converging to x, such that $\overline{F}(x) = \lim_{n\to\infty} F(x_n)$.*

Proof. We will distinguish three different cases.

- Assume, that $\overline{F}(x) = \infty$. This means, that for every sequence $(x_n)_n$, converging to x, we find, that $\liminf_{n\to\infty} F(x_n) = \infty$. Therefore, by just picking any such sequence $(x_n)_n$, we obtain the desired result.
- Assume, that $\overline{F}(x) = -\infty$. Then, for every $n \in \mathbb{N}$, we can find a sequence $(x_k^n)_k$, with $x_k^n \to x$ as $k \to \infty$, such that $\liminf_{k\to\infty} F(x_k^n) \leq -2n$. By taking subsequences, which we shall not relabel now, we can also obtain, that $\lim_{k\to\infty} F(x_k^n) = \liminf_{k\to\infty} F(x_k^n) \leq -2n$. Then, for every $n \in \mathbb{N}$, there exists $k_n^0 \in \mathbb{N}$, such that for $k \geq k_n^0$, $F(x_k^n) \leq -n$. Now, pick $k_n \geq k_n^0$, such that $\left\|x_{k_n}^n - x\right\| \leq \frac{1}{n}$. Then, the sequence $x_n := x_{k_n}^n$ satisfies

$$\|x_n - x\| = \|x_{k_n}^n - x\| \leq \frac{1}{n} \xrightarrow{n\to\infty} 0 \text{ and}$$
$$F(x_n) \leq -n \xrightarrow{n\to\infty} -\infty = \overline{F}(x).$$

- Assume, that $\overline{F}(x) \in (-\infty, \infty)$. As before, by definition of $\overline{F}$, we can find sequences $(x_k^n)_k$ with $x_k^n \xrightarrow{k\to\infty} x$ for all $n \in \mathbb{N}$, such that $\lim\limits_{k\to\infty} F(x_k^n) \le \overline{F}(x) + \frac{1}{n}$. Therefore, for every $n \in \mathbb{N}$, we can find $k_n^0 \in \mathbb{N}$, such that for $k \ge k_n^0$, $F(x_k^n) \le \overline{F}(x) + \frac{2}{n}$. Finally, let $k_n \ge k_n^0$, such that $\left\| x_{k_n}^n - x \right\| \le \frac{1}{n}$. Then for the sequence $x_n := x_{k_n}^n$, we find, that

$$\begin{aligned}
\|x_n - x\| = \|x_{k_n}^n - x\| \le \frac{1}{n} \xrightarrow{n\to\infty} 0 \text{ and} \\
\overline{F}(x) \underbrace{\le}_{\textit{Lemma } 2.50} \liminf_{n\to\infty} F(x_n) \le \limsup_{n\to\infty} F(x_n) = \\
= \limsup_{n\to\infty} F(x_{k_n}^n) \le \limsup_{n\to\infty} \left(\overline{F}(x) + \frac{2}{n} \right) \le \overline{F}(x),
\end{aligned} \tag{2.2}$$

yielding, that $\overline{F}(x) = \lim\limits_{n\to\infty} F(x_n)$.

□

Definition 2.52. *[Γ-regularization, [13, Sections 3.1, 3.2]] Let X be a normed vector space.*

- *We define $\Gamma(X)$ as the set of all functions $X \to \overline{\mathbb{R}}$, which are the pointwise supremum of a family of continuous, affine functions.*
- *Let $F : X \to \overline{\mathbb{R}}$, then the Γ-regularization of F is the pointwise supremum of all continuous, affine functions everywhere less than F.*

Note, that by Lemma 2.47, any function in $\Gamma(X)$, in particular the Γ-regularization, are convex and lsc. In order to get a better understanding of its utility, in the following, we will list some properties of the Γ-regularization.

Lemma 2.53. *[13, Definition 3.2] Let F and G be two functions mapping $X \to \overline{\mathbb{R}}$. The following are equivalent to each other:*

- *G is the largest function in $\Gamma(X)$, which is everywhere less than F.*
- *G is the Γ-regularization of F.*

Proposition 2.54. *[[13, Proposition 3.3]] Let $F : X \to \overline{\mathbb{R}}$ and G its Γ-regularization. If F is convex and admits a continuous affine minorant, then $\overline{F} = G$.*

Definition 2.55. *[Polar (or convex conjugate) function] Let X be a normed vector space and $F : X \to \overline{\mathbb{R}}$. We define the polar (or convex conjugate) function of F as*

$$\begin{aligned}
F^* : X^* &\to \overline{\mathbb{R}} \\
x^* &\mapsto \sup_{x\in X} \left[x^*(x) - F(x) \right]
\end{aligned}$$

Moreover, we define the bipolar of F as

$$\begin{aligned} F^{**} &: X \to \overline{\mathbb{R}} \\ x &\mapsto \sup_{x^* \in X^*} [x^*(x) - F^*(x^*)] \end{aligned}$$

Remark 2.56.

- *As the supremum of linear and continuous functions, by Lemma 2.47, F^{**} is convex and lsc.*
- *In the case that X is reflexive, $F^{**} = (F^*)^*$, which justifies the name bidual.*

Proposition 2.57. *[13, Proposition 4.1] Let $F : X \to \overline{\mathbb{R}}$. Then F^{**} is the Γ-regularization of F.*

2.2 The Problem in a Continuous Setting

In this section, we will formulate the proposed method in a continuous setting. Afterwards, we will analyze the method, the two main results being existence of a solution and stability of the method with respect to the data. In order to do so, we will need some preliminary definitions and results.

2.2.1 Preliminaries

Before introducing the model, we will state some definitions and results. In the following, $\Omega, \Sigma \subset \mathbb{R}^d$ are bounded domains, i.e., bounded, open and simply connected. Moreover, we denote $\Omega_\Sigma = \{x - y \mid x \in \Omega,\ y \in \Sigma\}$. For $q \in [1, \infty]$ used as a Hölder exponent (as in $L^q(\Omega)$) we will usually denote the conjugate exponent as q', i.e., $q' \in [1, \infty]$ is, such that $\frac{1}{q} + \frac{1}{q'} = 1$, using the convention, that $\frac{1}{\infty} = 0$.

Lemma 2.58. *Let $\mu \in \mathcal{M}(\Omega_\Sigma)$ and $\theta \in \mathcal{M}(\overline{\Sigma})$. Define*

$$\begin{aligned} \mu * \theta &: C_0(\Omega) \to \mathbb{R}, \\ f &\mapsto \int_{\Omega_\Sigma} \int_{\overline{\Sigma}} \tilde{f}(x + y)\, d\theta(x)\, d\mu(y) \end{aligned} \tag{2.3}$$

where we remember, that $\tilde{f}$ denotes the zero-extension of f outside of Ω. Then (2.3) is a well-defined, linear and bounded operator with respect to $\|\,.\,\|_\infty$, i.e., $\mu * \theta \in \mathcal{M}(\Omega) = (C_0(\Omega))^*$, and $\|\mu * \theta\|_{\mathcal{M}} \leq \|\mu\|_{\mathcal{M}} \|\theta\|_{\mathcal{M}}$.

Proof. Let $f \in C_0(\Omega)$. Then $\tilde{f} \in C(\mathbb{R}^d)$ and it is uniformly continuous as it has compact support in $\mathbb{R}^d$. For any $y \in \Omega_\Sigma$, $\tilde{f}(. + y) \in C(\overline{\Sigma})$. Hence,

$$\int_{\overline{\Sigma}} \tilde{f}(x+y)\, d\theta(x) = \theta(\tilde{f}(.+y))$$

is well-defined for any $y \in \Omega_\Sigma$. In order to to apply μ afterwards, we have to show, that the function

$$\begin{aligned} F : \Omega_\Sigma &\to \mathbb{R} \\ y &\mapsto \int_{\overline{\Sigma}} \tilde{f}(x+y)\, d\theta(x) \end{aligned}$$

is en element of $C_0(\Omega_\Sigma)$. Since $f \in C_0(\Omega)$, by definition, we can find a sequence $(f_n)_n \subset C_c(\Omega)$ converging to f with respect to $\| \,.\, \|_\infty$. We find, that for every $n \in \mathbb{N}$

$$\begin{aligned} F_n : \Omega_\Sigma &\to \mathbb{R} \\ y &\mapsto \int_{\overline{\Sigma}} \tilde{f}_n(x+y)\, d\theta(x) \end{aligned}$$

is an element of $C_c(\Omega_\Sigma)$. Indeed, to prove the continuity, let $\delta > 0$ and $y_1, y_2 \in \Omega_\Sigma$, such that $|y_1 - y_2| < \delta$. Then

$$\begin{aligned} |F_n(y_1) - F_n(y_2)| &= \left| \int_{\overline{\Sigma}} \tilde{f}_n(x+y_1) - \tilde{f}_n(x+y_2)\, d\theta(x) \right| \leq \\ &\leq \|\theta\|_{\mathcal{M}} \sup_{x \in \Sigma} |\tilde{f}_n(x+y_1) - \tilde{f}_n(x+y_2)| \leq \\ &\leq \|\theta\|_{\mathcal{M}} \sup_{z_1, z_2 :\, |z_1 - z_2| < \delta} |\tilde{f}_n(z_1) - \tilde{f}_n(z_2)| \to 0 \quad \text{as } \delta \to 0. \end{aligned}$$

The convergence to zero follows from the uniform continuity of $\tilde{f}_n$. Therefore F_n is continuous for any n. Now denote $K = \operatorname{supp}(f_n) \Subset \Omega$. Then $\tilde{f}_n(x+y) = 0$ for all $x \in \overline{\Sigma}$ if $y \notin K - \overline{\Sigma}$, i.e., $\int_{\overline{\Sigma}} \tilde{f}_n(x+y)\, d\theta(x) = 0$ if $y \notin K - \overline{\Sigma}$. $K - \overline{\Sigma}$ is compact since K and $\overline{\Sigma}$ are. What is left to show, is, that $K - \overline{\Sigma} \subset \Omega - \Sigma$ in order to obtain, that F_n is compactly supported in Ω_Σ. So let $y = z - x \in K - \overline{\Sigma}$ with $z \in K$ and $x \in \overline{\Sigma}$. Since $K \Subset \Omega$, there exists $\delta > 0$, such that the open ball with radius δ and center z is contained in the open set Ω. Moreover, we can find $\tilde{x} \in \Sigma$, such that $|x - \tilde{x}| < \delta$. Therefore, $z + (\tilde{x} - x) \in \Omega$. We find

$$y = z - x = z + (\tilde{x} - x) - \tilde{x} \in \Omega - \Sigma,$$

which shows, that $K - \overline{\Sigma} \subset \Omega - \Sigma$ and therefore F_n is compactly supported in Ω_Σ and altogether, $F_n \in C_c(\Omega_\Sigma)$ for all n. Moreover, for $y \in \Omega_\Sigma$

$$\begin{aligned} |F(y) - F_n(y)| &= \left| \int\limits_{\overline{\Sigma}} \tilde{f}(x+y) - \tilde{f}_n(x+y)\; d\theta(x) \right| \leq \\ &\leq \|\theta\|_{\mathcal{M}} \sup_{x\in\Sigma} |\tilde{f}(x+y) - \tilde{f}_n(x+y)| \leq \\ &\leq \|\theta\|_{\mathcal{M}} \sup_{z\in\Omega} |\tilde{f}(z) - \tilde{f}_n(z)| = \\ &= \|\theta\|_{\mathcal{M}} \sup_{z\in\Omega} |f(z) - f_n(z)| \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

The right-hand side tends to zero independently of y, since the f_n approximate f uniformly. Therefore F is the uniform limit of the sequence $(F_n)_n \subset C_c(\Omega_\Sigma)$ and accordingly an element of $C_0(\Omega_\Sigma)$. As a result, we may apply μ and find that $\mu * \theta$ is well-defined. The linearity of $\mu * \theta$ follows from the linearities of μ and θ. Moreover, we can bound the norm of $\mu * \theta$ as follows,

$$\begin{aligned} |\mu * \theta(f)| &\leq \|\mu\|_{\mathcal{M}} \sup_{y\in\Omega_\Sigma} |\theta(\tilde{f}(.+y))| \leq \\ \leq \|\mu\|_{\mathcal{M}} \sup_{y\in\Omega_\Sigma} &\left(\|\theta\|_{\mathcal{M}} \sup_{x\in\overline{\Sigma}} |(\tilde{f}(x+y))| \right) \leq \|\mu\|_{\mathcal{M}} \|\theta\|_{\mathcal{M}} \|f\|_\infty . \end{aligned}$$

Therefore, $\|\mu * \theta\|_{\mathcal{M}} \leq \|\mu\|_{\mathcal{M}} \|\theta\|_{\mathcal{M}} < \infty$ and $\mu * \theta \in \mathcal{M}(\Omega)$. □

Lemma 2.59. *Let $\mu \in \mathcal{M}(\Omega_\Sigma)$, $g \in L^s(\Sigma)$, $s \in (1,\infty]$, $q \in (1,s]$ and q' the Hölder conjugate exponent of q. Then*

$$\begin{aligned} \mu * g : C_c(\Omega) &\to \mathbb{R}, \\ f &\mapsto \int\limits_{\Omega_\Sigma} \int\limits_{\Sigma} \tilde{f}(x+y) g(x)\; dx\; d\mu(y) \end{aligned}$$

can be extended to a well-defined, linear and bounded operator on $L^{q'}(\Omega)$, i.e., $\mu * \theta \in (L^{q'}(\Omega))^* \simeq L^q(\Omega)$. Moreover, there exists $C = C(\Sigma, q, s)$, such that $\|\mu * g\|_q \leq C \|\mu\|_{\mathcal{M}} \|g\|_s$.

Proof. Let $f \in C_c(\Omega)$. We need to show, that $(y \mapsto \int_\Sigma \tilde{f}(x+y) g(x)\; dx) \in C_0(\Omega_\Sigma)$, before we can apply μ. Let $\delta > 0$, then we find for $y_1, y_2 \in \Omega$, such that $|y_1 - y_2| < \delta$

$$\begin{aligned} \left| \int_\Sigma \tilde{f}(x+y_1) g(x)\; dx - \int_\Sigma \tilde{f}(x+y_2) g(x)\; dx \right| &= \left| \int_\Sigma \left(\tilde{f}(x+y_1) - \tilde{f}(x+y_2) \right) g(x)\; dx \right| \leq \\ \leq \int_\Sigma \left| \tilde{f}(x+y_1) - \tilde{f}(x+y_2) \right| |g(x)|\; dx &\leq \sup_{x\in\Sigma} |\tilde{f}(x+y_1) - \tilde{f}(x+y_2)| \, \|g\|_1 \leq \\ \leq |\Sigma|^{\frac{1}{s'}} \|g\|_s \sup_{z_1,z_2:|z_1-z_2|<\delta} &|\tilde{f}(z_2) - \tilde{f}(z_1)| \to 0 \quad \text{as } \delta \to 0, \end{aligned}$$

where we used the fact that with s' the Hölder conjugate exponent of s, we have, using Hölder's inequality, $\|g\|_1 = \int_\Sigma 1 \cdot |g|\ dx \leq \|g\|_s \|1\|_{s'} = \|g\|_s |\Sigma|^{\frac{1}{s'}}$. The uniform continuity of the compactly supported, continuous function f then implies convergence to zero as $\delta \to 0$. Therefore, $(y \mapsto \int_\Sigma \tilde{f}(x+y)g(x)\ dx)$ is uniformly continuous. The compact support of f in Ω again implies that

$$y \mapsto \int_\Sigma \tilde{f}(x+y)g(x)\ dx$$

is compactly supported in Ω_Σ as shown in Lemma 2.58. Hence, we may apply μ, proving, that $\mu * g$ is well-defined. Linearity follows from the linearities of the integral and μ. Moreover, for $s < \infty$, since $\frac{s}{q} \geq 1$,

$$\int_\Sigma |\tilde{f}(x+y)g(x)|\ dx \underbrace{\leq}_{\text{Hölder}} \left(\int_\Sigma |\tilde{f}(x+y)|^{q'}\ dx\right)^{\frac{1}{q'}} \left(\int_\Sigma |g(x)|^q\ dx\right)^{\frac{1}{q}} \leq$$

$$\underbrace{\leq}_{\text{Hölder}} \|f\|_{q'} \left(\int_\Sigma |g(x)|^{q\frac{s}{q}}\ dx\right)^{\frac{1}{q}\frac{q}{s}} |\Sigma|^{\frac{s-q}{sq}} = |\Sigma|^{\frac{s-q}{sq}} \|f\|_{q'} \|g\|_s .$$

In the case $s = \infty$, we find

$$\int_\Sigma |\tilde{f}(x+y)g(x)|\ dx \underbrace{\leq}_{\text{Hölder}} \left(\int_\Sigma |\tilde{f}(x+y)|^{q'}\ dx\right)^{\frac{1}{q'}} \left(\int_\Sigma |g(x)|^q\ dx\right)^{\frac{1}{q}} \leq$$

$$\leq \|f\|_{q'} \left(\int_\Sigma \|g\|_\infty^q\ dx\right)^{\frac{1}{q}} = \|f\|_{q'} \|g\|_\infty |\Sigma|^{\frac{1}{q}}.$$

Therefore, in any case, there exists $C > 0$, such that

$$\int_\Sigma |\tilde{f}(x+y)g(x)|\ dx \leq C \|f\|_{q'} \|g\|_s .$$

Hence, we compute for the norm of $\mu * g$,

$$|\mu * g(f)| \leq \|\mu\|_{\mathcal{M}} \sup_{y \in \Omega_\Sigma} |\int_\Sigma \tilde{f}(x+y)g(x)\ dx| \leq C \|\mu\|_{\mathcal{M}} \|f\|_{q'} \|g\|_s .$$

This shows, that $\mu * g : C_c(\Omega) \to \mathbb{R}$ is a linear operator, which is bounded with respect to $\|\ .\ \|_{q'}$. Since $C_c(\Omega)$ is dense in $L^{q'}(\Omega)$ for $q' \in [1, \infty)$, by Lemma 2.1, we can extend $\mu * g$ to a linear, bounded operator in $L^{q'}(\Omega)$. Via the isomorphism $(L^{q'}(\Omega))^* \cong L^q(\Omega)$ for $1 \leq q' < \infty$ we get the desired result. □

This enables us to define the convolution of measures.

Definition 2.60. *[Convolution of measures]*

- *Let $\mu \in \mathcal{M}(\Omega_\Sigma)$ and $\theta \in \mathcal{M}(\overline{\Sigma})$. We define the convolution of μ and θ, $\mu * \theta \in \mathcal{M}(\Omega)$, via*
$$\mu * \theta(f) = \int\limits_{\Omega_\Sigma} \int\limits_{\Sigma} \tilde{f}(x+y)\, d\theta(x)\, d\mu(y), \quad \text{for } f \in C_0(\Omega).$$
- *Let $\mu \in \mathcal{M}(\Omega_\Sigma)$ and $g \in L^s(\Sigma)$ for $s \in (1,\infty]$. Then for $q \in (1,s]$, we define the convolution of μ and g, $\mu * g \in L^q(\Omega)$, as the unique function in $L^q(\Omega)$ satisfying*
$$\int\limits_{\Omega_\Sigma} \int\limits_{\Sigma} \tilde{f}(x+y) g(x)\, dx\, d\mu(y) = \int_\Omega f(z)(\mu * g)(z) dz \quad \forall f \in C_c(\Omega).$$
We will denote
$$\mu * g(f) \coloneqq \int_\Omega f(z)(\mu * g)(z) dz$$
for $f \in L^{q'}(\Omega)$.

Remark 2.61.

- *Assume, that μ and θ are induced by functions, as mentioned in Remark 2.33, i.e.,*
$$\begin{aligned} \mu : C_0(\Omega_\Sigma) &\to \mathbb{R} \\ \mu(f) &= \int\limits_{\Omega_\Sigma} f(x) g_\mu(x)\, dx, \\ \theta : C(\overline{\Sigma}) &\to \mathbb{R} \\ \theta(f) &= \int\limits_{\Sigma} f(x) g_\theta(x)\, dx, \end{aligned}$$
with $g_\mu \in L^1(\Omega)$ and $g_\theta \in L^1(\Sigma)$. Then the convolution of μ and θ reduces to the well-known convolution of two functions. Using Fubinis theorem and a change of variables for $f \in C_0(\Omega)$ we can compute
$$\mu * \theta(f) = \int\limits_{\Omega_\Sigma} \int\limits_{\Sigma} \tilde{f}(x+y)\, d\theta(x)\, d\mu(y) =$$
$$= \int\limits_{\Omega_\Sigma} \int\limits_{\Sigma} \tilde{f}(x+y) g_\theta(x)\, g_\mu(y)\, dx\, dy = \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \tilde{f}(x+y) \tilde{g_\theta}(x)\, \tilde{g_\mu}(y)\, dy\, dx =$$
$$= \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \tilde{f}(z) \tilde{g_\mu}(z-x)\, \tilde{g_\theta}(x)\, dz\, dx = \int\limits_{\mathbb{R}^d} \tilde{f}(z) \int\limits_{\mathbb{R}^d} \tilde{g_\mu}(z-x)\, \tilde{g_\theta}(x)\, dx\, dz =$$
$$= \int\limits_{\Omega} f(z) \int\limits_{\Sigma} g_\mu(z-x)\, g_\theta(x)\, dx\, dz = \int\limits_{\Omega} f(z)(g_\mu * g_\theta)(z)\, dz,$$
where $\tilde{g}$ denotes the zero extension of a function g outside of its domain of definition.

- *Assume, that* $g_\mu \in L^{s'}(\Omega)$ *and* $g_\theta \in L^s(\Sigma)$ *with* $s > 1$*. Then*

$$|(g_\mu * g_\theta)(z+h) - (g_\mu * g_\theta)(z)| \leq \left| \int_\Sigma \tilde{g_\mu}(z+h-x)\, g_\theta(x)\, dx - \int_\Sigma \tilde{g_\mu}(z-x)\, g_\theta(x)\, dx \right| \leq$$

$$\leq \int_\Sigma |\tilde{g_\mu}(z+h-x) - \tilde{g_\mu}(z-x)|\; |g_\theta(x)|\; dx \underbrace{\leq}_{Hölder} \|\tilde{g_\mu}(.+h) - \tilde{g_\mu}\|_{s'} \|g_\theta\|_s \to 0$$

as $|h| \to 0$ *as proven in Lemma A.1. So we find, that the convolution of* $g_\mu * g_\theta$ *is actually continuous, in other words, under some circumstances, the convolution increases the regularity of functions.*

2.2.2 Problem Formulation

We now propose a variational regularization method, whose goal is to narrow the gap between conventional, variational regularization methods and machine/deep learning methods in image processing. Variational methods, such as total variation regularization, typically are very well suited for reconstructing piecewise smooth images and, just as importantly, they are mathematically well-understood. Machine/deep learning methods, on the other hand, perform very well also on images containing texture, but they lack an underlying mathematical foundation in many cases. In order to combine the strengths of both types of methods, we aim to decompose an image into two parts, one being a piecewise smooth cartoon part and the other containing the texture of the given image. Each of the to parts is treated with an appropriate regularizing term according to their nature. This results in the following formulation.

$$\min_{u,v} \mathcal{D}(u, y_0) + \lambda_{\mathcal{R}} \mathcal{R}(u-v) + \lambda_{\mathcal{G}} \mathcal{G}(v), \tag{P}$$

where $\mathcal{D}$ is a data fidelity term with data $y_0 \in Y$, Y being a Banach space, $\mathcal{R}$ is a cartoon prior enforcing piecewise smoothness, $\mathcal{G}$ is a texture prior, motivated by generative neural networks as they are found in recent deep learning methods, $\lambda_{\mathcal{R}}, \lambda_{\mathcal{G}} > 0$ are parameters to balance the different regularizers, u is the image, v is the texture part of u and $u - v$ the remaining cartoon part of u. The image u and the texture part v are assumed to be elements of $L^q(\Omega)$, where for now $1 \leq q \leq \infty$ and Ω is the image domain, which is open, bounded and simply connected. The next sections are dedicated to the introduction and mathematical analysis of the the functionals $\mathcal{R}$ and $\mathcal{G}$.

2.2.3 The Texture Prior $\mathcal{G}$

In this section we motivate, introduce and analyze the texture prior $\mathcal{G}$ from (P). As already mentioned in the introduction, there is a lot of research empirically showing the ability of convolutional neural networks to generate and/or regularize texture heavy images.

Therefore, we model the texture part $v \in L^q(\Omega)$ of (P) to be the output of such a network. We define the convolutional network $\mathcal{N}(\mu, \theta)$ with parameters (μ, θ) as

$$\mathcal{N}(\mu, \theta) := \sum_{k=1}^{n_{f,1}} \mu_k^1 * \theta_k^1, \tag{2.4}$$

subject to:

$$\mu_l^{j-1} = \sum_{k=1}^{n_{f,j}} \mu_k^j * \theta_{k,l}^j, \quad \text{for } j = 2, 3, ..., n, \ l = 1, 2, ..., n_{f,j-1}.$$

We denote

$$\mu := (\mu_k^n)_{k=1,2,...,n_{f,n}} \quad \text{and}$$
$$\theta := ((\theta_k^1)_k, (\theta_{k,l}^j)_{j,k,l}).$$

We will refer to $\mu^j = (\mu_k^j)_k$ as the coefficients at layer j and $\theta^j = (\theta_{k,l}^j)_{k,l}$ for $j > 1$, $\theta^1 = (\theta_k^1)_k$ as the filter kernels at layer j and layer 1, respectively. The coefficients μ in the deepest layer and the filter kernels are parameters of this neural network. In order to get a better understanding, we refer to Figure 2.1. This is a simplified sketch of (2.4) with $n = 2$. The green nodes illustrate the coefficients $\mu = \mu^2$, the blue nodes μ^1 and the yellow node is the texture part v, which is the output of the network. The edges connecting the nodes are to be understood as the convolution with the according filter kernel (each edge corresponds to one of the kernels $\theta_{k,l}^j$). Adding the corresponding convolutions gives the coefficients on the next layer and the output of the network, respectively. The coefficients and the filter

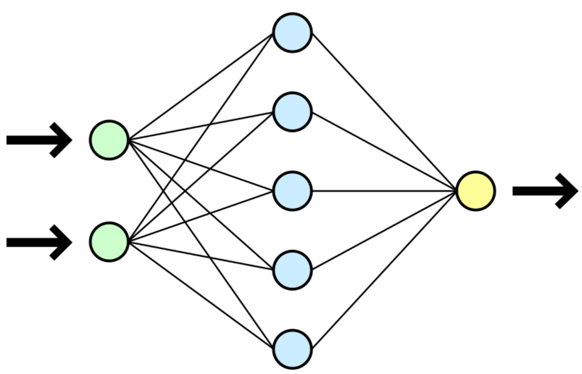


Figure 2.1: Sketch of a neural network [11]

kernels up to the first layer in (2.4) are modeled as Borel measures, since we also want to be able to capture, e.g., delta peaks, which are not in any L^p space. The texture prior $\mathcal{G}$ of our model enforces $v = \mathcal{N}(\mu, \theta)$ for some parameters (μ, θ). As a regularization, we penalize the norms of the coefficients μ as well as the filter kernels θ. Moreover, we impose additional constraints on the filter kernels. That is, in the deeper layers the kernels shall be non-negative and in the top layer they shall have zero mean. Our complete texture prior is defined as $\mathcal{G} : L^q(\Omega) \to [0, \infty]$,

$$\mathcal{G}(v) = \inf_{\mu,\theta} \left[\lambda_\mu \left\| \mu \right\|_{\mathcal{M}} + \lambda_\theta \sum_{i=2}^{n} \left(\left\| \theta^i \right\|_{\mathcal{M}} + \mathcal{I}_{\{. \geq 0\}}(\theta^i) \right) + \mathcal{I}_A(\theta^1) \right], \tag{TEX}$$

subject to:

$$\begin{cases} v & = \mathcal{N}(\mu, \theta), \\ \mu & \in [\mathcal{M}(\Omega_\mu)]^{n_{f,n}}, \\ \theta^1 & \in [L^s(\Sigma)]^{n_{f,1}}, \\ \theta^i & \in [\mathcal{M}(\overline{\Sigma})]^{n_{f,i} \times n_{f,i-1}}, \quad \text{for } i = 2, 3, ..., n, \end{cases}$$

where $s \in (1, \infty)$, $q \in (1, s]$, $\lambda_\mu, \lambda_\theta > 0$ and

$$A = \left\{ \theta^1 \in [L^s(\Sigma)]^{n_{f,1}} \;\middle|\; \forall k = 1, 2, ..., n_{f,1} : \left\| \theta^1_k \right\|_s \leq 1, \int_\Sigma \theta^1_k \, dx = 0 \right\}.$$

Here, Ω_μ denotes the domain of definition of the coefficients μ. The size of Ω_μ depends on the sizes of Ω, the image domain, and Σ, the domain of the filter kernels. Employing Lemma 2.58 and Lemma 2.59 we find, e.g., that the domain of definition of μ^1 is Ω_Σ and the domain of definition of μ^2 is $(\Omega_\Sigma)_\Sigma$ and so on. For $\eta \in \mathcal{M}(K)$, we say, that $\eta \geq 0$, if for all $f \in C(K)$, $f \geq 0$ it holds true , that $\eta(f) \geq 0$. Accordingly, $\theta^i \geq 0$, if all its components satisfy $\theta^i_{k,l} \geq 0$. As a norm of a vector of Borel measures we use the sum of the norms of its components, i.e., for $\eta = (\eta_i)_{i=1}^N \in [\mathcal{M}(K)]^N$,

$$\left\| \eta \right\|_{\mathcal{M}} \coloneqq \sum_{i=1}^{N} \left\| \eta_i \right\|_{\mathcal{M}},$$

but any norm in $\mathbb{R}^N$ applied to $(\|\eta_i\|)_{i=1}^N$ would yield an equivalent norm. In the case, that there are no parameters μ, θ, such that v can be expressed as the output of the given neural network with these parameters, $\mathcal{G}(v)$ is the infimum over the empty set and we therefore set $\mathcal{G}(v) = \infty$. For the sake of brevity, we introduce the following notation,

- $M_1 \coloneqq [L^s(\Sigma)]^{n_{f,1}}$,
- $M_i \coloneqq [\mathcal{M}(\overline{\Sigma})]^{n_{f,i} \times n_{f,i-1}}$ for $i > 1$,
- $M_\theta \coloneqq M_1 \times M_2 \times ... \times M_n$
- $M_\mu \coloneqq [\mathcal{M}(\Omega_\mu)]^{n_{f,n}}$.

The different treatment of the first layer of filter kernels θ^1 compared to the remaining layers is explained in detail in Remark 2.62. The reason for choosing the coefficients and kernels to be Borel measures on open and compact sets, respectively, and not both in the same manner will be explained later in Remark 2.66, since the explanation relies on a result, we have not discussed yet.

Remark 2.62. *There is a particular reason, why we only allow the filter kernels in the first layer θ^1 to be functions in $L^s(\Sigma)$ and choose all other kernels to be measures. Assume, that we would have chosen $L^s(\Sigma)$ as the underlying space for all filter kernels. Then, by Lemma 2.59, we would find that $\mu_l^{n-1} \in L^s(\Omega_{n-1})$, where Ω_i shall now denote the domain of definition of μ^i for all i. Recall, that Young's inequality [8, Satz 3.13] states, that for $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ and $f \in L^p(\mathbb{R}^d)$, $g \in L^q(\mathbb{R}^d)$ we have, $f * g \in L^r(\mathbb{R}^d)$. Since $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ implies, that $r \geq \max\{p, q\}$ with strict inequality if $p, q > 1$, we find that the convolution increases regularity in terms of increasing the Hölder exponent. For instance, we find for $s \geq 2$ after the second convolution in our network, that $\mu_l^{n-2} \in L^\infty(\Omega_{n-2})$ and for $s \in (1, 2)$, that $\mu_l^{n-2} \in L^{\frac{s}{2-s}}(\Omega_{n-2})$ with $\frac{s}{2-s} > s$. Iterating this procedure shows, that μ^i is of higher regularity, the lower i is. In particular, this would further restrict the the images, we can describe with our model.*

The following remark shows, that the assumption, that the texture part v is generated by a convolutional neural network is not too restrictive. That is, a network according to (2.4) of appropriate size is able to generate any given texture/image.

Remark 2.63. *For an appropriate choice of $(n_{f,i})_{i=1}^n$, for every $u \in L^s(\Omega)$, there exists (μ, θ), such that $v(\mu, \theta) = u$.*

Proof. In the following proof, we will denote the domain of definition of μ^i as Ω_i and we will assume for simplicity, $0 \in \Sigma$. First we note, that for any $x \in \Sigma$, $y \in \Omega_i$, $\delta_x \in \mathcal{M}(\overline{\Sigma})$ and $\delta_y \in \mathcal{M}(\Omega_i)$ and we find $\delta_y * \delta_x = \delta_{(x+y)} \in \mathcal{M}(\Omega_{i-1})$, which is zero in the case, that $x + y \notin \Omega_{i-1}$. Moreover, for $g \in L^s(\Sigma)$ and $y_0 \in \Omega_1$, we compute for the convolution $\delta_{y_0} * g \in L^s(\Omega)$

$$\delta_{y_0} * g(f) = \int_{\Omega_1} \int_{\Sigma} \tilde{f}(x + y) g(x) \, dx \, d\delta_{y_0} = \int_{\Sigma} \tilde{f}(x + y_0) g(x) \, dx =$$

$$= \int_{\mathbb{R}^d} \tilde{f}(x + y_0) \tilde{g}(x) \, dx = \int_{\mathbb{R}^d} \tilde{f}(z) \tilde{g}(z - y_0) \, dx = \int_{\Omega} f(z) \tilde{g}(z - y_0) \, dz,$$

i.e., $\delta_{y_0} * g = \tilde{g}(. - y_0)$. By boundedness of Ω and the fact, that Σ contains a neighborhood of 0, there exist $n_{f,1} \in \mathbb{N}$ and $(x_i)_{i=1}^{n_{f,1}} \subset \Omega$, such that

$$\Omega \subset \bigcup_{i=1}^{n_{f,1}} (x_i + \Sigma),$$

i.e., we can cover all of Ω with translated copies of Σ with 'centers' $x_i \in \Omega$. Moreover, let $A_i \subset \Sigma$, such that all $x_i + A_i$ are disjoint and $\Omega \subset \bigcap_{i=1}^{n_{f,1}} (x_i + A_i)$. This can be achieved, e.g., by defining $A_1 = \Sigma$ and for $i > 1$,

$$A_i = (x_i + \Sigma) \setminus \left(\bigcup_{j=1}^{i-1} (x_j + A_j) \right) - x_i.$$

Let $u \in L^s(\Omega)$ be arbitrary. Defining

$$g_i : \Sigma \to \mathbb{R}$$
$$g_i(x) = \begin{cases} \tilde{u}(x + x_i) & \text{if } x \in A_i \\ 0 & \text{else,} \end{cases}$$

we find

$$\delta_{x_i} * g_i(x) = \tilde{g}_i(x - x_i) = \begin{cases} \tilde{u}(x - x_i + x_i) = \tilde{u}(x) & \text{if } x \in x_i + A_i \\ 0 & \text{else.} \end{cases}$$

Hence, for $x \in \Omega$, we find

$$\sum_{i=1}^{n_{f,1}} \delta_{x_i} * g_i(x) = \sum_{i=1}^{n_{f,1}} \chi_{(x_i + A_i)}(x) \tilde{u}(x) = u(x),$$

Therefore, we have already determined $n_{f,1}$, $\mu^1 = (\delta_{x_i})_{i=1}^{n_{f,1}}$ and $\theta^1 = (g_i)_{i=1}^{n_{f,1}}$. The remaining part is now rather simple. As stated in the beginning, $\delta_y * \delta_x = \delta_{(x+y)}$, therefore we can successively determine the necessary μ^i and θ^i in the deeper layers. For all i, $x_i \in \Omega_1$ can be written as $x_i = x_i^2 + y_i^2$ with $x_i^2 \in \Sigma$ and $y_i^2 \in \Omega_2$, consequently, $\delta_{x_i} = \mu_i^1 = \delta_{y_i^2} * \delta_{x_i^2}$, hence we set $n_{f,2} = n_{f,1}$, $\mu_i^2 = \delta_{y_i^2}$, $\theta_{i,i}^2 = \delta_{x_i^2}$, $\theta_{j,i}^2 = 0$ for $j \neq i$ and obtain

$$\sum_{k=1}^{n_{f,2}} \mu_k^2 * \theta_{k,l}^2 = \mu_l^2 * \theta_{l,l}^2 = \delta_{y_l^2} * \delta_{x_l^2} = \delta_{(x_l^2 + y_l^2)} = \delta_{x_l} = \mu_l^1.$$

This procedure can be repeated until μ and θ are fully determined, such that $v(\mu, \theta) = u$. Also note, that $(n_{f,i})_{i=1}^n$ solely depends on the domains Ω and Σ. □

We now begin our analysis of the functional $\mathcal{G}$. Our main goals are to prove, that $\mathcal{G}$ is proper, coercive and lower semi-continuous, which makes it an appropriate regularizing functional. We start by proving a general continuity result, which is a key ingredient for well-posedness of our method.

Lemma 2.64. *[Sequential weak*-continuity of the convolution]*

1. *Let $\mu_n, \mu \in \mathcal{M}(\Omega_\Sigma)$, $\theta_n, \theta \in \mathcal{M}(\overline{\Sigma})$, such that $\mu_n \overset{*}{\rightharpoonup} \mu$ and $\theta_n \overset{*}{\rightharpoonup} \theta$. Then $\mu_n * \theta_n \overset{*}{\rightharpoonup} \mu * \theta$ in $\mathcal{M}(\Omega)$.*

2. *Let $s \in (1, \infty]$ and $q \in (1, s]$, $q < \infty$. Further, let $\mu_n, \mu \in \mathcal{M}(\Omega_\Sigma)$, $g_n, g \in L^s(\Sigma)$, such that $\mu_n \overset{*}{\rightharpoonup} \mu$ and $g_n \overset{*}{\rightharpoonup} g$. Then $\mu_n * g_n \rightharpoonup \mu * g$ in $L^q(\Omega)$.*

Proof. **Step 1:** We show the convergence of a subsequence $\mu_{n_k} * \theta_{n_k} \overset{*}{\rightharpoonup} \mu * \theta$:

First, we notice, that by Lemma 2.10, the sequences $(\mu_n)_n$ and $(\theta_n)_n$ are both bounded with respect to $\| \, . \, \|_{\mathcal{M}}$ as they are weak* convergent. Using Lemma 2.58, we also find,

$$\|\mu_n * \theta_n\|_{\mathcal{M}} \leq \|\mu_n\|_{\mathcal{M}} \|\theta_n\|_{\mathcal{M}} .$$

Therefore, $(\mu_n * \theta_n)_n$ is a bounded sequence in $\mathcal{M}(\Omega)$. Hence, Remark 2.32 implies, that there exist $\eta \in \mathcal{M}(\Omega)$ and a subsequence $(\mu_{n_k} * \theta_{n_k})_k$, such that

$$\mu_{n_k} * \theta_{n_k} \overset{*}{\rightharpoonup} \eta \text{ as } k \to \infty.$$

What is left to show, is, that $\eta = \mu * \theta$. To this aim, we will show, that for arbitrary $f \in C_0(\Omega)$, $\eta(f) = \mu * \theta(f)$. So let $f \in C_0(\Omega)$. We find for any $y \in \Omega_\Sigma$, due to the weak* convergence $\theta_n \overset{*}{\rightharpoonup} \theta$,

$$\int_{\overline{\Sigma}} \tilde{f}(x+y) \; d\theta_{n_k}(x) \to \int_{\overline{\Sigma}} \tilde{f}(x+y) \; d\theta(x) \text{ as } k \to \infty. \tag{2.5}$$

The crux of the proof will be, to show that the convergence in (2.5) is uniform in y. Denote

$$g_{n_k}(y) := \int_{\overline{\Sigma}} \tilde{f}(x+y) \; d\theta_{n_k}(x)$$

and

$$g(y) := \int_{\overline{\Sigma}} \tilde{f}(x+y) \; d\theta(x).$$

Then $g_{n_k}, g \in C_0(\Omega_\Sigma)$ as shown in Lemma 2.58 and $g_{n_k}(y) \to g(y)$ for any $y \in \mathbb{R}^d$. Moreover, for $\delta > 0$ and $y_1, y_2 \in \mathbb{R}^d$, $|y_1 - y_2| < \delta$, we find

$$\begin{aligned} |g_{n_k}(y_1) - g_{n_k}(y_2)| &= \left| \int_{\overline{\Sigma}} \tilde{f}(x+y_1) - \tilde{f}(x+y_2) \; d\theta_{n_k}(x) \right| \leq \\ &\leq \|\theta_{n_k}\|_{\mathcal{M}} \sup_{x \in \overline{\Sigma}} \left| \tilde{f}(x+y_1) - \tilde{f}(x+y_2) \right| \leq \\ &\leq \underbrace{\sup_{m \in \mathbb{N}} \|\theta_m\|_{\mathcal{M}}}_{\substack{<\infty \\ \textit{Lemma } 2.10}} \sup_{z_1, z_2 \colon |z_1 - z_2| < \delta} \left| \tilde{f}(z_1) - \tilde{f}(z_2) \right| \to 0 \quad \text{as } \delta \to 0. \end{aligned}$$

The convergence to zero is assured by the uniform continuity of $\tilde{f}$. Since the last expression tends to zero independently of k and y_1, y_2, we see, that the sequence

$(g_{n_k})_k \subset C_0(\Omega_\Sigma) \subset C(\overline{\Omega_\Sigma})$ is uniformly equicontinuous. Moreover, $(g_{n_k})_k$ is uniformly bounded, since

$$|g_{n_k}(y)| = |\int_{\overline{\Sigma}} \tilde{f}(x+y)\, d\theta_{n_k}(x)| \leq \|\theta_{n_k}\|_{\mathcal{M}} \|f\|_\infty \leq \sup_{l\in\mathbb{N}} \|\theta_l\|_{\mathcal{M}} \|f\|_\infty .$$

Hence, the Arzelà-Ascoli theorem, Corollary 2.21, states, that there is a subsequence $(g_{n_{k_l}})_l$ converging to g uniformly, i.e.

$$\sup_{y\in\Omega_\Sigma} |g_{n_{k_l}}(y) - g(y)| = \sup_{y\in\Omega_\Sigma} |\int_{\overline{\Sigma}} \tilde{f}(x+y)\, d\theta_{n_k}(x) - \int_{\overline{\Sigma}} \tilde{f}(x+y)\, d\theta(x)| \to 0 \quad \text{as } k \to 0.$$

This enables us, to make the following estimate.

$$\begin{aligned} &|\mu_{n_{k_l}} * \theta_{n_{k_l}}(f) - \mu * \theta(f)| \leq \\ &\leq |\mu_{n_{k_l}} * \theta_{n_{k_l}}(f) - \mu_{n_{k_l}} * \theta(f)| + |\mu_{n_{k_l}} * \theta(f) - \mu * \theta(f)| = \\ &= |\mu_{n_{k_l}}(g_{n_{k_l}} - g)| + |\mu_{n_{k_l}}(g) - \mu(g)| \leq \\ &\leq \left\|\mu_{n_{k_l}}\right\|_{\mathcal{M}} \sup_{y\in\Omega_\Sigma} |g_{n_{k_l}}(y) - g(y)| + |\mu_{n_{k_l}}(g) - \mu(g)| \leq \\ &\leq \underbrace{\sup_{l\in\mathbb{N}} \|\mu_l\|_{\mathcal{M}}}_{\substack{<\infty \\ \textit{Lemma } 2.10}} \underbrace{\sup_{y\in\Omega_\Sigma} |g_{n_{k_l}}(y) - g(y)|}_{(i)} + \underbrace{|\mu_{n_{k_l}}(g) - \mu(g)|}_{(ii)} \to 0 \quad \text{as } l \to 0. \end{aligned}$$

Here, (i) goes to zero as we have proven before and (ii) goes to zero by the weak* convergence $\mu_{n_{k_l}} \overset{*}{\rightharpoonup} \mu$. Thus, $\mu_{n_{k_l}} * \theta_{n_{k_l}}(f) \to \mu * \theta(f)$ as $l \to \infty$. But notice, that for this subsequence, we still have, that

$$\mu_{n_{k_l}} * \theta_{n_{k_l}} \overset{*}{\rightharpoonup} \eta,$$

i.e., also $\mu_{n_{k_l}} * \theta_{n_{k_l}}(f) \to \eta(f)$. Consequently, we find, that $\mu * \theta(f) = \eta(f)$ by uniqueness of the limit. Since $f \in C_0(\Omega)$ was arbitrary, this implies, that $\mu * \theta = \eta$ and $\mu_{n_k} * \theta_{n_k} \overset{*}{\rightharpoonup} \mu * \theta$.

Step 2: We prove convergence of the original sequence $(\mu_n * \theta_n)_n$:
Assume $\mu_n * \theta_n \overset{*}{\not\rightharpoonup} \mu * \theta$. In other words, there exists $f \in C_0(\Omega)$, $\epsilon > 0$ and a subsequence $(\mu_{n_k} * \theta_{n_k})_k$, such that

$$|\mu_{n_k} * \theta_{n_k}(f) - \mu * \theta(f)| > \epsilon$$

for all k. But for these subsequences we still have, that $\mu_{n_k} \overset{*}{\rightharpoonup} \mu$ and $\theta_{n_k} \overset{*}{\rightharpoonup} \theta$. Therefore, by the same procedure as in Step 1, we can extract a further subsequence $(\mu_{n_{k_l}} * \theta_{n_{k_l}})_l$, such that

$$\mu_{n_{k_l}} * \theta_{n_{k_l}} \overset{*}{\rightharpoonup} \mu * \theta$$

as $l \to \infty$, implying also $|\mu_{n_{k_l}} * \theta_{n_{k_l}}(f) - \mu * \theta(f)| \to 0$ and therefore contradicting our assumption. Hence, $\mu_n * \theta_n \overset{*}{\rightharpoonup} \mu * \theta$.

2. Notice, that $h \mapsto \int_\Sigma h(x)u(x)\, dx$ for $u \in L^s(\Sigma)$ is actually an element of $\mathcal{M}(\overline{\Sigma})$. Defining $\theta, \theta_n \in \mathcal{M}(\overline{\Sigma})$ as

$$\theta(h) := \int\limits_\Sigma h(x)g(x)\, dx$$
$$\theta_n(h) := \int\limits_\Sigma h(x)g_n(x)\, dx$$

for $h \in C(\overline{\Sigma})$, we see, that $\mu_n * \theta_n(f) = \mu_n * g_n(f)$ for any $f \in C_0(\Omega)$ and the convergence $g_n \overset{*}{\rightharpoonup} g$ in $L^s(\Sigma)$ implies $\theta_n \overset{*}{\rightharpoonup} \theta$ in $\mathcal{M}(\overline{\Sigma})$. Therefore, we do not need to repeat all the computations above and can already conclude, that $\mu_n * \theta_n \overset{*}{\rightharpoonup} \mu * \theta$ in $\mathcal{M}(\Omega)$, i.e., $\mu_n * g_n(f) \to \mu * g(f)$ for any $f \in C_0(\Omega)$. In order to prove the stronger notion of convergence, stated above, let now $f \in L^{q'}(\Omega)$ with $q' \in (1, \infty)$, such that $\frac{1}{q} + \frac{1}{q'} = 1$. By density, we can find $(f_m)_m \in C_c^\infty(\Omega)$ such that $f_m \to f$ in $L^{q'}(\Omega)$. Now, let $\epsilon > 0$ be arbitrarily small. Then we can pick $m_0 \in \mathbb{N}$, such that

$$\|f - f_{m_0}\|_{q'} < \frac{\epsilon}{2C\left(\sup\limits_{n\in\mathbb{N}} \|\mu_n\|_{\mathcal{M}} \sup\limits_{n\in\mathbb{N}} \|g_n\|_s + \|\mu\|_{\mathcal{M}} \|g\|_s\right)},$$

where C is the constant from Lemma 2.59. Note, that the denominator is only zero, in the case that all elements of the sequence $(\mu_n * g_n)_n$ are zero. In that case, the assertion, we want to prove, holds trivially. For this fixed m_0, we can then find $n_0 \in \mathbb{N}$, such that for $n > n_0$, it holds true that

$$|\mu_n * g_n(f_{m_0}) - \mu * g(f_{m_0})| < \frac{\epsilon}{2},$$

because $\mu_n * \theta_n \overset{*}{\rightharpoonup} \mu * \theta$. Accordingly, for all $n > n_0$

$$\begin{aligned} &|\mu_n * g_n(f) - \mu * g(f)| \leq \\ &\leq |\mu_n * g_n(f - f_{m_0})| + |\mu_n * g_n(f_{m_0}) - \mu * g(f_{m_0})| + |\mu * g(f_{m_0} - f)| \leq \\ &\underbrace{\leq}_{Lemma\ 2.59} C\|\mu_n\|_{\mathcal{M}} \|g_n\|_s \|f - f_{m_0}\|_{q'} + |\mu_n * g_n(f_{m_0}) - \mu * g(f_{m_0})| + \\ &+ C\|\mu\|_{\mathcal{M}} \|g\|_s \|f - f_{m_0}\|_{q'} \leq \\ &\leq C\left(\sup_{n\in\mathbb{N}} \|\mu_n\|_{\mathcal{M}} \sup_{n\in\mathbb{N}} \|g_n\|_s + \|\mu\|_{\mathcal{M}} \|g\|_s\right) \|f - f_{m_0}\|_{q'} + |\mu_n * g_n(f_{m_0}) - \mu * g(f_{m_0})| < \epsilon. \end{aligned}$$

Hence, $|\mu_n * g_n(f) - \mu * g(f)| \to 0$ as $n \to 0$. Since this is true for all $f \in L^{q'}(\Omega)$, we conclude, that $\mu_n * g_n \rightharpoonup \mu * g$ in $L^q(\Omega)$.

□

Corollary 2.65. *Let $s \in (1, \infty]$ and $q \in (1, s]$, $q < \infty$. Assume, that*

- $\mu_m \overset{*}{\rightharpoonup} \mu$ *in* M_μ,
- $(\theta_m)^i \overset{*}{\rightharpoonup} \theta^i$ *in* M_i *for* $i > 1$ *and*
- $(\theta_m)^1 \rightharpoonup \theta^1$ *in* M_1.

as $m \to \infty$. *Then also*

$$\mathcal{N}(\mu_m, \theta_m) \rightharpoonup \mathcal{N}(\mu, \theta) \quad \text{in } L^q(\Omega)$$

as $m \to \infty$.

Proof. The proof is simply a successive application of Lemma 2.64. □

Remark 2.66. *Having proven Lemma 2.64, we now explain, why we have chosen* $\theta \in \mathcal{M}(\overline{\Sigma})$ *and not* $\mathcal{M}(\Sigma)$, *which were less restrictive. The issues in this case will arise, when investigating continuity of the convolution as in Lemma 2.64. Consider as a counter example* $\Omega = (-2, 2)^2$, $\Sigma = (-1, 1)^2$ *and therefore,* $\Omega_\Sigma = (-3, 3)^2$. *Moreover, define the measures* $\theta_n = \delta_{(1-\frac{1}{n}, 1-\frac{1}{n})}$, $\theta = 0$, $\mu_n = \mu = \delta_{(0,0)}$ *and* $\lambda = \delta_{(1,1)}$, *where* δ_x *denotes the Dirac-measure in the point* x. *Then*

$$\begin{aligned}
\text{for } f \in C_0(\Omega_\Sigma)\text{:} \quad \mu_n(f) &= f(0,0) \to f(0,0) = \mu(f), \\
\text{for } f \in C_0(\Sigma)\text{:} \quad \theta_n(f) &= f(1 - \frac{1}{n}, 1 - \frac{1}{n}) \to f(1,1) = 0 = \theta(f), \\
\text{for } f \in C_0(\Omega)\text{:} \quad \mu_n * \theta_n(f) &= \int_{\Omega_\Sigma} \int_\Sigma \tilde{f}(x+y)\, d\theta_n(x)\, d\mu_n(y) = \\
&= \int_{\Omega_\Sigma} \tilde{f}((1 - \frac{1}{n}, 1 - \frac{1}{n}) + y)\, d\mu_n(y) = \\
&= f(1 - \frac{1}{n}, 1 - \frac{1}{n}) \to f(1,1) = \lambda(f) \text{ and} \\
\text{for } f \in C_0(\Omega)\text{:} \quad \mu * \theta(f) &= \int_{\Omega_\Sigma} \int_\Sigma \tilde{f}(x+y)\, d\theta(x)\, d\mu(y) = 0
\end{aligned}$$

Therefore, $\mu_n \overset{*}{\rightharpoonup} \mu$ *in* $\mathcal{M}(\Omega_\Sigma)$, $\theta_n \overset{*}{\rightharpoonup} \theta$ *in* $\mathcal{M}(\Sigma)$ *and* $\mu_n * \theta_n \overset{*}{\rightharpoonup} \lambda$ *in* $\mathcal{M}(\Omega)$, *but* $\lambda = \delta_{(1,1)} \neq 0 = \mu * \theta$ *in* $\mathcal{M}(\Omega)$. *The problem lies in the fact, that the convergence of* θ_n *depends on the space we consider. Applied to functions with zero boundary condition,* $\theta_n \overset{*}{\rightharpoonup} 0$, *but applied to functions in* $C(\overline{\Sigma})$, $\theta_n \overset{*}{\rightharpoonup} \delta_{(1,1)} \neq 0$. *Since, in the convolution, we apply* θ_n *to* $\tilde{f}(. + y)$, *which is not necessarily zero on the boundary of* Σ, *problems will occur, when choosing* $\theta_n \in \mathcal{M}(\Sigma) = (C_0(\Sigma))^*$.

In the following, in order to simplify notation, we denote for $\theta \in M_\theta$ and $\mu \in M_\mu$,

$$G(\mu,\theta) = \lambda_\mu \|\mu\|_{\mathcal{M}} + \lambda_\theta \sum_{i=2}^{n} \left(\left\|\theta^i\right\|_{\mathcal{M}} + \mathcal{I}_{\{.\geq 0\}}(\theta^i) \right) + \mathcal{I}_A(\theta^1). \tag{2.6}$$

This implies, that $\mathcal{G}(v) = \inf_{\substack{\mu,\theta: \\ v=\mathcal{N}(\mu,\theta)}} G(\mu,\theta)$, where the infimum is taken over $\mu \in M_\mu$, and $\theta \in M_\theta$. Recall, that we assumed $s \in (1,\infty)$ and $q \in (1,s]$, therefore we are allowed to use Lemma 2.64 and reflexivity of $L^s(\Sigma)$ in the following.

Lemma 2.67. *Let $v \in L^q(\Omega)$ and assume*

- $\mu_m \overset{*}{\rightharpoonup} \mu$ *in* M_μ,
- $(\theta_m)^i \overset{*}{\rightharpoonup} \theta^i$ *in* M_i *for* $i > 1$ *and*
- $(\theta_m)^1 \rightharpoonup \theta^1$ *in* M_1.

as $m \to \infty$. Then

$$G(\mu,\theta) \leq \liminf_{m\to\infty} G(\mu_m,\theta_m).$$

Proof. We will consider all terms of G separately at first.

$\|\,.\,\|_{\mathcal{M}}$: By Lemma 2.45, $\|\,.\,\|_{\mathcal{M}}$ is lsc with respect to weak*-convergence, as it can be viewed as a dual norm (Theorem 2.30). Therefore we find,

$$\|\mu_k\|_{\mathcal{M}} \leq \liminf_{m\to\infty} \|(\mu_m)_k\|_{\mathcal{M}} \text{ for } k = 1,2,...n_{f,n} \text{ and}$$
$$\left\|\theta^i_{k,l}\right\|_{\mathcal{M}} \leq \liminf_{m\to\infty} \left\|(\theta_m)^i_{k,l}\right\|_{\mathcal{M}}, \text{ for } i > 1 \text{ and all } k \text{ and } l.$$

$\mathcal{I}_{\{.\geq 0\}}$: We want to show, that $\mathcal{I}_{\{.\geq 0\}}(\theta^i) \leq \liminf_{m\to\infty} \mathcal{I}_{\{.\geq 0\}}((\theta_m)^i)$. If the right-hand side is ∞, we are done. Otherwise, assume $\liminf_{m\to\infty} \mathcal{I}_{\{.\geq 0\}}((\theta_m)^i) = \lim_{m\to\infty} \mathcal{I}_{\{.\geq 0\}}((\theta_m)^i) < \infty$, which we could always achieve by taking a subsequence. This means, that for sufficiently large m, we find that for all $f \in C(\overline{\Sigma})$, $f \geq 0$, it holds true that $(\theta_m)^i_{k,l}(f) \geq 0$ for $i > 1$ and all k and l. Due to weak* convergence, also

$$\theta^i_{k,l}(f) = \lim_{m\to\infty} (\theta_m)^i_{k,l}(f) \geq 0,$$

i.e., $\theta^i \geq 0$, implying $\mathcal{I}_{\{.\geq 0\}}(\theta^i) = 0 \leq \liminf_{m\to\infty} \mathcal{I}_{\{.\geq 0\}}((\theta_m)^i) = 0$.

$\mathcal{I}_A$: Finally, we want to show, that $\mathcal{I}_A(\theta_k^1) \leq \liminf_{m\to\infty} \mathcal{I}_A((\theta_m)_k^1)$. As above, assume, that the right-hand side of the inequality is finite, because otherwise, the result follows trivially and again we suppose, that $\liminf_{m\to\infty} \mathcal{I}_A((\theta_m)_k^1) = \lim_{m\to\infty} \mathcal{I}_A((\theta_m)_k^1)$. Therefore, by definition of A, we know, that $\left\|(\theta_m)_k^1\right\|_s \leq 1$ and $\int_\Sigma (\theta_m)_k^1 \, dx = 0$ for m sufficiently large. Since $\| \, . \, \|_s$ is convex and continuous in L^s, it is also weakly lsc. Thus,

$$\left\|\theta_k^1\right\|_s \leq \liminf_{m\to\infty} \left\|(\theta_m)_k^1\right\|_s \leq 1.$$

Moreover, due to the fact, that the constant 1 is in $L^{s'}(\Sigma) = (L^s(\Sigma))^*$ with s' the Hölder conjugate exponent of s, because $|\Sigma| < \infty$, we also obtain

$$\int_\Sigma 1 \cdot \theta_k^1 \, dx = \lim_{m\to\infty} \int_\Sigma 1 \cdot (\theta_m)_k^1 \, dx = 0,$$

Altogether, $\theta^1 \in A$ and consequently $\mathcal{I}_A(\theta_k^1) = 0 \leq \liminf_{m\to\infty} \mathcal{I}_A((\theta_m)_k^1)$.

Now it is time, to deduce the desired result. For real sequences $(a_m)_m$, $(b_m)_m$, we always have

$$\liminf_{m\to\infty} a_m + \liminf_{m\to\infty} b_m = \lim_{m\to\infty} \left(\inf_{k\geq m} a_k + \inf_{k\geq m} b_k \right) \leq \lim_{m\to\infty} \inf_{k\geq m} (a_k + b_k) = \liminf_{m\to\infty} (a_m + b_m).$$

Noting, that $\lambda_\mu, \lambda_\theta > 0$, a successive application of this inequality eventually yields

$$G(\mu,\theta) \leq \liminf_{m\to\infty} G(\mu_m, \theta_m).$$

□

Lemma 2.68. *Let $v \in L^q(\Omega)$ and assume, that $\mathcal{G}(v) < \infty$. Then the infimum in $\mathcal{G}(v)$ is attained.*

Proof. By assumption $\mathcal{G}(v) < \infty$, that is, there exist parameters $\mu \in M_\mu$ and $\theta \in M_\theta$, such that $v = \mathcal{N}(\mu,\theta)$ and $G(\mu,\theta) < \infty$. We can therefore pick minimizing sequences $(\mu_m)_m \subset M_\mu$, $(\theta_m)_m \subset M_\theta$, such that for all m, $v = \mathcal{N}(\mu_m, \theta_m)$ and $\mathcal{G}(v) = \lim_{m\to\infty} G(\mu_m, \theta_m)$. The definition of G immediately implies, that the sequences μ_m and θ_m are bounded in their spaces of definition. By reflexivity of $L^s(\Sigma)$ for $s \in (1,\infty)$ and the theorem of Banach-Alaoglu, Remark 2.32, we can therefore find a subsequence again labeled m and μ, θ, such that $\mu_m \overset{*}{\rightharpoonup} \mu$ in M_μ, $(\theta_m)^1 \rightharpoonup \theta^1$ in M_1 and for $i > 1$, $(\theta_m)^i \overset{*}{\rightharpoonup} \theta^i$ in M_i as $m \to \infty$. By Corollary 2.65, $\mathcal{N}(\mu,\theta) = \lim_{m\to\infty} \mathcal{N}(\mu_m,\theta_m) = v$. Note, that this subsequence is still a minimizing sequence of G. Hence, by Lemma 2.67 we find

$$G(\mu,\theta) \leq \liminf_{m\to\infty} G(\mu_m,\theta_m) = \mathcal{G}(v),$$

which shows, that (μ,θ) is the desired minimizer and concludes the proof. □

Lemma 2.69. *The texture prior* $\mathcal{G} : L^q(\Omega) \to [0, \infty]$ *is weakly lsc.*

Proof. Let $v_m \rightharpoonup v$ in $L^q(\Omega)$. We want to show, that $\mathcal{G}(v) \le \liminf_{m\to\infty} \mathcal{G}(v_m)$. Let us assume, that $\liminf_{m\to\infty} \mathcal{G}(v_m) < \infty$, because otherwise, there is nothing to prove. Assume further, that $\liminf_{m\to\infty} \mathcal{G}(v_m) = \lim_{m\to\infty} \mathcal{G}(v_m)$ and that $\mathcal{G}(v_m) < \infty$ for all m, which could always be achieved by extracting a subsequence. By Lemma 2.68, for every m, there exist $\mu_m \in M_\mu$ and $\theta_m \in M_\theta$, such that $v_m = \mathcal{N}(\mu_m, \theta_m)$ and $\mathcal{G}(v_m) = G(\mu_m, \theta_m)$. The fact, that $(\mathcal{G}(v_m))_m$ is bounded as a convergent sequence and the definition of G imply, that the sequences $(\mu_m)_m$ and $(\theta_m)_m$ are bounded in the respective spaces. As in the proof of Lemma 2.68, we can extract a subsequence again labelled with m and find μ, θ, such that $\mu_m \overset{*}{\rightharpoonup} \mu$ in M_μ, $(\theta_m)^1 \rightharpoonup \theta^1$ in M_1 and for $i > 1$ $(\theta_m)^i \overset{*}{\rightharpoonup} \theta^i$ in M_i as $m \to \infty$. By Corollary 2.65, $v = \lim_{m\to\infty} v_m = \lim_{m\to\infty} \mathcal{N}(\mu_m, \theta_m) = \mathcal{N}(\mu, \theta)$. Therefore,

$$\mathcal{G}(v) \le G(\mu, \theta) \underbrace{\le}_{\text{Lemma 2.67}} \liminf_{m\to\infty} G(\mu_m, \theta_m) = \liminf_{m\to\infty} \mathcal{G}(v_m) = \lim_{m\to\infty} \mathcal{G}(v_m).$$

□

Lemma 2.70. $\mathcal{G} : L^q(\Omega) \to [0, \infty]$ *is proper and coercive.*

Proof. In order to show, that $\mathcal{G}$ is proper, we simply note, that $\mathcal{G}(0) = 0$, since $\mathcal{N}(0, 0) = 0$ and $G(0, 0) = 0$. To prove coercivity, let $(v_m)_m \subset L^q(\Omega)$, such that $\|v_m\|_q \to \infty$. We have to show, that $\mathcal{G}(v_m) \to \infty$ as well. Assume to the contrary, $\mathcal{G}(v_m) \not\to \infty$, i.e., there is a subsequence m_k, such that $(\mathcal{G}(v_{m_k}))_k$ is bounded. By Lemma 2.68, we can pick μ_{m_k} and θ_{m_k}, such that $v_{m_k} = \mathcal{N}(\mu_{m_k}, \theta_{m_k})$ and $\mathcal{G}(v_{m_k}) = G(\mu_{m_k}, \theta_{m_k})$. Therefore, as before μ_{m_k} and θ_{m_k} are bounded by the very definition of G and the boundedness $\mathcal{G}(v_{m_k}) = G(\mu_{m_k}, \theta_{m_k})$. Employing Lemma 2.59 and Lemma 2.58, this implies, that also $v_{m_k} = \mathcal{N}(\mu_{m_k}, \theta_{m_k})$ is bounded in $L^q(\Omega)$, contradicting our assumption. □

2.2.4 The Cartoon Prior $\mathcal{R}$

This section is dedicated to the introduction and mathematical analysis of the cartoon prior $\mathcal{R}$ from (P). Luckily, regularizing functionals for piecewise smooth images are already well-understood. A very well-known example of such regularizing functionals is the total variation functional TV (see Definition and Proposition 2.36). The total variation $L^1(\Omega) \to [0, \infty], w \mapsto \mathrm{TV}(w)$ is a lsc generalization of $w \mapsto \|\nabla w\|_1$ on the space $L^1(\Omega)$. Unfortunately, the numerical solution algorithm, we will apply later, relies on certain stronger smoothness properties, which $\|\,.\,\|_1$ lacks. Therefore, instead of TV we will from now on consider the more general functional

$$\begin{aligned} \mathcal{R} : L^1(\Omega) &\to [0, \infty] \\ w &\mapsto J^{**}(w), \end{aligned} \tag{CAR}$$

where $J^{**} : L^1(\Omega) \to [0,\infty]$ is the bipolar of

$$J : L^1(\Omega) \to [0,\infty]$$
$$w \mapsto \begin{cases} \int\limits_\Omega j(\nabla u(x))\,dx & \text{if } u \in W^{1,1}(\Omega) \\ +\infty & \text{else.} \end{cases} \tag{2.7}$$

with a function $j : \mathbb{R}^d \to [0,\infty)$. This way, under some assumptions on j, J^{**} will be lsc and convex (and hence also weakly lsc) and the discrete analogue will be smooth. We impose the following assumptions on j.

Assumption 2.71. *The function $j : \mathbb{R}^d \to [0,\infty)$ shall satisfy the following conditions:*

1. *j is convex.*
2. *j has linear growth, i.e., there exists $\gamma > 0$, such that for any $p \in \mathbb{R}^d$, we have*
$$\frac{1}{\gamma}(|p| - 1) \leq j(p) \leq \gamma(|p| + 1).$$
3. *j is Lipschitz continuous, i.e., there exists $L > 0$, such that for any $p, q \in \mathbb{R}^d$,*
$$|j(p) - j(q)| \leq L|p - q|.$$

We will now present a few examples, satisfying our assumptions, in order to show, that they are not too restrictive.

Example 2.72. *[Examples for possible j]*

The following functions satisfy Assumption 2.71:

1. $$j : \mathbb{R}^d \to [0,\infty),$$
$$j(p) = |p|.$$
2. *For $\epsilon > 0$,*
$$j : \mathbb{R}^d \to [0,\infty),$$
$$j(p) = \sqrt{|p|^2 + \epsilon}.$$
3. *(Huber TV) For $\epsilon > 0$,*
$$j : \mathbb{R}^d \to [0,\infty),$$
$$j(p) = \begin{cases} \frac{|p|^2}{2\epsilon} & \text{if } |p| < \epsilon \\ |p| - \frac{\epsilon}{2} & \text{else.} \end{cases}$$

Proof. 1. The proof is straightforward and will therefore be omitted here.

2.
- Convexity:
  Defining $h : [0,\infty) \to [0,\infty)$, $h(t) = \sqrt{t^2+\epsilon}$, we can write $j(p) = h(|p|)$. Notice, that h is monotonically increasing and $h''(t) = \frac{\epsilon}{(t^2+\epsilon)^{\frac{3}{2}}} > 0$ for all t, hence h is also strictly convex. Let $p, q \in \mathbb{R}^d$ and $\lambda \in (0,1)$. Then, by convexity of $|\,.\,|$ on $\mathbb{R}^d$, we get

$$j(\lambda p + (1-\lambda)q) = h(|\lambda p + (1-\lambda)q|) \underbrace{\leq}_{h \text{ increasing}} h(\lambda|p| + (1-\lambda)|q|) \leq$$
$$\underbrace{\leq}_{h \text{ convex}} \lambda h(|p|) + (1-\lambda)h(|q|) = \lambda j(p) + (1-\lambda)j(q).$$

- Linear growth:
  Let $p \in \mathbb{R}^d$ be arbitrary.

$$|p| = \sqrt{|p|^2} \leq \sqrt{|p|^2+\epsilon} = j(p) \leq \sqrt{|p|^2+\epsilon+2|p|\sqrt{\epsilon}} = \sqrt{(|p|+\sqrt{\epsilon})^2} = |p| + \sqrt{\epsilon} \leq$$
$$\leq \max\left\{1, \sqrt{\epsilon}\right\}(|p|+1).$$

- Lipschitz continuity:
  Using again the notation $h(t) = \sqrt{t^2+\epsilon}$ for $t \in [0,\infty)$, we find, that $h'(t) = \frac{t}{\sqrt{t^2+\epsilon}} = \frac{1}{\sqrt{1+\frac{\epsilon}{t^2}}} \leq 1$. This implies, that h is Lipschitz continuous with Lipschitz constant 1. Thus, for $p, q \in \mathbb{R}^d$,

$$|j(p) - j(q)| = |h(|p|) - h(|q|)| \leq |\,|p| - |q|\,| \leq |p-q|\,.$$

3.
- Convexity:
  Defining

$$h : [0,\infty) \to [0,\infty)$$
$$h(t) = \begin{cases} \frac{t^2}{2\epsilon} & \text{if } t < \epsilon \\ t - \frac{\epsilon}{2} & \text{else,} \end{cases}$$

  we can write $j(p) = h(|p|)$ as before. It follows

$$h'(t) = \begin{cases} \frac{t}{\epsilon} & \text{if } t < \epsilon \\ 1 & \text{else.} \end{cases}$$

  We find, that h and h' are monotonically increasing, i.e., h is increasing and convex. Let $p, q \in \mathbb{R}^d$ and $\lambda \in (0,1)$. Then, by convexity of $|\,.\,|$ on $\mathbb{R}^d$, we get as above

$$j(\lambda p + (1-\lambda)q) = h(|\lambda p + (1-\lambda)q|) \underbrace{\leq}_{h \text{ increasing}} h(\lambda|p| + (1-\lambda)|q|) \leq$$

$$\underbrace{\leq}_{h \text{ convex}} \lambda h(|p|) + (1-\lambda)h(|q|) = \lambda j(p) + (1-\lambda)j(q).$$

- Linear growth:
  It holds true, that for all $t > 0$,
  $$t - \frac{\epsilon}{2} \leq h(t) \leq t.$$
  Thus, for $p \in \mathbb{R}^d$ arbitrary, we compute
  $$|p| - \frac{\epsilon}{2} \leq h(|p|) = j(p) \leq |p|.$$
- Lipschitz continuity:
  We have, that $h'(t) \leq 1$ for all $t > 0$. Hence, h is Lipschitz continuous with Lipschitz constant 1, and we find for $p, q \in \mathbb{R}^d$,
  $$|j(p) - j(q)| = |h(|p|) - h(|q|)| \leq |\ |p| - |q|\ | \leq |p - q|.$$

□

As in Section 2.2.3, we will no mathematically analyze the cartoon prior $\mathcal{R} = J^{**}$ with the main goal being, to establish lower semi-continuity of the functional and a Poincaré type inequality.

Lemma 2.73. *Under Assumption 2.71, $J : L^1(\Omega) \to [0, \infty]$, defined as in (2.7), is convex.*

Proof. Let $u, w \in L^1(\Omega)$. We want to show, that for any $\lambda \in (0,1)$,

$$J(\lambda u + (1-\lambda)w) \leq \lambda J(u) + (1-\lambda)J(w). \tag{2.8}$$

If one of the functions u, w is not in $W^{1,1}(\Omega)$, the right-hand side of (2.8) is ∞ and we are done. So assume now $u, w \in W^{1,1}(\Omega)$ and $\lambda \in (0,1)$. Then, by linearity of ∇ and linearity and monotonicity of the integral, we find

$$J(\lambda u + (1-\lambda)w) = \int\limits_\Omega j(\lambda \nabla u(x) + (1-\lambda)\nabla w(x))\, dx \leq$$

$$\underbrace{\leq}_{j \text{ convex}} \int\limits_\Omega \lambda j(\nabla u(x)) + (1-\lambda) j(\nabla w(x))\, dx = \int\limits_\Omega \lambda j(\nabla u(x))\, dx + \int\limits_\Omega (1-\lambda) j(\nabla w(x))\, dx =$$

$$= \lambda J(u) + (1-\lambda)J(w).$$

□

Corollary 2.74. *It holds true, that*

$$J^{**} = \overline{J} = \inf \left\{ \liminf_{n\to\infty} J(u_n) \,\middle|\, u_n \to u \text{ in } L^1(\Omega) \right\}.$$

In particular, we find, that J^{**} is the lsc regularization of J on the space $L^1(\Omega)$.

Proof. Since J is convex and bounded from below by zero, Proposition 2.54 implies, that $\overline{J}$ is equal to the Γ-regularization of J, which again is equal to J^{**} according to Proposition 2.57. This proves the first inequality. The second one is an application of Lemma 2.50. □

Lemma 2.75. *Under Assumption 2.71, $J^{**} : L^1(\Omega) \to [0,\infty]$ satisfies*

$$J^{**}(u) < \infty \Longleftrightarrow u \in \mathrm{BV}(\Omega).$$

Proof." ⇒ " Assume $J^{**}(u) < \infty$. By Corollary 2.74 and Lemma 2.51, we can find a sequence $(u_n)_n \subset L^1(\Omega)$ converging to u in $L^1(\Omega)$, such that

$$J^{**}(u) = \lim_{n\to\infty} J(u_n).$$

Since $J^{**}(u) < \infty$, we can assume, without loss of generality, that $J(u_n) < \infty$ for all n, i.e., $(u_n)_n \subset W^{1,1}(\Omega)$. By density, we can also find a sequence $(w_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, such that $\|u_n - w_n\|_{1,1} \le \frac{1}{n}$. Then, $\|w_n - u\|_1 \le \|w_n - u_n\|_1 + \|u_n - u\|_1 \to 0$. We also find, that

$$\begin{aligned} \limsup_{n\to\infty} \int_\Omega |\nabla w_n|\, dx &\le \limsup_{n\to\infty} \left(\int_\Omega |\nabla w_n - \nabla u_n|\, dx + \int_\Omega |\nabla u_n|\, dx \right) \le \\ &\underbrace{\le}_{\|\nabla w_n - \nabla u_n\|_1 \to 0} \limsup_{n\to\infty} \int_\Omega |\nabla u_n|\, dx \le \\ &\underbrace{\le}_{\textit{Assumption } 2.71} \limsup_{n\to\infty} \int_\Omega \gamma(j(\nabla u_n) + 1)\, dx = \gamma J^{**}(u) + \gamma|\Omega| < \infty. \end{aligned}$$

We can further extract a subsequence $(w_{n_k})_k$, satisfying

$$\lim_{k\to\infty} \int_\Omega |\nabla w_{n_k}|\, dx = \limsup_{n\to\infty} \int_\Omega |\nabla w_n|\, dx < \infty.$$

Then, Proposition 2.40 implies, that $u \in \mathrm{BV}(\Omega)$.

"$\Leftarrow$" Assume now $u \in \mathrm{BV}(\Omega)$. Using Proposition 2.40, we can find a sequence $(u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, converging to u in $L^1(\Omega)$, such that

$$\lim_{n\to\infty} \int_\Omega |\nabla u_n| \, dx < \infty.$$

This implies,

$$\liminf_{n\to\infty} J(u_n) = \liminf_{n\to\infty} \int_\Omega j(\nabla u_n) \, dx \underbrace{\leq}_{Assumption\ 2.71} \liminf_{n\to\infty} \int_\Omega \gamma(|\nabla u_n| + 1) \, dx < \infty,$$

As a result,

$$J^{**}(u) = \inf \left\{ \liminf_{n\to\infty} J(w_n) \;\middle|\; (w_n)_n \subset L^1(\Omega), w_n \to u \text{ in } L^1(\Omega) \right\} \leq \liminf_{n\to\infty} J(u_n) < \infty.$$

□

Lemma 2.76. *Under Assumption 2.71, for every $u \in \mathrm{BV}(\Omega)$*

$$J^{**}(u) = \inf \left\{ \liminf_{n\to\infty} J(u_n) \;\middle|\; (u_n)_n \subset W^{1,1}(\Omega) \cap C^\infty(\Omega),\ u_n \to u \text{ in } L^1(\Omega) \right\}.$$

Proof. From Corollary 2.74, we already know, that

$$J^{**}(u) = \inf \left\{ \liminf_{n\to\infty} J(u_n) \;\middle|\; (u_n)_n \subset L^1(\Omega),\ u_n \to u \text{ in } L^1(\Omega) \right\}.$$

Therefore, we can already conclude, that

$$J^{**}(u) \leq \inf \left\{ \liminf_{n\to\infty} J(u_n) \;\middle|\; (u_n)_n \subset W^{1,1}(\Omega) \cap C^\infty(\Omega),\ u_n \to u \text{ in } L^1(\Omega) \right\}.$$

In order to prove the remaining inequality, let $(u_n)_n \subset L^1(\Omega)$ be a sequence, converging to u in $L^1(\Omega)$, such that

$$J^{**}(u) = \lim_{n\to\infty} J(u_n),$$

which exists according to Corollary 2.74 and Lemma 2.51. Since we assumed, $u \in \mathrm{BV}(\Omega)$, $J^{**}(u) < \infty$ and we can suppose without loss of generality, that $J(u_n) < \infty$ for all n. By definition of J this means, $u_n \in W^{1,1}(\Omega)$ for all n. By density (see Theorem 2.35), we can then find a sequence $(\phi_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, such that

$$\|\phi_n - u_n\|_{1,1} < \frac{1}{n}.$$

As a result, $\phi_n \to u$ in $L^1(\Omega)$ and

$$|J(\phi_n) - J(u_n)| \leq \int_\Omega |j(\nabla \phi_n) - j(\nabla u_n)| \, dx \leq$$

$$\underbrace{\leq}_{Assumption\ 2.71} \int_\Omega L \, |\nabla \phi_n - \nabla u_n| \, dx = L \, \|\nabla \phi_n - \nabla u_n\|_1 \leq L \, \|\phi - u\|_{1,1} \leq \frac{L}{n}.$$

Therefore,

$$|J^{**}(u) - J(\phi_n)| \le |J^{**}(u) - J(u_n)| + |J(u_n) - J(\phi_n)| \to 0$$

as $n \to \infty$. Hence, $J^{**}(u) = \lim_{n\to\infty} J(\phi_n) = \liminf_{n\to\infty} J(\phi_n)$ with a sequence $(\phi_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$ converging to u in $L^1(\Omega)$, implying

$$J^{**}(u) \ge \inf \left\{ \liminf_{n\to\infty} J(u_n) \;\middle|\; (u_n)_n \subset W^{1,1}(\Omega) \cap C^\infty(\Omega),\ u_n \to u \text{ in } L^1(\Omega) \right\}$$

and concluding the proof. □

Corollary 2.77. *Under Assumption 2.71, for every $u \in \mathrm{BV}(\Omega)$, there exists a sequence $(u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$ converging to u in $L^1(\Omega)$, such that*

$$J^{**}(u) = \lim_{n\to\infty} J(u_n).$$

Proof. By Lemma 2.76, for every $n \in \mathbb{N}$, we can find a sequence $(u_k^n)_k \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, such that $u_k^n \to u$ as $k \to \infty$ and

$$\left| \lim_{k\to\infty} J(u_k^n) - J^{**}(u) \right| \le \frac{1}{n}.$$

For any $n \in \mathbb{N}$, pick $k_n \in \mathbb{N}$, such that

$$\left\| u_{k_n}^n - u \right\|_1 \le \frac{1}{n}$$

and

$$\left| J(u_{k_n}^n) - J^{**}(u) \right| \le \frac{2}{n}.$$

Then, the sequence $(u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, defined as $u_n := u_{k_n}^n$ for $n \in \mathbb{N}$ satisfies $u_n \to u$ in $L^1(\Omega)$ and

$$\lim_{n\to\infty} J(u_n) = J^{**}(u).$$

□

Lemma 2.78. *Under Assumption 2.71, for every $u \in \mathrm{BV}(\Omega)$, we have*

$$\frac{1}{\gamma}(\mathrm{TV}(u) - |\Omega|) \le J^{**}(u) \le \gamma(\mathrm{TV}(u) + |\Omega|).$$

Proof. Since we only consider $u \in \mathrm{BV}(\Omega)$, Corollary 2.77 states, that we can find a sequence $(u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, converging to u in $L^1(\Omega)$, such that

$$J^{**}(u) = \lim_{n\to\infty} J(u_n).$$

Therefore,

$$J^{**}(u) = \lim_{n\to\infty} J(u_n) = \lim_{n\to\infty} \int\limits_\Omega j(\nabla u_n)\, dx \underbrace{\geq}_{Assumption\ 2.71} \limsup_{n\to\infty} \int\limits_\Omega \frac{1}{\gamma}(|\nabla u_n| - 1)\, dx =$$
$$= \frac{1}{\gamma}\left(\limsup_{n\to\infty} \int\limits_\Omega |\nabla u_n|\, dx - |\Omega|\right) \underbrace{\geq}_{Proposition\ 2.40} \frac{1}{\gamma}\left(\mathrm{TV}(u) - |\Omega|\right).$$

Therefore, $\frac{1}{\gamma}(\mathrm{TV} - |\Omega|) \leq J^{**}$. On the other hand, as stated in Proposition 2.40, we can also find a sequence $(u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega)$, such that $u_n \to u$ in $L^1(\Omega)$ as $n \to \infty$ and

$$\mathrm{TV}(u) = \lim_{n\to\infty} \int\limits_\Omega |\nabla u_n|\, dx.$$

Then, we find, by definition of $J^{**} = \overline{J}$ and Lemma 2.50

$$J^{**}(u) \leq \liminf_{n\to\infty} J(u_n) = \liminf_{n\to\infty} \int\limits_\Omega j(\nabla u_n)\, dx \underbrace{\leq}_{Assumption\ 2.71} \liminf_{n\to\infty} \int\limits_\Omega \gamma(|\nabla u_n| + 1)\, dx =$$
$$= \gamma(\mathrm{TV}(u) + |\Omega|).$$

□

Lemma 2.79. *[Translation invariance of J^{**}] Let*

$$K(\Omega) = \{f : \Omega \to \mathbb{R} \mid \exists w \in \mathbb{R} : f(x) = w \text{ for a.e. } x \text{ in } \Omega\}.$$

Then, for every $u \in \mathrm{BV}(\Omega)$ and $g \in K(\Omega)$ it holds true, that

$$J^{**}(u + g) = J^{**}(u).$$

Proof. Let $u \in \mathrm{BV}(\Omega)$ and $g \in K(\Omega)$. First, we note, that for $w \in W^{1,1}(\Omega)$, it holds true, that

$$J(w) = \int\limits_\Omega j(\nabla w)\, dx = \int\limits_\Omega j(\nabla(w + g))\, dx = J(w + g).$$

By Lemma 2.76, we can compute

$$J^{**}(u+g) = \inf\left\{\liminf_{n\to\infty} J(w_n) \;\middle|\; (w_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega),\ w_n \to u + g \text{ in } L^1(\Omega)\right\} =$$
$$= \inf\left\{\liminf_{n\to\infty} J(w_n - g) \;\middle|\; (w_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega),\ w_n \to u + g \text{ in } L^1(\Omega)\right\} =$$
$$= \inf\left\{\liminf_{n\to\infty} J(u_n) \;\middle|\; (u_n)_n \subset C^\infty(\Omega) \cap W^{1,1}(\Omega),\ u_n \to u \text{ in } L^1(\Omega)\right\} = J^{**}(u).$$

□

Lemma 2.80. *[Poincaré inequality] There exists a constant $C_P > 0$, such that for all p with $1 \leq p \leq \frac{d}{d-1}$*

$$\|u - u_\Omega\|_p \leq C_P \left(J^{**}(u) + 1\right) \quad \forall u \in \mathrm{BV}(\Omega).$$

Proof. This follows from Lemma 2.78 together with Lemma 2.42. □

2.2.5 Existence and Stability Analysis

In this section, we will briefly recall the proposed problem including the data fidelity term and then prove, that the problem admits a solution and that it is stable under certain conditions. The proposed problem reads as follows:

$$\min_{u,v \in L^q(\Omega)} F(u, v, y_0) := \mathcal{D}(u, y_0) + \lambda_{\mathcal{R}} \mathcal{R}(u - v) + \mathcal{G}(v), \tag{$P(y_0)$}$$

where Y is a Banach space, $y_0 \in Y$ is the data and $\lambda_{\mathcal{R}} > 0$. The functionals $\mathcal{R}$ and $\mathcal{G}$ can be found in (CAR) and (TEX), respectively. For the data term, we will consider two different cases. For $A \in \mathcal{L}(L^q(\Omega), Y)$, a forward operator, either

- $\mathcal{D}(u, y_0) = \lambda_D \psi(\|Au - y_0\|)$ with $\psi : [0, \infty) \to [0, \infty)$ continuous, increasing and coercive (linear inverse problem) or
- $\mathcal{D}(u, y_0) = \mathcal{I}_{\{.=y_0\}}(Au)$ (inpainting).

Usually, $\psi(t) = \frac{t^q}{q}$ for some $q \in [1, \infty)$.

Example 2.81. *[Forward operator for inpainting] What we typically mean by inpainting can be obtained by defining the forward operator as the restriction*

$$\begin{aligned} A : L^q(\Omega) &\to L^p(M) \\ u &\mapsto u\big|_M. \end{aligned}$$

for $M \subset \Omega$, such that $|M| > 0$ and $p \in [1, q]$.

Before stating the first main result, we prove lower semi-continuity of the objective functional in the following sense:

Lemma 2.82. *Let*

- $u_m \rightharpoonup u$ *in* $L^q(\Omega)$,
- $y_0^m \rightharpoonup y_0$ *in* Y,
- $v_m \rightharpoonup v$ *in* $L^q(\Sigma)$

as $m \to \infty$. Then,

$$F(u, v, y_0) \leq \liminf_{m \to \infty} F(u_m, v_m, y_0^m).$$

Proof. $\mathcal{D}$: Let us start by investigating the data term $\mathcal{D}$. We have to distinguish the two possible cases.

- Linear inverse problem: $\mathcal{D}(u, y_0) = \lambda_{\mathcal{D}}\psi(\|Au - y_0\|)$
  $\|\,.\,\|$ is continuous and convex in Y and therefore, also lsc with respect to weak convergence in Y (see Lemma 2.46). Lemma A.2 implies, that as a result also $\psi(\|\,.\,\|)$ is weakly lsc. Continuous linear operators between Banach spaces are also weakly continuous (see Lemma 2.9), thus $u_m \rightharpoonup u$ implies $Au_m \rightharpoonup Au$. Moreover, the sum/difference of weakly convergent sequences is again weakly convergent, therefore, $Au_m - y_0^m \rightharpoonup Au - y_0$ in Y. Altogether, we find

  $$\psi(\|Au - y_0\|) \leq \liminf_{m\to\infty} \psi(\|Au_m - y_0^m\|).$$

  Noting, that $\lambda_{\mathcal{D}} > 0$, we obtain $\mathcal{D}(u, y_0) \leq \liminf_{m\to\infty} \mathcal{D}(u_m, y_0^m)$.

- Inpainting: $\mathcal{D}(u, y_0) = \mathcal{I}_{\{.=y_0\}}(Au)$
  Again we find, that $\mathcal{D}$ is convex. In order to prove lower semi-continuity of $\mathcal{D}$, let $w_m \to w$ in $L^q(\Omega)$ and $z_0^m \to z_0$ in Y as $m \to \infty$. Our goal is, to show, that $\mathcal{D}(w, z_0) \leq \liminf_{m\to\infty} \mathcal{D}(w_m, z_0^m)$. We may assume, $\liminf_{m\to\infty} \mathcal{D}(w_m, z_0^m) < \infty$, because otherwise, there is nothing to prove. In this case, let $(w_{m_k})_k$, $(z_0^{m_k})_k$ be, such that $\liminf_{m\to\infty} \mathcal{D}(w_m, z_0^m) = \lim_{k\to\infty} \mathcal{D}(w_{m_k}, z_0^{m_k}) < \infty$. This means, for k large enough, $Aw_{m_k} = z_0^{m_k}$. Hence,

  $$Aw = \lim_{k\to\infty} Aw_{m_k} = \lim_{k\to\infty} z_0^{m_k} = z_0.$$

  Therefore, $\mathcal{D}(w, z_0) = 0$ and $\mathcal{D}$ is lsc. Again, the convexity and lower semi-continuity imply lower semi-continuity with respect to weak convergence and we conclude, $\mathcal{D}(u, y_0) \leq \liminf_{m\to\infty} \mathcal{D}(u_m, y_0^m)$.

$\mathcal{R}$: Recall, that $\mathcal{R} = J^{**}$, i.e., $\mathcal{R}$ is defined as the bipolar of the functional J. By definition of the bipolar function (Γ-regularization) J^{**} is convex and lsc in $L^1(\Omega)$ and consequently weakly lsc in $L^1(\Omega)$. For $q \geq 1$, weak convergence in L^q implies weak convergence in L^1, therefore, $u_m - v_m \rightharpoonup u - v$ in $L^1(\Omega)$. Consequently,

$$J^{**}(u - v) \leq \liminf_{m\to\infty} J^{**}(u_m - v_m).$$

Noting, that $\lambda_{\mathcal{R}} > 0$, we also obtain

$$\lambda_{\mathcal{R}}\mathcal{R}(u - v) \leq \liminf_{m\to\infty} \lambda_{\mathcal{R}}\mathcal{R}(u_m - v_m).$$

$\mathcal{G}$: We have already shown weak lower semi-continuity of $\mathcal{G}$ in Lemma 2.69. Therefore,

$$\mathcal{G}(v) \leq \liminf_{m\to\infty} \mathcal{G}(v_m).$$

For real sequences $(a_m)_m$, $(b_m)_m$, it holds true, that

$$\liminf_{m\to\infty} a_m + \liminf_{m\to\infty} b_m \leq \liminf_{m\to\infty} (a_m + b_m) .$$

A successive application of this elementary fact eventually yields

$$F(u, v, y_0) \leq \liminf_{m\to\infty} F(u_m, v_m, y_0^m).$$

□

Theorem 2.83. *[Existence of solutions] Consider $(P(y_0))$ with data $y_0 \in Y$. Suppose, that Assumption 2.71 holds true and that $q \leq \frac{d}{d-1}$. In the case of inpainting, further assume, that there exists a function $\hat{u_0} \in \mathrm{BV}(\Omega)$ such that $A\hat{u_0} = y_0$. Then, there exists a solution to $(P(y_0))$.*

Proof. First of all, we notice, that by the assumption on q, we have the continuous embedding $\mathrm{BV}(\Omega) \subset L^q(\Omega)$, Lemma 2.41, and the Poincaré inequality, Lemma 2.80, on $L^q(\Omega)$. Now let us begin with the proof, which we will divide in 3 steps, namely, showing, that the problem has a finite value, finding a convergent minimizing sequence and proving optimality of the corresponding limit.

Step 1: We show, that the problem has a finite value:

The objective functional is bounded from below by zero. Moreover, we will now show, that we can pick (u, v), such that $F(u, v, y_0) < \infty$, proving that the objective functional is proper. First let us choose $v = 0 = \mathcal{N}(0, 0)$. Therefore, $\mathcal{G}(v) = 0$ as well and $F(u, 0, y_0) = \mathcal{D}(u, y_0) + \lambda_{\mathcal{R}}\mathcal{R}(u)$. In order to obtain a finite value of the objective functional, we only need to find $u \in L^q(\Omega)$, such that $\mathcal{D}(u, y_0) < \infty$ and $\mathcal{R}(u) = J^{**}(u) < \infty$. The second condition is equivalent to asking $u \in \mathrm{BV}(\Omega)$ as stated in Lemma 2.75.

- Linear inverse problem: $\mathcal{D}(u, y_0) = \lambda_{\mathcal{D}}\psi(\|Au - u_0\|)$
  Since $\mathcal{D}(u, u_0) < \infty$ for every $u \in L^q(\Omega)$, any element of $\mathrm{BV}(\Omega)$, e.g., $u = 0$ will serve the purpose.
- Inpainting: $\mathcal{D}(u, y_0) = \mathcal{I}_{\{.=y_0\}}(Au)$
  In this case, we may pick $u = \hat{u_0} \in \mathrm{BV}(\Omega)$, as stated in the theorem.

In both cases, we find that the objective functional is proper, implying, that $(P(y_0))$ has a finite value. Hence, we can pick a minimizing sequence $(u_m, v_m)_m$.

Step 2: Using typical compactness arguments, we show, that there is a convergent minimizing sequence:

Since $\mathcal{G}$ is coercive, as proven in Lemma 2.70, and $\mathcal{D}$ and $\mathcal{R}$ are non-negative, we find, that the sequence $(v_m)_m$ is bounded in $L^q(\Omega)$. Unfortunately, boundedness of $(u_m)_m$ is not as trivial. Lemma 2.80, Poincaré's inequality, implies, that

$$\begin{aligned}\|(u_m - v_m) - (u_m - v_m)_\Omega\|_q &\leq C_P(J^{**}(u_m - v_m) + 1) \leq \\ &\leq C(F(u_m, v_m, y_0) + 1),\end{aligned}$$

where in the constant C, also $\lambda_\mathcal{R}$ is taken into account. Hence,

$$((u_m - v_m) - (u_m - v_m)_\Omega)_m$$

is bounded in $L^q(\Omega)$. Since $(v_m)_m$ is already known to be bounded in $L^q(\Omega)$, also $(v_m - (v_m)_\Omega)_m$ is bounded, because for any $w \in L^q(\Omega)$, we find

$$\begin{aligned}|w_\Omega| = \frac{1}{|\Omega|}\Big|\int\limits_\Omega w\, dx\Big| \underbrace{\leq}_{\text{Hölder}} \frac{1}{|\Omega|}\|w\|_q \|1\|_{q'} = |\Omega|^{\frac{1-q'}{q'}}\|w\|_q \\ \Rightarrow \|v_m - (v_m)_\Omega\|_q \leq \|v_m\|_q (1 + |\Omega|^{\frac{1-q'}{q'}}),\end{aligned}$$

where q' denotes the Hölder conjugate exponent of q. As a result, also $(u_m - (u_m)_\Omega)_m$ is bounded in $L^q(\Omega)$. In the following, we will alter the sequence $(u_m)_m$ in a way, such that the evaluation of the objective functional remains unchanged and the sequence becomes bounded. Denote $U = K(\Omega) \cap \ker(A)$ with $K(\Omega)$ again the finite dimensional space of functions defined on Ω, which are constant a.e.. As a subspace of $K(\Omega)$, U is finite dimensional. Let $U^\perp$ be a complement of U in $K(\Omega)$, and $P : K(\Omega) \to U^\perp$ be the continuous, linear projection onto $U^\perp$. Instead of $(u_m)_m$, we will now consider the sequence

$$w_m := u_m - (I - P)(u_m)_\Omega.$$

Since $(I - P)(u_m)_\Omega \in U = K(\Omega) \cap \ker(A)$, we find, that $Au_m = Aw_m$ and by Lemma 2.79, that $J^{**}(u_m - v_m) = J^{**}(w_m - v_m)$. In other words, $(w_m, v_m)_m$ is a minimizing sequence as well. We will now show boundedness of $(w_m)_m$. Notice, that

$$w_m = u_m - (u_m)_\Omega + P(u_m)_\Omega.$$

We have already shown, that $(u_m - (u_m)_\Omega)_m$ is bounded in $L^q(\Omega)$, therefore, it only remains to prove, that $(P(u_m)_\Omega)_m$ is bounded as well. In order to do so, we introduce the restriction

$$\begin{aligned}A|_{U^\perp} : U^\perp &\to A(U^\perp) \\ u &\mapsto Au.\end{aligned}$$

This operator is actually a bijective operator between finite dimensional spaces. Surjectivity follows from the restriction of the image set. To see injectivity, assume $Au = 0$ for $u \in U^\perp \subset K(\Omega)$. This implies, that $u \in \ker(A) \cap U^\perp =$

$(\ker(A) \cap K(\Omega)) \cap U^\perp = U \cap U^\perp = \{0\}$. Therefore, we can compute

$$\|P(u_m)_\Omega\|_q = \left\|(A|_{U^\perp})^{-1} AP(u_m)_\Omega\right\|_q \leq \left\|(A|_{U^\perp})^{-1}\right\| \|AP(u_m)_\Omega\| =$$
$$= \left\|(A|_{U^\perp})^{-1}\right\| \|A(u_m)_\Omega\| \leq \left\|(A|_{U^\perp})^{-1}\right\| (\|A(u_m - (u_m)_\Omega)\| + \|Au_m\|) \leq$$
$$\leq \left\|(A|_{U^\perp})^{-1}\right\| \left(\|A\| \underbrace{\|(u_m - (u_m)_\Omega)\|_q}_{(i)} + \underbrace{\|Au_m\|}_{(ii)} \right)$$

We have already shown, that (i) is bounded. To show boundedness of (ii) we distinguish the two possible data terms:

- Linear inverse problem: $\mathcal{D}(u, y_0) = \lambda_{\mathcal{D}} \psi(\|Au - y_0\|)$
  Assume to the contrary, (ii) was unbounded, i.e., we could find a subsequence, such that $\|Au_{m_k}\| \xrightarrow{k\to\infty} \infty$ and therefore, also $\|Au_{m_k} - y_0\| \xrightarrow{k\to\infty} \infty$. Then the coercivity of ψ would imply, that $\psi(\|Au_{m_k} - y_0\|) \to \infty$, which contradicts the boundedness of $(F(u_m, v(\mu_m, \theta_m), \mu_m, \theta_m, y_0))_m$, ergo (ii) is bounded as well.
- Inpainting: $\mathcal{D}(u, y_0) = \mathcal{I}_{\{.=y_0\}}(Au)$
  This already implies, that $Au_m = y_0$, and therefore (ii) is bounded

In any case, we find, that $(P(u_m)_\Omega)_m$ is bounded as well. As a result, the sequence $(w_m)_m$ is bounded in $L^q(\Omega)$ and we can therefore extract a subsequence, such that $w_m \rightharpoonup u$ and $v_m \rightharpoonup v$ for some $u, v \in L^q(\Omega)$. We consider as our new minimizing sequence $(w_m, v_m)_m$. What is now left to prove, is, that the obtained limit (u, v) is a solution to $(P(y_0))$.

Step 3: We show optimality of the limit:

We are now exactly in the setting of Lemma 2.82, implying

$$F(u, v, y_0) \leq \liminf_{m\to\infty} F(w_m, v_m, y_0).$$

Since $(w_m, v_m)_m$ is a minimizing sequence, $(F(w_m, v_m, y_0))_m$ converges to the infimum of $(P(y_0))$, proving, that (u, v) is a solution of $(P(y_0))$.

□

The next result shows stability of the proposed method with the data term of a linear inverse problem.

Theorem 2.84. *[Stability of the linear inverse problem] Suppose, that Assumption 2.71 holds true and consider $(P(y_0))$ with $q \leq \frac{d}{d-1}$. Let $\mathcal{D}(u, y_0) = \lambda_{\mathcal{D}} \psi(\|Au - y_0\|)$ (linear* *inverse problem). Further, let $y_0^m, y_0 \in Y$, such that $y_0^m \to y_0$ as $m \to \infty$. Let $(u_m, v_m)_m$ be a sequence of solutions to $(P(y_0))$ with data y_0^m, i.e.,*

$$(u_m, v_m) \in \operatorname*{argmin}_{u,v \in L^q(\Omega)} F(u, v, y_0^m).$$

Then, there exists a subsequence $(u_{m_k}, v_{m_k})_k$ and $u, v \in L^q(\Omega)$, such that $u_{m_k} \rightharpoonup u$ and $v_{m_k} \rightharpoonup v$ in $L^q(\Omega)$, where (u, v) is a solution to $(P(y_0))$ with data y_0, i.e.,

$$F(u, v, y_0) \in \operatorname*{argmin}_{\tilde{u},\tilde{v} \in L^q(\Omega)} F(\tilde{u}, \tilde{v}, y_0).$$

Proof.

Step 1: We show, that the sequence of solutions is bounded and find a convergent subsequence:

Since $(u_m, v_m)_m$ are minimizers of $F(., ., y_0^m)$, we find, that

$$F(u_m, v_m, y_0^m) \le F(0, 0, y_0^m) = \lambda_{\mathcal{D}}\psi(\|y_0^m\|) + \lambda_{\mathcal{R}}\mathcal{R}(0) \le \\ \underbrace{\le}_{\psi \text{ increasing}} \lambda_{\mathcal{D}}\psi(\sup_{l \in \mathbb{N}} \left\|y_0^l\right\|) + \lambda_{\mathcal{R}}\mathcal{R}(0) < \infty,$$

where we used, that $(y_0^m)_m$ is convergent and therefore bounded. As a result, we find, that the sequence of objective functional evaluations $F(u_m, v_m, y_0^m))_m$ is bounded. As before, by the coercivity of $\mathcal{G}$ and non-negativity of $\mathcal{D}$ and $\mathcal{R}$, we find that the sequence $(v_m)_m$ is bounded in $L^q(\Omega)$. By coercivity of ψ, we obtain boundedness of $(Au_m - y_0^m)_m$ and by the fact, that $(y_0^m)_m$ is convergent in Y, boundedness of $(Au_m)_m$. Since $L^q(\Omega)$ is reflexive for $1 < q < \infty$, we can find $v \in L^q(\Omega)$ and a subsequence, again labeled with m, such that $v_m \rightharpoonup v$ in $L^q(\Omega)$. The Poincaré inequality, Lemma 2.80, implies, that

$$\left\|(u_m - v_m) - (u_m - v_m)_\Omega\right\|_q \le C_P(J^{**}(u_m - v_m) + 1) \le \\ \le C(\underbrace{F(u_m, v_m, y_0^m)}_{\text{bounded}} + 1). \tag{2.9}$$

The boundedness of $(v_m)_m$ in $L^q(\Omega)$ implies, that also $(v_m - (v_m)_\Omega)_m$ is bounded in $L^q(\Omega)$ as shown in Theorem 2.83. Together with (2.9), this shows boundedness of $(u_m - (u_m)_\Omega)_m$. In the same way as in Theorem 2.83, we alter our sequence to

$$w_m := u_m - (u_m)_\Omega + P(u_m)_\Omega.$$

Again, for all m, (w_m, v_m) is a solution to $(P(y_0))$ with data y_0^m. In order to show boundedness of $(w_m)_m$, it is sufficient, to show, that $(P(u_m)_\Omega)_m$ is bounded. We use the same notation as in Theorem 2.83 and estimate

$$\|P(u_m)_\Omega\|_q = \left\|(A|_{U^\perp})^{-1}AP(u_m)_\Omega\right\|_q \leq \left\|(A|_{U^\perp})^{-1}\right\| \|AP(u_m)_\Omega\| =$$
$$= \left\|(A|_{U^\perp})^{-1}\right\| \|A(u_m)_\Omega\| \leq \left\|(A|_{U^\perp})^{-1}\right\| (\|A(u_m - (u_m)_\Omega)\| + \|Au_m\|) \leq$$
$$\leq \left\|(A|_{U^\perp})^{-1}\right\| \left(\|A\| \underbrace{\|(u_m - (u_m)_\Omega)\|_q}_{(i)} + \underbrace{\|Au_m\|}_{(ii)} \right)$$

We have already shown both, boundedness of (i) and (ii). As a result also $(w_m)_m$ is bounded in $L^q(\Omega)$. We can therefore extract a further subsequence again labeled with m, such that also $w_m \rightharpoonup u$ in $L^q(\Omega)$ for some $u \in L^q(\Omega)$ and we consider as a new sequence of solutions to $(P(y_0))$ with data y_0^m the sequence $(w_m, v_m)_m$.

Step 2: We show, that the limit of our sequence is a solution to $(P(y_0))$ with data y_0:

As in Theorem 2.83, we use the lower semi-continuity of the objective functional in order to obtain the desired result. To be precise, using Lemma 2.82, we find for $(\tilde{u}, \tilde{v})$ arbitrary,

$$F(u, v, y_0) \underbrace{\leq}_{\text{lower semi-continuity}} \liminf_{m\to\infty} F(w_m, v_m, y_0^m)$$
$$\underbrace{\leq}_{\text{optimality}} \liminf_{m\to\infty} F(\tilde{u}, \tilde{v}, y_0^m) \underbrace{=}_{y_0^m \to y_0} F(\tilde{u}, \tilde{v}, y_0).$$

Thus, (u, v) is optimal, concluding the proof.

□

2.2.6 Explicit Representation of J^{**}

In the following section, we will derive an explicit formula for the functional J^{**} for the examples mentioned in Example 2.72.

Definition 2.85. *[Recession function, [4, 2.5. The integrand f and its recession function]] Let $j : \mathbb{R}^d \to [0, \infty)$ be Borel measurable and convex. Then we define the recession function of j as*

$$j^\infty : \mathbb{R}^d \to \mathbb{R}$$
$$j^\infty(p) = \lim_{t\to\infty} \frac{j(tp)}{t} = \sup_{t>0} \frac{j(tp) - j(0)}{t}.$$

Example 2.86. *As an example and for later use, we will now compute the recession functions of the functions presented in Example 2.72.*

1.
$$j : \mathbb{R}^d \to [0, \infty)$$
$$j(p) = |p|.$$

Then for any $p \in \mathbb{R}^d$,
$$j^\infty(p) = \lim_{t\to\infty} \frac{|tp|}{t} = |p| = j(p).$$
Hence, $j^\infty = j$.

2. *For* $\epsilon > 0$,
$$j : \mathbb{R}^d \to [0, \infty)$$
$$j(p) = \sqrt{|p|^2 + \epsilon}.$$

Then for any $p \in \mathbb{R}^d$,
$$j^\infty(p) = \lim_{t\to\infty} \frac{\sqrt{|tp|^2 + \epsilon}}{t} = \lim_{t\to\infty} \sqrt{|p|^2 + \frac{\epsilon}{t^2}} = |p|.$$

3. *For* $\epsilon > 0$,
$$j : \mathbb{R}^d \to [0, \infty),$$
$$j(p) = \begin{cases} \frac{|p|^2}{2\epsilon} & \text{if } |p| < \epsilon \\ |p| - \frac{\epsilon}{2} & \text{else.} \end{cases}$$

Let $p \in \mathbb{R}^d$, $p \neq 0$. *Then for* t *large enough, we find, that* $|tp| > \epsilon$ *and therefore*
$$j^\infty(p) = \lim_{t\to\infty} \frac{j(tp)}{t} = \lim_{t\to\infty} \frac{|tp| - \frac{\epsilon}{2}}{t} = |p|.$$

For $p = 0$, *we find*
$$j^\infty(0) = \lim_{t\to\infty} \frac{j(t0)}{t} = 0.$$
Therefore, altogether again we find $j^\infty(p) = |p|$.

Theorem 2.87. *[Explicit formula for* J^{**}*, [4, Theorem 4.1]] Let* $j : \mathbb{R}^d \to [0, \infty)$ *be continuous and convex. Assume moreover, there exists* $\Lambda > 0$, *such that for all* $p \in \mathbb{R}^d$
$$j(p) \leq \Lambda(1 + |p|).$$

Let $J : L^1(\Omega) \to [0, \infty]$ *be defined as*
$$J(u) = \begin{cases} \int_\Omega j(\nabla u(x))\, dx & \text{if } u \in W^{1,1}(\Omega) \\ \infty & \text{else.} \end{cases}$$

Then for every $u \in \mathrm{BV}(\Omega)$,

$$\overline{J}(u) = \int\limits_{\Omega} j(\nabla u(x))\, dx + \int\limits_{\Omega} j^{\infty}(\frac{D^c u}{|D^c u|}) d|D^c u| + \int\limits_{J_u \cap \Omega} d\mathcal{H}^{d-1} \int\limits_{u^-(x)}^{u^+(x)} j^{\infty}(\nu_u)\, ds,$$

where $\mathcal{H}^{d-1}$ *denotes the* $(d-1)$ *dimensional Hausdorff measure (see [5, Definition 2.46]), which coincides with the* $(d-1)$ *dimensional Lebesgue measure on 'nice' surfaces/manifolds (see [5, Remark 2.72]).*

Example 2.88. *Let us apply Theorem 2.87 to the examples we presented for* j*. Since for all the examples, we have shown, that* $j^{\infty}(.) = |.|$*, we find for all functions* j *from Example 2.72, that for* $u \in \mathrm{BV}(\Omega)$

$$\begin{aligned}\overline{J}(u) = J^{**}(u) &= \int\limits_{\Omega} j(\nabla u(x))\, dx + \int\limits_{\Omega} |\frac{D^c u}{|D^c u|}|\, d|D^c u| + \int\limits_{J_u \cap \Omega} d\mathcal{H}^{d-1} \int\limits_{u^-(x)}^{u^+(x)} \underbrace{|\nu_u|}_{=1}\, ds = \\ &= \int\limits_{\Omega} j(\nabla u(x))\, dx + \int\limits_{\Omega} \underbrace{|\frac{D^c u}{|D^c u|}|}_{\substack{=1 \\ [5, Corollary 1.29]}} d|D^c u| + \int\limits_{J_u \cap \Omega} d\mathcal{H}^{d-1}(u^+(x) - u^-(x)) = \\ &= \int\limits_{\Omega} j(\nabla u(x))\, dx + |D^c u|(\Omega) + \int\limits_{J_u \cap \Omega} d\mathcal{H}^{d-1}(u^+(x) - u^-(x)) =\end{aligned}$$

Here, $\int\limits_{J_u \cap \Omega} d\mathcal{H}^{d-1}(u^+(x) - u^-(x))$ *can be interpreted as a measure for jumps of* u*, since we integrate* $u^+(x) - u^-(x)$*, which is the height of a jump at the point* $x \in J_u$*.*

Chapter 3

Practical Results

This chapter is dedicated to the investigation of the proposed method in a discrete setting, the end goal being, to apply the method to real images. In this chapter, for $1 \leq p < \infty$ and $z \in \mathbb{R}^d$, we will denote

$$\|z\|_p := \left(\sum_{i=1}^{d} |z_i|^p \right)^{\frac{1}{p}}$$

and

$$\|z\|_\infty := \max_{i=1,2,\dots d} |z_i|.$$

If $z \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_n}$ for $n \in \mathbb{N}$, then $\|z\|_p$ is the norm as above evaluated at z, understood as an element of $\mathbb{R}^{d_1 d_2 \dots d_n}$. Moreover, for a matrix $A \in \mathbb{R}^{d_2 \times d_1}$ (a map $A : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, respectively) and $1 \leq p, q \leq \infty$, we will denote the induced norm as

$$\|A\|_{p,q} := \max_{\substack{z \in \mathbb{R}^{d_1} \\ \|z\|_p \leq 1}} \|Az\|_q .$$

3.1 The Algorithm

In this section, we introduce an algorithm for solving non-convex minimization problems, which will afterwards be applied to an instance of the regularization method proposed in Chapter 2. The presented algorithm is a special case of the Inertial Proximal Alternating Linearized Minimization (iPALM) algorithm [20]. We consider problems of the form

$$\min F(x) := H(x) + \sum_{i=1}^{n} f_i(x_i) \text{ over all } x = (x_1, x_2, ..., x_n) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times ... \times \mathbb{R}^{d_n}, \quad (3.1)$$

where f_i for $i = 1, 2, ..., n$ are extended valued, potentially non-smooth functions and H is a smooth coupling function. Typically, the functions f_i contain constraints and/or regularizing terms, whereas H is a term to be minimized, that arises from the modeling process. Before discussing the method in detail, we have to state some necessary definitions and results, in particular the notion of the subdifferential, which is a generalization of the gradient for non-differentiable functions.

Definition 3.1.

- *Let* $\sigma : \mathbb{R}^d \to (-\infty, \infty]$*. We define the domain of* σ *as*

$$\operatorname{dom}(\sigma) := \left\{ x \in \mathbb{R}^d \;\middle|\; \sigma(x) < \infty \right\}.$$

- *Let* $G : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$*, i.e.,* G *is a function, whose values are subsets of* $\mathbb{R}^d$*, then we define the domain of* G *as*

$$\operatorname{dom}(G) := \left\{ x \in \mathbb{R}^d \;\middle|\; G(x) \neq \emptyset \right\}.$$

Definition 3.2. *[Normal cone, [17, Definition 1.1]] Let* $\emptyset \neq \Omega \subset \mathbb{R}^d$ *and* $\overline{x} \in \Omega$*.*

- *We define for* $\epsilon \geq 0$

$$\hat{N}_\epsilon(\overline{x}; \Omega) := \left\{ y \in \mathbb{R}^d \;\middle|\; \limsup_{\substack{u \in \Omega \\ u \to \overline{x}}} \langle y, \frac{u - \overline{x}}{\|u - \overline{x}\|_2} \rangle \leq \epsilon \right\}.$$

In the case that $\epsilon = 0$*, we denote* $\hat{N}_0(\overline{x}; \Omega) = \hat{N}(\overline{x}; \Omega)$*.*

- *We further define the normal cone to* Ω *at* $\overline{x}$ *as*

$$N(\overline{x}; \Omega) := \left\{ y \in \mathbb{R}^d \;\middle|\; \exists \epsilon_k \downarrow 0,\ (x_k)_k \subset \Omega,\ x_k \to \overline{x},\ y_k \in \hat{N}_{\epsilon_k}(x_k; \Omega) :\ y_k \to y \right\}.$$

Proposition 3.3. *[17, Proposition 1.5] Let* $\Omega \subset \mathbb{R}^d$ *be locally convex at* $x \in \Omega$*, i.e., there exists a neighborhood* U *of* x*, such that* $U \cap \Omega$ *is convex. Then* $\hat{N}(x; \Omega) = N(x; \Omega)$*.*

Definition 3.4. *[Subdifferential, [17, Definition 1.77]] Let* $\sigma : \mathbb{R}^d \to \overline{\mathbb{R}}$*. We define the subdifferential of* σ *at* x *as*

$$\partial\sigma(x) = \left\{ y \in \mathbb{R}^d \;\middle|\; (y, -1) \in N((x, \sigma(x))); \operatorname{epi}(\sigma)) \right\}$$

if $|\sigma(x)| < \infty$ *and* $\partial\sigma(x) = \emptyset$ *else.*

Remark 3.5. *The subdifferential can be interpreted well in simple cases.*

- *If σ is continuously differentiable in a neighborhood of $\overline{x}$, then the subdifferential reduces to the classical gradient, i.e.,*

$$\partial\sigma(\overline{x}) = \{\nabla\sigma(\overline{x})\}\,.$$

For a proof, see [17, Corollary 1.82].

- *If σ is convex, the subdifferential coincides with the well-known subdifferential of convex analysis (see [17, Theorem 1.93]). That is, if $\sigma : \mathbb{R}^d \to \overline{\mathbb{R}}$ is convex and finite at $\overline{x}$, we find that*

$$\partial\sigma(\overline{x}) = \left\{ y \in \mathbb{R}^d \;\middle|\; \forall x \in \mathbb{R}^d :\ \sigma(\overline{x}) + \langle y, x - \overline{x}\rangle \leq \sigma(x) \right\}. \tag{3.2}$$

Note, that in this case, $0 \in \partial\sigma(\overline{x})$ is a sufficient condition for $\overline{x}$ being a minimum of σ. As an example, let $\sigma : \mathbb{R} \to \mathbb{R}$, $\sigma(x) = |x|$. We would like to compute the subdifferential at $x = 0$. By (3.2), it consists of the slopes of all affine functions, which are everywhere less than σ and 0 at $x = 0$. In Figure 3.1 one can observe two of those functions with different slopes. Altogether, we find $\partial\sigma(0) = [-1, 1]$.

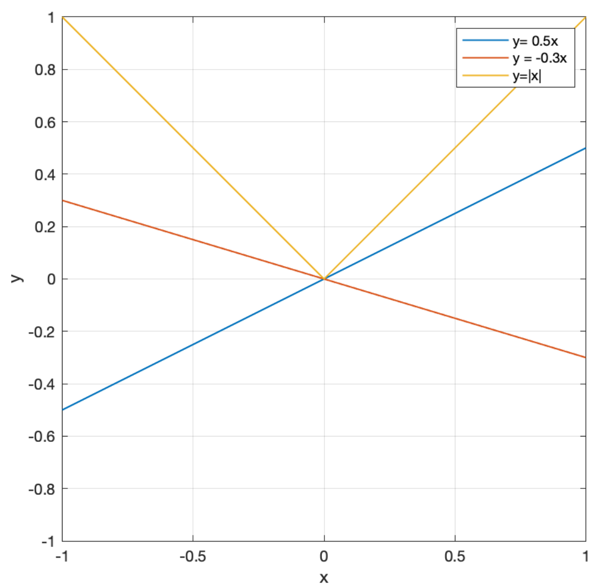


Figure 3.1: Affine minorants of σ.

The proof methodology used in the convergence analysis in [20] is based on the concept of KL functions, which shall be introduced in the following.

Definition 3.6. *[Kurdyka-Łojasiewicz property, KL functions, [7, Definition 3]] For* $\eta \in (0, \infty]$*, we define*

$$\Phi_\eta = \left\{ \phi \in C([0,\eta),\mathbb{R}) \,\middle|\, \phi \text{ is concave, } \phi(0) = 0,\ \phi \in C^1((0,\eta),\mathbb{R}),\ \forall s \in (0,\eta) : \phi'(s) > 0 \right\}.$$

A function $\sigma : \mathbb{R}^d \to (-\infty, \infty]$ *is said to have the Kurdyka-Łojasiewicz (KL) property, at* $\overline{u} \in \mathrm{dom}(\partial\sigma)$*, if there exist* $\eta \in (0, \infty]$*, a neighborhood* U *of* $\overline{u}$ *and a function* $\phi \in \Phi_\eta$*, such that for all*

$$u \in U \cap \left\{ w \in \mathbb{R}^d \,\middle|\, \sigma(\overline{u}) < \sigma(w) < \sigma(\overline{u}) + \eta \right\}$$

the following inequality holds

$$\phi'(\sigma(u) - \sigma(\overline{u}))\, \mathrm{dist}(0, \partial\sigma(u)) \geq 1, \tag{3.3}$$

where for $S \subset \mathbb{R}^d$ *and* $x \in \mathbb{R}^d$*,*

$$\mathrm{dist}(x, S) = \inf \left\{ \|x - y\|_2 \mid y \in S \right\}.$$

If σ *satisfies the KL property at each point of* $\mathrm{dom}(\partial\sigma)$*, then* σ *is called a KL function.*

Remark 3.7. *Let us take a brief moment to interpret the definition of a KL function. In the case, that* σ *is continuously differentiable,* (3.3) *reads as*

$$\phi'(\sigma(u) - \sigma(\overline{u}))\, \|\nabla\sigma(u)\|_2 \geq 1.$$

The KL property is mostly interesting at critical points. Assume for simplicity $\sigma(\overline{u}) = 0$ *and that* $\overline{u}$ *is a critical point, i,e.,* $\nabla\sigma(\overline{u}) = 0$*. Then, the KL property states, that*

$$\phi'(\sigma(u))\, \|\nabla\sigma(u)\|_2 = \|\nabla(\phi \circ \sigma(u))\|_2 \geq 1.$$

This means, that up to the reparametrisation ϕ*, the function* σ *is sharp at the critical point (similar to* $x \mapsto |x|$ *at* $x = 0$*).*

Considering the problem (3.1) again, we aim to find critical points of F, that is, we seek to find $x \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times ... \times \mathbb{R}^{d_n}$, such that $0 \in \partial F(x)$. This is a meaningful generalization of the gradient vanishing. We take the following basic assumptions on the functions involved in (3.1).

Assumption 3.8.

1. *For* $i = 1, 2, ..., n$*,* $f_i : \mathbb{R}^{d_i} \to (-\infty, \infty]$ *is proper (not* ∞ *everywhere), lower semi-continuous, convex and* $\inf_{\mathbb{R}^{d_i}} f_i > -\infty$*.*
2. $H : \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n} \to \mathbb{R}$ *is differentiable and* $\inf_{\mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}} F > -\infty$*.*
3. F *is a KL function.*

4. ∇H *is Lipschitz continuous on bounded sets, i.e., for any bounded subset* $B_1 \times ... \times B_n \subset \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}$*, there exists* $L > 0$*, such that for all* $x, y \in B_1 \times ... \times B_n$*,*

$$\|\nabla H(x) - \nabla H(y)\|_2 \leq L \|x - y\|_2 .$$

Further, let us introduce the notion of the proximal mapping.

Definition 3.9. *[Proximal mapping] Let* $\sigma : \mathbb{R}^d \to (-\infty, \infty]$ *be proper. The proximal mapping of* σ *is defined as*

$$\operatorname{prox}_t^{\sigma}(p) := \operatorname{argmin}\left\{\sigma(q) + \frac{t}{2}\|q-p\|_2^2 \mid q \in \mathbb{R}^d\right\},$$

for $t > 0$*.*

Note, that in general $\operatorname{prox}_t^{\sigma}(p)$ might be the empty set. The following lemma provides a sufficient condition for $\operatorname{prox}_t^{\sigma}(p)$ to contain at least one element.

Lemma 3.10. *Let* $\sigma : \mathbb{R}^d \to (-\infty, \infty]$ *be proper, lower semi-continuous and* $\inf_{\mathbb{R}^d} \sigma > -\infty$*. Then for every* $t > 0$ *and* $p \in \mathbb{R}^d$*,* $\operatorname{prox}_t^{\sigma}(p) \neq \emptyset$*. Moreover, if* σ *is also convex, then* $\operatorname{prox}_t^{\sigma}(p)$ *consists of a single element.*

Proof. Since $\inf_{\mathbb{R}^d} \sigma > -\infty$, also $\inf\left\{\sigma(q) + \frac{t}{2}\|q-p\|_2^2 \mid q \in \mathbb{R}^d\right\} > -\infty$. Let $(q_n)_n \subset \mathbb{R}^d$ be a minimizing sequence, i.e., $\inf\left\{\sigma(q) + \frac{t}{2}\|q-p\|_2^2 \mid q \in \mathbb{R}^d\right\} = \lim_{n\to\infty}\left(\sigma(q_n) + \frac{t}{2}\|q_n - p\|_2^2\right)$. Then $(q_n)_n$ is bounded, because $t > 0$ and

$$\frac{t}{2}\|q_n - p\|_2^2 = \left(\sigma(q_n) + \frac{t}{2}\|q_n - p\|_2^2\right) - \sigma(q_n) \leq \underbrace{\left(\sigma(q_n) + \frac{t}{2}\|q_n - p\|_2^2\right)}_{(i)} - \underbrace{\inf_{\mathbb{R}^d} \sigma}_{(ii)},$$

where (i) is bounded, since $(q_n)_n$ is a minimizing sequence of expression (i), and (ii) is bounded as an assumption of this lemma. Therefore, we can find a convergent subsequence $(q_{n_k})_k$ and $\hat{q} \in \mathbb{R}^d$, such that $q_{n_k} \to \hat{q}$ as $k \to \infty$. Since σ is assumed to be lower semi-continuous, we find, that

$$\begin{aligned}
\sigma(\hat{q}) + \frac{t}{2}\|\hat{q} - p\|_2^2 &\leq \liminf_{k\to\infty} \sigma(q_{n_k}) + \liminf_{k\to\infty} \frac{t}{2}\|q_{n_k} - p\|_2^2 = \\
= \lim_{k\to\infty}\left(\inf_{l\geq k} \sigma(q_{n_l}) + \inf_{l\geq k}\frac{t}{2}\|q_{n_l} - p\|_2^2\right) &\leq \lim_{k\to\infty}\inf_{l\geq k}\left(\sigma(q_{n_l}) + \frac{t}{2}\|q_{n_l} - p\|_2^2\right) = \\
= \liminf_{k\to\infty}\left(\sigma(q_{n_k}) + \frac{t}{2}\|q_{n_k} - p\|_2^2\right) &= \inf\left\{\sigma(q) + \frac{t}{2}\|q - p\|_2^2 \mid q \in \mathbb{R}^d\right\}.
\end{aligned}$$

Therefore, $\hat{q} \in \operatorname{prox}_t^{\sigma}(p)$, which shows, that $\operatorname{prox}_t^{\sigma}(p) \neq \emptyset$.

Now, assume that σ is convex, then $q \mapsto f(q) := \sigma(q) + \frac{t}{2}\|q-p\|_2^2$ is strictly convex, since for $q, \tilde{q} \in \mathbb{R}^d$, $q \neq \tilde{q}$ and $\lambda \in (0,1)$ we find, that

$$
\begin{aligned}
&\lambda f(q) + (1-\lambda) f(\tilde{q}) - f(\lambda q + (1-\lambda)\tilde{q}) = \\
&= \lambda\sigma(q) + \lambda\frac{t}{2}\|q-p\|_2^2 + (1-\lambda)\sigma(\tilde{q}) + (1-\lambda)\frac{t}{2}\|\tilde{q}-p\|_2^2 - \\
&\left(\sigma(\lambda q + (1-\lambda)\tilde{q}) + \frac{t}{2}\|\lambda q + (1-\lambda)\tilde{q} - p\|_2^2\right) \geq \\
&\underbrace{\geq}_{\sigma \text{ convex}} \frac{t}{2}\left(\lambda\|q-p\|_2^2 + (1-\lambda)\|\tilde{q}-p\|_2^2 - \lambda^2\|q-p\|_2^2 - 2\lambda(1-\lambda)\langle q-p, \tilde{q}-p\rangle - (1-\lambda)^2\|\tilde{q}-p\|_2^2\right) = \\
&= \frac{t\lambda(1-\lambda)}{2}\left(\|q-p\|_2^2 + \|\tilde{q}-p\|_2^2 - 2\langle q-p, \tilde{q}-p\rangle\right) = \\
&= \frac{t\lambda(1-\lambda)}{2}\|(q-p) - (\tilde{q}-p)\|_2^2 = \\
&\frac{t\lambda(1-\lambda)}{2}\|q-\tilde{q}\|_2^2 > 0.
\end{aligned}
$$

The strict convexity implies uniqueness of the minimizer, which proves that $\mathrm{prox}_t^{\sigma}(p)$ contains only a single element. ☐

The proposed algorithm reads as follows.

Algorithm 3.11 iPALM

initialize: $x^0 = (x_1^0, x_2^0, ..., x_n^0) \in \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}$
for $m = 1, 2, ...$: **do**
  for $i = 1, 2, ..., n$: **do**
    Take $\alpha_i^m, \beta_i^m \in [0,1]$ and $\tau_i^m > 0$ and compute

$$
\begin{aligned}
y_i^m &= x_i^m + \alpha_i^m (x_i^m - x_i^{m-1}), \\
z_i^m &= x_i^m + \beta_i^m (x_i^m - x_i^{m-1}),
\end{aligned}
$$

$$
x_i^{m+1} \in \mathrm{prox}_{\tau_i^m}^{f_i}\left(y_i^m - \frac{1}{\tau_i^m}\nabla_{x_i} H(x_1^{m+1}, ..., x_{i-i}^{m+1}, z_i^m, x_{i+1}^m, ..., x_n^m)\right). \tag{3.4}
$$

  end for
end for

The proof of convergence of the sequence, generated by Algorithm 3.11 can be found in [20]. Therefore, we will not present most of the details, but the following two results, which play a crucial role in the proof.

Lemma 3.12. *[Descent lemma] Let $h : \mathbb{R}^d \to \mathbb{R}$ be differentiable and $u, v \in \mathbb{R}^d$. Further assume that ∇h is Lipschitz continuous with Lipschitz constant L on the line connecting u and v, given as $\{tu + (1-t)v \mid t \in [0,1]\}$. Then*

$$h(v) \leq h(u) + \langle v - u, \nabla h(u) \rangle + \frac{L}{2} \|u - v\|_2^2 .$$

Proof. Using the fundamental theorem of calculus, we obtain

$$\begin{aligned} h(v) - h(u) &= \int_0^1 \frac{d}{dt} h(u + t(v-u)) dt = \int_0^1 \langle v - u, \nabla h(u + t(v-u)) \rangle dt = \\ &= \int_0^1 \langle v - u, \nabla h(u + t(v-u)) - \nabla h(u) \rangle dt + \langle v - u, \nabla h(u) \rangle \leq \\ &\leq \int_0^1 |\langle v - u, \nabla h(u + t(v-u)) - \nabla h(u) \rangle| dt + \langle v - u, \nabla h(u) \rangle \leq \\ &\underbrace{\leq}_{\text{Cauchy-Schwartz}} \int_0^1 \|v-u\|_2 \|\nabla h(u + t(v-u)) - \nabla h(u)\|_2 \, dt + \langle v - u, \nabla h(u) \rangle \leq \\ &\underbrace{\leq}_{\nabla h \text{ Lipsachitz}} \int_0^1 \|v-u\|_2 \, Lt \, \|v-u\|_2 \, dt + \langle v - u, \nabla h(u) \rangle = \frac{L}{2} \|v-u\|_2^2 + \langle v - u, \nabla h(u) \rangle . \end{aligned}$$

□

Lemma 3.13. *[Proximal inequality] Let $h : \mathbb{R}^d \to \mathbb{R}$ be differentiable with ∇h Lipschitz continuous with Lipschitz constant L and let $\sigma : \mathbb{R}^d \to (-\infty, \infty]$ be proper, lower semi-continuous, convex and $\inf_{\mathbb{R}^d} \sigma > -\infty$. Further, let $v, w \in \mathbb{R}^d$, $t > 0$ and*

$$u^+ = \operatorname{prox}_t^\sigma (v - \frac{1}{t} \nabla h(w)).$$

Then for every $u \in dom(\sigma)$ and $s > 0$, we have

$$\begin{gathered} h(u^+) + \sigma(u^+) \leq h(u) + \sigma(u) + \\ + \frac{L + s - t}{2} \left\| u^+ - u \right\|_2^2 + \frac{t}{2} \|v - u\|_2^2 - \frac{t}{2} \left\| v - u^+ \right\|_2^2 + \frac{1}{2s} L^2 \|u - w\|_2^2 . \end{gathered}$$

Proof. First, note that we can write the proximal mapping in a more convenient way.

$$\begin{aligned} \operatorname{prox}_t^\sigma (v - \frac{1}{t} \nabla h(w)) &= \operatorname{argmin} \left\{ \sigma(\xi) + \frac{t}{2} \left\| v - \frac{1}{t} \nabla h(w) - \xi \right\|_2^2 \mid \xi \in \mathbb{R}^d \right\} = \\ = \operatorname{argmin} &\left\{ \sigma(\xi) + \frac{t}{2} \left(\|v - \xi\|_2^2 - 2 \langle v - \xi, \frac{1}{t} \nabla h(w) \rangle + \left\| \frac{1}{t} \nabla h(w) \right\|_2^2 \right) \mid \xi \in \mathbb{R}^d \right\} = \\ &= \operatorname{argmin} \left\{ \underbrace{\sigma(\xi) + \frac{t}{2} \|v - \xi\|_2^2 - \langle v - \xi, \nabla h(w) \rangle}_{=: f(\xi)} \mid \xi \in \mathbb{R}^d \right\} . \end{aligned} \tag{3.5}$$

The function f to be minimized is convex, since σ is assumed to be convex. In fact, f is even strictly convex, since for $\eta, \xi \in \mathbb{R}^d$ and $\lambda \in (0,1)$, similar to the proof of Lemma 3.10 we find

$$\begin{aligned}
&\lambda f(\xi) + (1-\lambda) f(\eta) - f(\lambda\xi + (1-\lambda)\eta) \geq \\
&\underbrace{\geq}_{\sigma \text{ convex}} \frac{t}{2}\Big(\lambda \|v-\xi\|_2^2 + (1-\lambda)\|v-\eta\|_2^2 - \|v - \lambda\xi - (1-\lambda)\eta\|_2^2\Big) = \\
&= \frac{t}{2}\Big(\lambda \|v-\xi\|_2^2 + (1-\lambda)\|v-\eta\|_2^2 - \lambda^2\|v-\xi\|_2^2 - 2\lambda(1-\lambda)\langle v-\xi, v-\eta\rangle - (1-\lambda)^2\|v-\eta\|_2^2\Big) = \\
&= \frac{t}{2}\Big(\lambda(1-\lambda)\|v-\xi\|_2^2 + (1-\lambda)(1-(1-\lambda))\|v-\eta\|_2^2 - 2\lambda(1-\lambda)\langle v-\xi, v-\eta\rangle\Big) = \\
&= \frac{t\lambda(1-\lambda)}{2}\Big(\|v-\xi\|_2^2 + \|v-\eta\|_2^2 - 2\langle v-\xi, v-\eta\rangle\Big) = \frac{t\lambda(1-\lambda)}{2}\|(v-\xi)-(v-\eta)\|_2^2 = \\
&= \frac{t\lambda(1-\lambda)}{2}\|\xi-\eta\|_2^2 .
\end{aligned} \tag{3.6}$$

We are now able, to quantify the growth of the function f at its minimum. With $u^+ \in \mathbb{R}^d$, the unique minimizer of f, $u \in \mathbb{R}^d$ arbitrary and $\lambda \in (0,1)$, by setting $\xi = u$ and $\eta = u^+$ in (3.6), we obtain

$$\begin{aligned}
& f(\lambda u + (1-\lambda)u^+) + \frac{t\lambda(1-\lambda)}{2}\left\|u^+ - u\right\|_2^2 \leq \lambda f(u) + (1-\lambda) f(u^+) \\
\Rightarrow & \underbrace{\frac{f(\lambda u + (1-\lambda)u^+) - f(u^+)}{\lambda}}_{\geq 0} + \frac{t(1-\lambda)}{2}\left\|u^+ - u\right\|_2^2 \leq f(u) - f(u^+) \\
\Rightarrow & \frac{t(1-\lambda)}{2}\left\|u^+ - u\right\|_2^2 \leq f(u) - f(u^+).
\end{aligned}$$

Since this is true for all $\lambda \in (0,1)$, letting $\lambda \to 0^+$ yields

$$\frac{t}{2}\left\|u^+ - u\right\|_2^2 \leq f(u) - f(u^+).$$

By plugging in the definition of f, we obtain

$$\begin{aligned}
& \sigma(u^+) + \frac{t}{2}\left\|v - u^+\right\|_2^2 - \langle v - u^+, \nabla h(w)\rangle + \frac{t}{2}\left\|u^+ - u\right\|_2^2 \leq \sigma(u) + \frac{t}{2}\|v-u\|_2^2 - \langle v-u, \nabla h(w)\rangle \\
\Rightarrow & \sigma(u^+) \leq \sigma(u) + \frac{t}{2}\|v-u\|_2^2 - \langle v-u, \nabla h(w)\rangle - \frac{t}{2}\left\|v-u^+\right\|_2^2 + \langle v-u^+, \nabla h(w)\rangle - \frac{t}{2}\left\|u^+ - u\right\|_2^2 .
\end{aligned}$$

Combining this inequality with Lemma 3.12, we obtain

$$\begin{aligned} h(u^+) + \sigma(u^+) \le h(u) + \langle u^+ - u, \nabla h(u)\rangle + \frac{L}{2}\left\|u^+ - u\right\|_2^2 + \\ \sigma(u) + \frac{t}{2}\|v-u\|_2^2 - \langle v-u, \nabla h(w)\rangle - \frac{t}{2}\left\|v-u^+\right\|_2^2 + \langle v-u^+, \nabla h(w)\rangle - \frac{t}{2}\left\|u^+-u\right\|_2^2 = \\ = h(u) + \sigma(u) + \langle u^+ - u, \nabla h(u) - \nabla h(w)\rangle + \frac{L-t}{2}\left\|u^+-u\right\|_2^2 + \frac{t}{2}\|v-u\|_2^2 - \frac{t}{2}\left\|v-u^+\right\|_2^2. \end{aligned} \tag{3.7}$$

We note, that for $p, q \in \mathbb{R}^d$ and any $s > 0$, we have

$$\begin{aligned} 0 \le \left\|\sqrt{s}p - \frac{1}{\sqrt{s}}q\right\|_2^2 = s\|p\|_2^2 - 2\langle p, q\rangle + \frac{1}{s}\|q\|_2^2 \\ \Rightarrow \langle p, q\rangle \le \frac{s}{2}\|p\|_2^2 + \frac{1}{2s}\|q\|_2^2. \end{aligned}$$

Therefore, we can compute

$$\begin{aligned} \langle u^+ - u, \nabla h(u) - \nabla h(w)\rangle \le \frac{s}{2}\left\|u^+-u\right\|_2^2 + \frac{1}{2s}\|\nabla h(u) - \nabla h(w)\|_2^2 \underbrace{\le}_{\nabla h \text{ Lipschitz}} \\ \le \frac{s}{2}\left\|u^+-u\right\|_2^2 + \frac{L^2}{2s}\|u-w\|_2^2. \end{aligned}$$

Plugging this into (3.7) gives the desired result. □

Remark 3.14. *Carefully reviewing the proof of Lemma 3.13, we notice, that the assumptions on h may be weakened. To be precise, we do not need global Lipschitz continuity of ∇h. Indeed, it is sufficient, that only in the points u, u^+, w the following conditions are satisfied:*

(i) Descent property:

$$h(u^+) \le h(u) + \langle u^+ - u, \nabla h(u)\rangle + \frac{L}{2}\left\|u^+-u\right\|_2^2$$

(ii) Lipschitz property:

$$\|\nabla h(u) - \nabla h(w)\|_2 \le L\|u-w\|_2$$

In the proof of convergence of Algorithm 3.11 presented in [20], Lemma 3.13 is applied at the step (3.4), where the old and new iterates will take the place of u and u^+. Combined with the presented results in [20], this motivates the following assumption, which also covers the choice of the parameters α_i^m, β_i^m and τ_i^m in Algorithm 3.11.

Assumption 3.15. *Using the notations of Algorithm 3.11, we make the following assumptions:*

1. *There exists* $\epsilon \in (0,1)$*, such that for all* $i = 1, 2, ..., n$ *there are* $\overline{\alpha_i}$*,* $\overline{\beta_i} \in (0, 1-\epsilon)$*, such that for all* $m \in \mathbb{N}$*,* $0 \leq \alpha_i^m \leq \overline{\alpha_i}$ *and* $0 \leq \beta_i^m \leq \overline{\beta_i}$*.*

2. *Using* ϵ *from item 1 above, for all* $i = 1, 2, ..., n$*,*

$$\delta_i = \frac{\overline{\alpha_i} + 2\overline{\beta_i}}{2(1-\epsilon-\overline{\alpha_i})} L_i,$$
$$\tau_i^m = \frac{(1+\epsilon)\delta_i + (1+\beta_i^m)L_i^m}{2-\alpha_i^m},$$

where $0 < L_i^m \leq L_i$ *for all* m *and* L_i^m *satisfies the following properties:*

i) Descent property:

$$H(x_1^{m+1}, ..., x_{i-i}^{m+1}, x_i^{m+1}, x_{i+1}^m, ..., x_n^m) \leq$$
$$\leq H(x_1^{m+1}, ..., x_{i-i}^{m+1}, x_i^m, x_{i+1}^m, ..., x_n^m) +$$
$$+\langle x_i^{m+1} - x_i^m, \nabla_{x_i} H(x_1^{m+1}, ..., x_{i-i}^{m+1}, x_i^m, x_{i+1}^m, ..., x_n^m)\rangle + \frac{L_i^m}{2} \left\| x_i^{m+1} - x_i^m \right\|_2^2$$

ii) Lipschitz property:

$$\left\| \nabla_{x_i} H(x_1^{m+1}, ..., x_{i-i}^{m+1}, z_i^m, x_{i+1}^m, ..., x_n^m) - \nabla_{x_i} H(x_1^{m+1}, ..., x_{i-i}^{m+1}, x_i^m, x_{i+1}^m, ..., x_n^m) \right\|_2 \leq$$
$$\leq L_i^m \left\| z_i^m - x_i^m \right\|_2$$

Remark 3.16. *If the sequence* $(x^m)_m$*, generated by Algorithm 3.11, is bounded, it is possible, to find* L_i^m *satisfying the descent and the Lipschitz property. In order to see this, let* $B = B_1 \times ... \times B_n$ *be a bounded, convex set, such that* $(x_i^m)_m, (y_i^m)_m, (z_i^m)_m \subset B_i$ *for all* i*. Then, by Assumption 3.8,* $\nabla_{x_i} H$ *is Lipschitz continuous on* B*. If now* L_i^m *is a Lipschitz constant for* $\nabla_{x_i} H$ *on* B*, it fulfills both properties (see Lemma 3.12).*

This enables us, to present the main result of this section.

Theorem 3.17. *Suppose, that Assumption 3.8 and Assumption 3.15 hold true. Then, if the sequence* $(x^m)_m$*, generated by Algorithm 3.11, is bounded, it converges to a critical point of* F*.*

Proof. The proof is done analogously to the one of [20, Theorem 4.1] and we only sketch the necessary modifications to cover the setting considered here:

Firstly, we should note, that the results in [20] are presented for the case $n = 2$, but can be extended to the general case without significant modifications. As stated in Remark 3.14, the descent and Lipschitz property in Assumption 3.15 enable us, to prove and use the proximal inequality, Lemma 3.13, at all iterations in Algorithm 3.11. Using this, all proofs in [20] can be carried through with the assumptions we have made, eventually leading to the desired result. □

3.2 The Problem in a Discrete Setting

3.2.1 Preliminaries

In the following we will introduce some notations and present several results, which will be used afterwards.

Definition 3.18. *[Discrete convolution] Let $\mu \in \mathbb{R}^{n\times m}$ and $\theta \in \mathbb{R}^{r\times s}$, such that $r \leq n$ and $s \leq m$. We define the discrete convolution $\mu * \theta \in \mathbb{R}^{(n-r+1)\times(m-s+1)}$ as*

$$(\mu * \theta)_{i,j} \coloneqq \sum_{\substack{k=1,2,...,r\\ l=1,2,...,s}} \mu_{i+r-k,j+s-l}\theta_{k,l}$$

for $i = 1, 2, ..., n - r + 1$ and $j = 1, 2, ...m - s + 1$.

Remark 3.19. *Most readers may be familiar with the convolution. The distinctive property of the convolution in Definition 3.18 is just, that we compute exactly those values, where no boundary extension of μ is needed. For θ, effectively we use a zero extension.*

Definition 3.20. *[Strided upconvolution] Let $\mu \in \mathbb{R}^{n\times m}$, $\theta \in \mathbb{R}^{r\times s}$ and $\sigma \in \mathbb{N}$. Further, let $\tilde{n}, \tilde{m} \in \mathbb{N}$, such that $\sigma(n-1)+1 \leq \tilde{n} \leq \sigma n$ and $\sigma(m-1)+1 \leq \tilde{m} \leq \sigma m$. We define the strided upconvolution $\mu *_\sigma \theta$ as*

$$\mu *_\sigma \theta \coloneqq \tilde{\mu}^\sigma * \theta,$$

where the zero interpolation $\tilde{\mu}^\sigma \in \mathbb{R}^{\tilde{n}\times\tilde{m}}$ is defined as

$$\tilde{\mu}^\sigma_{i,j} \coloneqq \begin{cases} \mu_{k,l} & \text{if } \exists k \in \{1,2,...,n\},\ l \in \{1,2,...,m\}: \\ & i = \sigma*(k-1)+1, j = \sigma*(l-1)+1\ . \\ 0 & \text{else} \end{cases}$$

Remark 3.21.

- *The strided upconvolution can also be interpreted as the adjoint operator of a down-sampled convolution. In order to see this, let us compute the adjoint of*

$$\begin{aligned} *_\sigma\theta : \mathbb{R}^{n\times m} &\to \mathbb{R}^{N\times M} \\ \mu &\mapsto \mu *_\sigma \theta \end{aligned}$$

  (N, M can be computed according to the definitions above). Let $\mu \in \mathbb{R}^{n\times m}$ and $u \in \mathbb{R}^{N\times M}$ be arbitrary. In the following, a tilde denotes the zero extension, i.e., for $i, j \in \mathbb{Z}$

$$\tilde{\theta}_{i,j} = \begin{cases} \theta_{i,j} & \text{if } 1 \leq i \leq r,\ 1 \leq j \leq s \\ 0 & \text{else.} \end{cases}$$

We compute

$$\langle \mu *_\sigma \theta, u\rangle = \sum_{\substack{i=1,2,..,N\\ j=1,2,..,M}} (\mu *_\sigma \theta)_{i,j} u_{i,j} = \sum_{\substack{i=1,2,..,N\\ j=1,2,..,M}} u_{i,j} \sum_{\substack{k=1,2,..,r\\ l=1,2,..,s}} \tilde{\mu}^\sigma_{i+r-k,j+s-l}\theta_{k,l} =$$

$$= \sum_{\substack{i=1,2,..,N\\ j=1,2,..,M}} u_{i,j} \sum_{\substack{k'=i,i+1,..,i+r-1\\ l'=j,j+1,..,j+s-1}} \tilde{\mu}^\sigma_{k',l'}\theta_{i+r-k',j+s-l'} =$$

$$= \sum_{\substack{i=1,2,..,N\\ j=1,2,..,M}} u_{i,j} \sum_{\substack{k'=1,2,..,\tilde{n}\\ l'=1,2,..,\tilde{m}}} \tilde{\mu}^\sigma_{k',l'}\tilde{\theta}_{i+r-k',j+s-l'} =$$

$$= \sum_{\substack{k'=1,2,..,\tilde{n}\\ l'=1,2,..,\tilde{m}}} \tilde{\mu}^\sigma_{k',l'} \sum_{\substack{i=1,2,..,N\\ j=1,2,..,M}} u_{i,j}\tilde{\theta}_{i+r-k',j+s-l'} =$$

$$= \sum_{\substack{k=1,2,..,n\\ l=1,2,..,m}} \mu_{k,l} \sum_{\substack{i=1,2,..,N\\ j=1,2,..,M}} u_{i,j}\tilde{\theta}_{i+r-\sigma(k-1)-1,j+s-\sigma(l-1)-1}.$$

Hence, we see, that the adjoint of $*_\sigma\theta$ *is a convolution operation with the convolution kernel* $(\theta_{-i,-j})_{i,j}$ *(up to a shift) evaluated at every* σ*-th entry.*

- *Note that, there is some wiggle room to choose the size of* $\tilde{\mu}$ *without losing any information about* μ*. This will be relevant later, when we are given an image size and want to fit the architecture of a convolutional network accordingly, such that we can generate images of the given size.*
- *As the name suggests, the strided upconvolution is a method of up sampling. Figuratively speaking,* $\tilde{\mu}$ *is the result of adding zeros in between the values of* μ*, such that only every* σ^{th} *entry of* $\tilde{\mu}$ *is non-zero. This increases the size of* μ *approximately by the factor* σ*. Afterwards, we perform the regular convolution of Definition 3.18 on* $\tilde{\mu}$*.*
- *If* $\sigma = 1$*, we recover the regular convolution.*
- *We will refer to* θ *as the (filter) kernel,* μ *as coefficient and* σ *as the stride of the convolution.*

Definition 3.22. *Let* $N, M, m_f, \sigma \in \mathbb{N}$, $\mu \in \mathbb{R}^{n\times m\times n_f}$, $\theta_l \in \mathbb{R}^{s\times s\times n_f}$ *(we use square filter kernels now, to simplify writing) for* $l = 1, 2, ..., m_f$*, such that* $N = \tilde{n} - s + 1$ *and* $M = \tilde{m} - s + 1$ *with* $\sigma(n-1) + 1 \leq \tilde{n} \leq \sigma n$ *and* $\sigma(m-1) + 1 \leq \tilde{m} \leq \sigma m$ *(compare to Definition 3.20 and Remark 3.21). We define*

$$K^\sigma_\theta : \mathbb{R}^{n\times m\times n_f} \to \mathbb{R}^{N\times M\times m_f}$$

$$p = (p_k)_{k=1}^{n_f} \mapsto \left(\sum_{k=1}^{n_f} p_k *_\sigma (\theta_l)_k\right)_{l=1}^{m_f}$$

and

$$\begin{aligned} C^\sigma_\mu : \mathbb{R}^{s\times s\times n_f} &\to \mathbb{R}^{N\times M} \\ D = (D_k)_{k=1}^{n_f} &\mapsto \sum_{k=1}^{n_f} \mu_k *_\sigma D_k. \end{aligned}$$

Lemma 3.23. *With the notation of Definition 3.22 it holds true, that*

$$\|K^\sigma_\theta\|_{2,2} \leq \sqrt{\sum_{l=1}^{m_f}\sum_{k=1}^{n_f} \|(\theta_l)_k\|_1^2}$$

and

$$\left\|C^\sigma_\mu\right\|_{2,2} \leq \sqrt{\sum_{k=1}^{n_f} \|\mu_k\|_1^2}.$$

Proof. Let $p \in \mathbb{R}^{n\times m\times n_f}$ be arbitrary. We compute

$$\begin{aligned} \|p_k *_\sigma (\theta_l)_k\|_2^2 = \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} \left|(\tilde{p_k}^\sigma * (\theta_l)_k)_{i,j}\right|^2 = \sum_{\substack{i=s+1,\dots,\tilde{n}+1\\ j=s+1,\dots,\tilde{m}+1}} \left|\sum_{r,t=1,\dots,s} (\tilde{p_k}^\sigma)_{i-r,j-t}((\theta_l)_k)_{r,t}\right|^2 \leq \\ \leq \sum_{i,j}\left(\sum_{r,t} |(\tilde{p_k}^\sigma)_{i-r,j-t}| \sqrt{|((\theta_l)_k)_{r,t}|}\sqrt{|((\theta_l)_k)_{r,t}|}\right)^2 \underbrace{\leq}_{\text{Hölder}} \\ \leq \sum_{i,j}\left(\sum_{r,t} |(\tilde{p_k}^\sigma)_{i-r,j-t}|^2 \, |((\theta_l)_k)_{r,t}|\right)\left(\sum_{r,t} |((\theta_l)_k)_{r,t}|\right) \leq \\ \leq \|\theta_l\|_1^2 \, \|\tilde{p_k}^\sigma\|_2^2 = \|\theta_l\|_1^2 \, \|p_k\|_2^2\,. \end{aligned} \tag{3.8}$$

Therefore,

$$\left\|\sum_{k=1}^{n_f} p_k *_\sigma (\theta_l)_k\right\|_2 \leq \sum_{k=1}^{n_f} \|p_k *_\sigma (\theta_l)_k\|_2 \leq \sum_{k=1}^{n_f} \|(\theta_l)_k\|_1 \|p_k\|_2 \underbrace{\leq}_{\text{Hölder}} \|p\|_2 \sqrt{\sum_{k=1}^{n_f} \|(\theta_l)_k\|_1^2}.$$

Altogether, we find

$$\|K^\sigma_\theta(p)\|_2 = \sqrt{\sum_{l=1}^{m_f}\left\|\sum_{k=1}^{n_f} p_k * (\theta_l)_k\right\|_2^2} \leq \sqrt{\sum_{l=1}^{m_f}\sum_{k=1}^{n_f} \|(\theta_l)_k\|_1^2 \, \|p\|_2^2} = \|p\|_2 \sqrt{\sum_{l=1}^{m_f}\sum_{k=1}^{n_f} \|(\theta_l)_k\|_1^2}.$$

Analogously to (3.8), we obtain for arbitrary $D \in \mathbb{R}^{s\times s\times n_f}$

$$\begin{aligned}\|\mu_k *_\sigma D_k\|_2^2 &= \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} |(\tilde{\mu_k}^\sigma * D_k)_{i,j}|^2 = \sum_{\substack{i=s+1,\dots,\tilde{n}+1\\ j=s+1,\dots,\tilde{m}+1}} \left| \sum_{r,t=1,\dots,s} (\tilde{\mu_k}^\sigma)_{i-r,j-t} (D_k)_{r,t} \right|^2 \leq \\ &\leq \sum_{i,j} \left(\sum_{r,t} \sqrt{|(\tilde{\mu_k}^\sigma)_{i-r,j-t}|} \sqrt{|(\tilde{\mu_k}^\sigma)_{i-r,j-t}|}\, |(D_k)_{r,t}| \right)^2 \underbrace{\leq}_{\text{Hölder}} \\ &\leq \sum_{i,j} \left(\sum_{r,t} |(\tilde{\mu_k}^\sigma)_{i-r,j-t}|\, |(D_k)_{r,t}|^2 \right) \left(\sum_{r,t} |(\tilde{\mu_k}^\sigma)_{i-r,j-t}| \right) \leq \\ &\leq \|\tilde{\mu_k}^\sigma\|_1^2 \, \|D_k\|_2^2 = \|\mu_k\|_1^2 \, \|D_k\|_2^2 \,. \end{aligned} \tag{3.9}$$

This enables us to compute

$$\left\| C_\mu^\sigma(D) \right\|_2 = \left\| \sum_{k=1}^{n_f} \mu_k *_\sigma D_k \right\|_2 \leq \sum_{k=1}^{n_f} \|\mu_k *_\sigma D_k\| \leq \sum_{k=1}^{n_f} \|\mu_k\|_1 \, \|D_k\|_2 \underbrace{\leq}_{\text{Hölder}} \|D\|_2 \sqrt{\sum_{k=1}^{n_f} \|\mu_k\|_1^2}, \tag{3.10}$$

which concludes the proof. □

Definition 3.24. *[Discrete gradient] We define the discrete gradient operator as*

$$\begin{aligned} \nabla : \mathbb{R}^{N\times M} &\to \mathbb{R}^{N\times M\times 2} \\ u &\mapsto \nabla u, \end{aligned}$$

with

$$(\nabla u)_{i,j}^1 := \begin{cases} u_{i+1,j} - u_{i,j} & \text{for } i < N \\ 0 & \text{for } i = N, \end{cases}$$

$$(\nabla u)_{i,j}^2 := \begin{cases} u_{i,j+1} - u_{i,j} & \text{for } j < M \\ 0 & \text{for } j = M. \end{cases}$$

Lemma 3.25. *With the notation of Definition 3.24, it holds true, that*

$$\|\nabla\|_{2,2} \leq \sqrt{8}.$$

Proof. Using Young's inequality, we find for $a, b \in \mathbb{R}$,

$$(a+b)^2 = a^2 + b^2 + \underbrace{2ab}_{\leq a^2+b^2} \leq 2\left(a^2+b^2\right).$$

Deploying this elementary inequality, we compute

$$\begin{aligned}
\|\nabla u\|_2^2 = &\sum_{\substack{i=1,\dots,N-1\\ j=1,\dots,M-1}} \left[(u_{i+1,j}-u_{i,j})^2 + (u_{i,j+1}-u_{i,j})^2\right] + \\
&+ \sum_{i=1,\dots,N-1} (u_{i+1,M}-u_{i,M})^2 + \sum_{j=1,\dots,M-1} (u_{N,j+1}-u_{N,j})^2 \leq \\
&\leq 2 \sum_{\substack{i=1,\dots,N-1\\ j=1,\dots,M-1}} \left[u_{i+1,j}^2 + u_{i,j}^2 + u_{i,j+1}^2 + u_{i,j}^2\right] + \\
&+2 \sum_{i=1,\dots,N-1} \left(u_{i+1,M}^2 + u_{i,M}^2\right) + 2 \sum_{j=1,\dots,M-1} \left(u_{N,j+1}^2 + u_{N,j}^2\right) = \\
&= 4 \sum_{\substack{i=1,\dots,N-1\\ j=1,\dots,M-1}} u_{i,j}^2 + \\
&+2 \sum_{\substack{i=1,\dots,N-1\\ j=1,\dots,M}} u_{i+1,j}^2 + 2 \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M-1}} u_{i,j+1}^2 + 2 \sum_{i=1,\dots,N-1} u_{i,M}^2 + 2 \sum_{j=1,\dots,M-1} u_{N,j}^2 \leq \\
&\leq 4 \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} u_{i,j}^2 + 2 \sum_{\substack{i=1,\dots,N-1\\ j=1,\dots,M}} u_{i+1,j}^2 + 2 \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M-1}} u_{i,j+1}^2 \leq \\
&\leq 8 \|u\|_2^2,
\end{aligned}$$

which proves, that $\|\nabla\|_{2,2} \leq \sqrt{8}$. □

Definition 3.26. *[Discrete divergence] We define the discrete divergence operator as*

$$\begin{aligned}
\operatorname{div} : \mathbb{R}^{N\times M\times 2} &\to \mathbb{R}^{N\times M}\\
p &\mapsto \operatorname{div}(p),
\end{aligned}$$

with, for $1 \leq i \leq N$, $1 \leq j \leq M$,

$$(\operatorname{div} p)_{i,j} \coloneqq \begin{cases} p_{i,j}^1 - p_{i-1,j}^1 + p_{i,j}^2 - p_{i,j-1}^2 & \text{for } i < N, j < M \\ p_{i,M}^1 - p_{i-1,M}^1 - p_{i,M-1}^2 & \text{for } i < N, j = M \\ -p_{N-1,j}^1 + p_{N,j}^2 - p_{N,j-1}^2 & \text{for } i = N, j < M \\ -p_{N-1,M}^1 - p_{N,M-1}^2 & \text{for } i = N, j = M, \end{cases}$$

where we use the convention $p_{i,j}^k = 0$ if $i = 0$ or $j = 0$.

Lemma 3.27. *Using the scalar product*

$$\langle p, q\rangle = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} p_{i,j}^k q_{i,j}^k$$

on $\mathbb{R}^{N\times M\times L}$, it holds true, that

$$\nabla^* = -\operatorname{div}.$$

Proof. The proof is done by elementary computations. Let $u \in \mathbb{R}^{N\times M}$ and $p \in \mathbb{R}^{N\times M\times 2}$ be arbitrary. Using a zero extension of p, i.e., $p^k_{i,j} = 0$ if $(i,j) \notin \{1,2,...N\} \times \{1,2,...M\}$, we compute

$$\begin{aligned}
\langle \nabla u, p\rangle &= \sum_{i=1}^{N}\sum_{j=1}^{M} (\nabla u)^1_{i,j} p^1_{i,j} + (\nabla u)^2_{i,j} p^2_{i,j} = \sum_{i=1}^{N-1}\sum_{j=1}^{M} (u_{i+1,j} - u_{i,j}) p^1_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{M-1} (u_{i,j+1} - u_{i,j}) p^2_{i,j} = \\
&= \sum_{i=1}^{N-1}\sum_{j=1}^{M} -u_{i,j} p^1_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{M} u_{i,j} p^1_{i-1,j} + \sum_{i=1}^{N}\sum_{j=1}^{M-1} -u_{i,j} p^2_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{M} u_{i,j} p^2_{i,j-1} = \\
&= \sum_{i=1}^{N-1}\sum_{j=1}^{M} -u_{i,j}(p^1_{i,j} - p^1_{i-1,j}) + \sum_{j=1}^{M} u_{N,j} p^1_{N-1,j} + \sum_{i=1}^{N}\sum_{j=1}^{M-1} -u_{i,j}(p^2_{i,j} - p^2_{i,j-1}) + \sum_{i=1}^{N} u_{i,M} p^2_{i,M-1} = \\
&= \sum_{i=1}^{N-1}\sum_{j=1}^{M-1} -u_{i,j}(p^1_{i,j} - p^1_{i-1,j} + p^2_{i,j} - p^2_{i,j-1}) + \\
&+ \sum_{i=1}^{N-1} -u_{i,M}(p^1_{i,M} - p^1_{i-1,M}) + \sum_{i=1}^{N} u_{i,M} p^2_{i,M-1} + \sum_{j=1}^{M-1} -u_{N,j}(p^2_{N,j} - p^2_{N,j-1}) + \sum_{j=1}^{M} u_{N,j} p^1_{N-1,j} = \\
&= \sum_{i=1}^{N-1}\sum_{j=1}^{M-1} -u_{i,j}\operatorname{div}(p)_{i,j} + \sum_{i=1}^{N-1} -u_{i,M}(p^1_{i,M} - p^1_{i-1,M} - p^2_{i,M-1}) \\
&+ u_{N,M} p^2_{N,M-1} + \sum_{j=1}^{M-1} -u_{N,j}(p^2_{N,j} - p^2_{N,j-1} - p^1_{N-1,j}) + u_{N,M} p^1_{N-1,M} = \\
&= \sum_{i=1}^{N-1}\sum_{j=1}^{M-1} -u_{i,j}\operatorname{div}(p)_{i,j} + \\
&+ \sum_{i=1}^{N-1} -u_{i,M}(p^1_{i,M} - p^1_{i-1,M} - p^2_{i,M-1}) + \sum_{j=1}^{M-1} -u_{N,j}(-p^1_{N-1,j} + p^2_{N,j} - p^2_{N,j-1}) \\
&- u_{N,M}(-p^1_{N-1,M} - p^2_{N,M-1}) = \sum_{i=1}^{N}\sum_{j=1}^{M} -u_{i,j}\operatorname{div}(p)_{i,j} = \langle u, -\operatorname{div}(p)\rangle.
\end{aligned}$$

□

3.2.2 Minimization Problem and Solution Algorithm

In this section, we introduce a discrete analogue of the continuous model $(P(y_0))$ from Chapter 2. We will focus on the task of inpainting, since in this case we require on parameter less than for the linear inverse problem. This is due to the nature of the data fidelity term for inpainting, which has values in $\{0, \infty\}$, hence a multiplicative factor has no effect on the value of the data fidelity term. The choice of such a parameter could significantly impact the results of the method, making it more difficult to compare the performance to existing methods. Let $u_0 \in \mathbb{R}^{N \times M}$ be the unknown, original image, we would like to reconstruct. There are some pixels of u_0, which are known and others, which are not, formally we are given an image $\mathcal{M}$, called mask, of the same size as u_0, such that $\mathcal{M}_{i,j} = 1$, when $(u_0)_{i,j}$ is known and $\mathcal{M}_{i,j} = 0$ else. To simplify notation, we introduce the set

$$\mathfrak{M}(u_0) := \left\{ v \in \mathbb{R}^{N \times M} \;\middle|\; v_{i,j} = (u_0)_{i,j} \text{ if } \mathcal{M}_{i,j} = 1 \right\}.$$

We propose the following minimization problem, in order to obtain a reconstruction of u_0:

$$\begin{aligned} \min_{u,\mu,\theta} F(u, \mu, \theta) &:= \mathcal{I}_{\mathfrak{M}(u_0)}(u) + \lambda_{\mathrm{TV}} \mathrm{TV}_\epsilon(u - v) + \\ &+ \lambda_\mu \|\mu\|_1 + \lambda_\theta \sum_{i=2}^{n} \left(\left\|\theta^i\right\|_1 + \mathcal{I}_{\{.\geq 0\}}(\theta^i) \right) + \mathcal{I}_A(\theta^1), \end{aligned} \tag{DP}$$

subject to

$$\begin{aligned} v = v(\mu, \theta) &= \sum_{k=1}^{n_{f,1}} \mu_k^1 *_\sigma \theta_k^1 \\ \text{and for } j = 2, 3, ..., n, l = 1, 2, ..., n_{f,j-1} : \quad \mu_l^{j-1} &= \sum_{k=1}^{n_{f,j}} \mu_k^j *_\sigma \theta_{k,l}^j, \end{aligned}$$

where the filter kernels $\theta_k^1 \in \mathbb{R}^{ksz \times ksz}$ for $k = 1, 2, ...n_{f,1}$, $\theta_{k,l}^j \in \mathbb{R}^{ksz \times ksz}$ for $j = 2, 3, ..., n$, $k = 1, 2, ..., n_{f,j}$ and $l = 1, 2, ..., n_{f,j-1}$ (ksz is short for kernel size) and and the closed and convex set A is defined as

$$A = \left\{ \theta^1 \in \mathbb{R}^{ksz \times ksz \times n_{f,1}} \;\middle|\; \forall k = 1, 2, ...n_{f,1} : \left\|\theta_k^1\right\|_2 \leq 1, \sum_{i,j=1}^{ksz} \theta_{i,j}^1 = 0 \right\}$$

The superscripts indicate the layer of our convolutional model and $n_{f,j}$ is the number of filters/the filter depth in layer j. We denote

$$\begin{aligned} \mu &:= (\mu_k^n)_{k=1,2,...,n_{f,n}}, \\ \mu^j &:= (\mu_k^j)_{k=1,2,...n_{f,j}}, \\ \theta &:= ((\theta_k^1)_k, (\theta_{k,l}^j)_{j,k,l}) \quad \text{and,} \\ \theta_l^j &:= (\theta_{k,l}^j)_{k=1,2,...n_{f,j}}. \end{aligned}$$

In practice, the sizes of the filter kernels and the stride may be chosen arbitrarily, but the size of the coefficient image μ is then determined according to Definition 3.18 and

Definition 3.20. TV_ϵ is the following discrete analogue of J from Chapter 2.

$$\begin{aligned} \mathrm{TV}_\epsilon : \mathbb{R}^{N\times M} &\to [0,\infty) \\ u &\mapsto \phi_\epsilon(\nabla u), \end{aligned} \tag{3.11}$$

with

$$\begin{aligned} \phi_\epsilon : \mathbb{R}^{N\times M\times 2} &\to [0,\infty) \\ p = (p^k_{i,j})_{\substack{i=1,2,\dots,N\\ j=1,2,\dots,M\\ k=1,2}} &\mapsto \sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon}. \end{aligned} \tag{3.12}$$

The problem (DP) is the discrete analogue to the continuous problem investigated in Section 2.2.2 in the case of inpainting. Clearly, the goal is to apply Algorithm 3.11 to (DP). Accordingly, we have to distinguish different blocks of variables, which are updated separately in the algorithm, i.e., $x_1 \coloneqq u$, $x_2 \coloneqq \mu$ and $x_3 \coloneqq \theta$. Further we identify the smooth coupling function

$$H(x) = H(u,\mu,\theta) = \lambda_{\mathrm{TV}}\mathrm{TV}_\epsilon(u - v(\mu,\theta))$$

and the non-smooth functions

$$\begin{aligned} f_1(x_1) &= f_1(u) = \mathcal{I}_{\mathfrak{M}(u_0)}(u), \\ f_2(x_2) &= f_2(\mu) = \lambda_\mu \left\|\mu\right\|_1, \\ f_3(x_3) &= f_3(\theta) = \lambda_\theta \sum_{i=2}^{n}\left(\left\|\theta^i\right\|_1 + \mathcal{I}_{\{.\geq 0\}}(\theta^i)\right) + \mathcal{I}_A(\theta^1). \end{aligned}$$

In order to apply the algorithm, we need the gradient of H and the proximal mappings of all f_j. Using, that the adjoint operator of ∇, satisfies $\nabla^* = -\operatorname{div}$, with div the discrete divergence (see Lemma 3.27), we obtain

$$\begin{aligned} \nabla_u H(u,\mu,\theta) &= \lambda_{\mathrm{TV}}\ \nabla^*\ \nabla\phi_\epsilon(\nabla(u-v)) = \\ &= -\lambda_{\mathrm{TV}} \operatorname{div}\left(\underbrace{\left(\frac{(\nabla(u-v)^1_{i,j}, \nabla(u-v)^2_{i,j})}{\sqrt{(\nabla(u-v)^1_{i,j})^2 + (\nabla(u-v)^2_{i,j})^2+\epsilon}}\right)_{i,j}}_{\nabla\phi_\epsilon(\nabla(u-v))}\right). \end{aligned} \tag{3.13}$$

Moreover, by the linearity of the convolution we get for the derivative with respect to μ, denoted as D_μ,

$$\begin{aligned} D_\mu v = D_\mu K^\sigma_{\theta^1}(\mu^1) = K^\sigma_{\theta^1} D_\mu \mu^1 = K^\sigma_{\theta^1} D_\mu K^\sigma_{\theta^2}(\mu^2) = K^\sigma_{\theta^1}K^\sigma_{\theta^2} D_\mu \mu^2 = \\ = \dots = K^\sigma_{\theta^1}K^\sigma_{\theta^2}K^\sigma_{\theta^3}\dots K^\sigma_{\theta^n} \end{aligned}$$

and consequently, with * denoting the adjoint of an operator

$$\nabla_\mu v = (K^\sigma_{\theta^n})^*(K^\sigma_{\theta^{n-1}})^*...(K^\sigma_{\theta^2})^*(K^\sigma_{\theta^1})^*.$$

Therefore, we can compute

$$\nabla_\mu H = -\lambda_{\mathrm{TV}}\ \nabla_\mu v\ \nabla^*\ \nabla\phi_\epsilon(\nabla(u-v)) =$$
$$= \lambda_{\mathrm{TV}}(K^\sigma_{\theta^n})^*(K^\sigma_{\theta^{n-1}})^*...(K^\sigma_{\theta^2})^*(K^\sigma_{\theta^1})^* \operatorname{div}\left(\left(\frac{(\nabla(u-v)^1_{i,j},\nabla(u-v)^2_{i,j})}{\sqrt{(\nabla(u-v)^1_{i,j})^2+(\nabla(u-v)^2_{i,j})^2+\epsilon}}\right)_{i,j}\right). \tag{3.14}$$

A similar procedure yields for the derivatives with respect to θ^j_l,

$$D_{\theta^j_l} v = K^\sigma_{\theta^1}K^\sigma_{\theta^2}...K^\sigma_{\theta^{j-1}}D_{\theta^j_l}(C^\sigma_{\mu^j}(\theta^j_{i_{j-1}}))_{i_{j-1}}$$

and hence

$$\begin{aligned}\nabla_{\theta^j_l} H &= -\lambda_{\mathrm{TV}}\ \nabla_{\theta^j_l} v\ \nabla^*\ \nabla\phi_\epsilon(\nabla(u-v)) =\\ &= \lambda_{\mathrm{TV}}(0,0,...,\underbrace{(C^\sigma_{\mu^j})^*}_{\text{position } l},0,...,0)(K^\sigma_{\theta^{j-1}})^*(K^\sigma_{\theta^{j-2}})^*...\\ &...(K^\sigma_{\theta^2})^*(K^\sigma_{\theta^1})^* \operatorname{div}\left(\left(\frac{(\nabla(u-v)^1_{i,j},\nabla(u-v)^2_{i,j})}{\sqrt{(\nabla(u-v)^1_{i,j})^2+(\nabla(u-v)^2_{i,j})^2+\epsilon}}\right)_{i,j}\right).\end{aligned} \tag{3.15}$$

Let us now compute the proximal mappings of the functions $(f_j)_j$. We start by computing $\operatorname{prox}^{f_1}_\tau$.

$$\begin{aligned}\operatorname{prox}^{f_1}_\tau(w) &= \operatorname{argmin}\left\{\mathcal{I}_{\mathfrak{M}(u_0)}(u)+\frac{\tau}{2}\|w-u\|_2^2 \;\middle|\; u\right\} =\\ &= \begin{cases}(u_0)_{i,j} & \text{for } \mathcal{M}_{i,j}=1\\ w_{i,j} & \text{else}\end{cases}.\end{aligned} \tag{3.16}$$

The computation of $\operatorname{prox}^{f_2}_\tau$ can be done by invoking some elementary analysis.

$$\begin{aligned}\operatorname{prox}^{f_2}_\tau(\eta) &= \operatorname{argmin}\left\{\lambda_\mu\|\mu\|_1+\frac{\tau}{2}\|\eta-\mu\|_2^2 \;\middle|\; \mu\right\} =\\ &= \operatorname{argmin}\left\{\sum_{k,i,j}\lambda_\mu|\mu_{k,i,j}|+\frac{\tau}{2}(\eta_{k,i,j}-\mu_{k,i,j})^2 \;\middle|\; \mu\right\}.\end{aligned}$$

In this case, we can minimize with respect to all entries of μ separately, yielding the following result,

$$\operatorname{prox}^{f_2}_\tau(\eta) = \begin{cases}\eta_{k,i,j}-\frac{\lambda_\mu}{\tau} & \text{for } \eta_{k,i,j}>\frac{\lambda_\mu}{\tau}\\ \eta_{k,i,j}+\frac{\lambda_\mu}{\tau} & \text{for } \eta_{k,i,j}<-\frac{\lambda_\mu}{\tau}\\ 0, & \text{else}\end{cases}. \tag{3.17}$$

Next, let us consider the proximal mapping of f_3.

$$\operatorname{prox}_\tau^{f_3}(\rho) = \operatorname{argmin}\left\{ \lambda_\theta \sum_{i=2}^{n} \left(\left\| \theta^i \right\|_1 + \mathcal{I}_{\{.\geq 0\}}(\theta^i) \right) + \mathcal{I}_A(\theta^1) + \frac{\tau}{2} \left\| \rho - \theta \right\|_2^2 \;\middle|\; \theta \right\}.$$

Again, the solution $\hat{\theta} := \operatorname{prox}_\tau^{f_3}(\rho)$ is obtained by minimizing with respect to all layers θ^i separately. Starting with θ^1, we find by definition of A

$$\hat{\theta}^1 = \operatorname{argmin}\left\{ \left\| \rho^1 - \theta^1 \right\|_2^2 \;\middle|\; \theta^1 : \forall k = 1, 2, ...n_{f,1} : \left\| \theta_k^1 \right\|_2 \leq 1, \sum_{i,j} (\theta_k^1)_{i,j} = 0 \right\} \tag{3.18}$$

According to equation (3.18), $\hat{\theta}_k^1$ is the projection of ρ_k^1 onto the closed and convex set $\left\{ \theta_k^1 \mid \|\theta_k^1\|_2 \leq 1, \sum_{i,j} (\theta_k^1)_{i,j} = 0 \right\}$. Invoking Lemma A.3, we see, that we obtain $\operatorname{prox}_\tau^{f_3}(\rho^1)_k$ by projecting ρ_k^1 first onto the linear space $\left\{ \theta_k^1 \mid \sum_{i,j} (\theta_k^1)_{i,j} = 0 \right\}$ and then the result onto the unit ball. Altogether, we find, that

$$(\hat{\theta}_k^1)_{r,s} = \frac{(\rho_k^1)_{r,s} - \frac{1}{(ksz_1)^2} \sum_{i,j} (\rho_k^1)_{i,j}}{\max\left\{ 1, \left\| \left((\rho_k^1)_{r',s'} - \frac{1}{(ksz_1)^2} \sum_{i,j} (\rho_k^1)_{i,j} \right)_{r',s'} \right\|_2 \right\}}. \tag{3.19}$$

The remaining layers $\hat{\theta}^i$ for $1 < i \leq n$ are all computed identically and similarly to $\operatorname{prox}_\tau^{f_2}$,

$$\hat{\theta}^i = \operatorname{argmin}\left\{ \lambda_\theta \left\| \theta^i \right\|_1 + \mathcal{I}_{\{.\geq 0\}}(\theta^i) + \frac{\tau}{2} \left\| \rho^i - \theta^i \right\|_2^2 \mid \theta^i \right\} =$$

$$\operatorname{argmin}\left\{ \sum_{k,l,r,s} \left(\lambda_\theta \left\| (\theta_{k,l}^i)_{r,s} \right\|_1 + \frac{\tau}{2} ((\rho_{k,l}^i)_{r,s} - (\theta_{k,l}^i)_{r,s})^2 \right) \mid \theta^i : (\theta_{k,l}^i)_{r,s} \geq 0 \right\}.$$

Again, the minimization can be performed for each component of θ^i separately and the result is

$$(\hat{\theta}_{k,l}^i)_{r,s} = \begin{cases} (\rho_{k,l}^i)_{r,s} - \frac{\lambda_\theta}{\tau} & \text{for } (\rho_{k,l}^i)_{r,s} > \frac{\lambda_\theta}{\tau} \\ 0 & \text{else} \end{cases}. \tag{3.20}$$

This finishes our computations and we are now prepared to write down the algorithm. Please note, that in the following, the sub- and superscripts denoted as 'm' always refer to the iteration number of the algorithm.

Algorithm 3.28 GENERATIVE REGULARIZATION

inputs: corrupted input image u_0, image with info about known pixels M, $\epsilon \in (0,1)$, $\overline{\beta_u}, \overline{\beta_\mu}, \overline{\beta_\theta} \in (0, 1-\epsilon)$

initialize: $u^0 = u^1, \mu^0 = \mu^1, \theta^0 = \theta^1$

for $m = 1, 2, ...$ **do**

update u**:**

Take $\beta_u^m \in [0, \overline{\beta_u}]$ and compute

$$\tilde{u}^m = u^m + \beta_u^m (u^m - u^{m-1}).$$

Find L_u^m satisfying the descent and Lipschitz property with respect to u (see Assumption 3.15, 3), then compute step length τ_u^m and update u.

$$\begin{aligned}
\delta_u^m &= \frac{3\overline{\beta_u}}{2(1-\epsilon-\overline{\beta_u})} L_u^m, \\
\tau_u^m &= \frac{(1+\epsilon)\delta_u^m + (1+\beta_u^m)L_u^m}{2-\beta_u^m}, \\
\overline{u}^m &= \tilde{u}^m - \frac{1}{\tau_u^m} \nabla_u H(\tilde{u}^m, \mu^m, \theta^m) \\
(u^{m+1})_{i,j} &= \begin{cases} (u_0)_{i,j} & \text{if } \mathcal{M}_{i,j} = 1 \\ (\overline{u}^m)_{i,j} & \text{else} \end{cases}
\end{aligned} \tag{3.21}$$

update μ**:**

Take $\beta_\mu^m \in [0, \overline{\beta_\mu}]$ and compute

$$\tilde{\mu}^m = \mu^m + \beta_\mu^m (\mu^m - \mu^{m-1}).$$

Find L_μ^m satisfying the descent and Lipschitz property with respect to μ (see Assumption 3.15, 3), then compute step length τ_μ^m and update μ.

$$\begin{aligned}
\delta_\mu^m &= \frac{3\overline{\beta_\mu}}{2(1-\epsilon-\overline{\beta_\mu})} L_\mu^m, \\
\tau_\mu^m &= \frac{(1+\epsilon)\delta_\mu^m + (1+\beta_\mu^m)L_\mu^m}{2-\beta_\mu^m}, \\
\overline{\mu}^m &= \tilde{\mu}^m - \frac{1}{\tau_\mu^m} \nabla_\mu H(u^{m+1}, \tilde{\mu}^m, \theta^m) \\
(\mu^{m+1})_{k,i,j} &= \begin{cases} (\overline{\mu}^m)_{k,i,j} - \frac{\lambda_\mu}{\tau_\mu^m} & \text{for } (\overline{\mu}^m)_{k,i,j} > \frac{\lambda_\mu}{\tau_\mu^m} \\ (\overline{\mu}^m)_{k,i,j} + \frac{\lambda_\mu}{\tau_\mu^m} & \text{for } (\overline{\mu}^m)_{k,i,j} < -\frac{\lambda_\mu}{\tau_\mu^m} \\ 0, & \text{else} \end{cases}
\end{aligned} \tag{3.22}$$

update θ**:**

Take $\beta_\theta^m \in [0, \overline{\beta_\theta}]$ and compute

$$\tilde{\theta}^m = \theta^m + \beta_\theta^m(\theta^m - \theta^{m-1}).$$

Find L_θ^m satisfying the descent and Lipschitz property with respect to θ (see Assumption 3.15, 3), then compute step length τ_θ^m and update θ.

$$\begin{aligned}
\delta_\theta^m &= \frac{3\overline{\beta_\theta}}{2(1-\epsilon-\overline{\theta_\mu})} L_\theta^m,\\
\tau_\theta^m &= \frac{(1+\epsilon)\delta_\theta^m + (1+\beta_\theta^m)L_\theta^m}{2-\beta_\theta^m},\\
\overline{\theta}^m &= \tilde{\theta}^m - \frac{1}{\tau_\theta^m}\nabla_\theta H(u^{m+1}, \mu^{m+1}, \tilde{\theta}^m)\\
((\theta^{m+1})_k^1)_{r,s} &= \frac{((\overline{\theta}^m)_k^1)_{r,s} - \frac{1}{(ksz_1)^2}\sum_{i,j}((\overline{\theta}^m)_k^1)_{i,j}}{\max\left\{1, \left\|((\overline{\theta}^m)_k^1)_{r',s'} - \frac{1}{(ksz_1)^2}\sum_{i,j}((\overline{\theta}^m)_k^1)_{i,j})_{r',s'}\right\|_2\right\}},
\end{aligned} \tag{3.23}$$

$$\text{and for } i = 2, 3, ..., n:$$

$$((\theta^{m+1})_{k,l}^i)_{r,s} = \begin{cases} ((\overline{\theta}^m)_{k,l}^i)_{r,s} - \frac{\lambda_\theta}{\tau_\theta^m} & \text{for } ((\overline{\theta}^m)_{k,l}^i)_{r,s} > \frac{\lambda_\mu}{\tau_\theta^m} \\ 0 & \text{else} \end{cases}$$

end for

Remark 3.29. *Algorithm 3.28 slightly differs from Algorithm 3.11. The parameters δ_u^m, δ_μ^m and δ_θ^m are not constant, as Assumption 3.15 suggests. In order to compute $\delta_u^m, \delta_\mu^m, \delta_\theta^m$ according to Assumption 3.15, we would need upper bounds for L_u^m, L_μ^m and L_θ^m, which are not easily accessible.*

Remark 3.30. *The computation of L_u^m, L_μ^m and L_θ^m is not mentioned explicitly in Algorithm 3.28. Exemplarily, we will show it now for L_u^m in the form of a pseudo algorithm.*

Algorithm 3.31 BACKTRACKING

Compute L_u^m*, satisfying the Lipschitz property (Assumption 3.15)*
descent = False
while *not descent* ***do***
    Compute an update u^{m+1} *according to* (3.21) *using* L_u^m
    if *Descent property satisfied for* u^m, u^{m+1} *(Assumption 3.15)* ***then***
        descent = True
    else
        $L_u^m = 2L_u^m$
    end if
end while

In other words, we increase L_u^m*, until it satisfies both, the Lipschitz and the descent property. By Remark 3.16, this process terminates after a finite amount of iterations.*

It would also be possible, to obtain admissible L_u^m, L_μ^m *and* L_θ^m *by explicitly computing the Lipschitz constants of the partial gradients of* H *(see Lemma 3.37). In practice, however, this took longer than performing the backtracking above. Moreover, by Remark 3.14, the obtained values will in general be greater than necessary resulting in a smaller step size in the algorithm.*

Convergence of the Algorithm

In this section, we will show, that Algorithm 3.28 indeed converges to a critical point of the objective functional F, as stated in Theorem 3.17. Assumption 3.15 is fulfilled by the parameter choices and computations done in Algorithm 3.28 (up to the difference mentioned in Remark 3.29). Therefore, we are left to verify only the conditions stated in Assumption 3.8.

First of all, the functions $(f_i)_i$ are proper, convex and greater than or equal to zero. Moreover, they are lower semi-continuous, since indicator functions on closed sets are lower semi-continuous, multiplication with a positive real number does not affect lower semi-continuity and sums of lower semi-continuous functions are also lower semi-continuous. Also the coupling function H is differentiable and the objective function $F = H + f_1 + f_2 + f_3$ is greater than or equal to zero. The remaining tasks, in order to verify Assumption 3.8, are to investigate local Lipschitz continuity of the gradient of H, as well as showing that the objective function is a KL function (Definition 3.6). We start by showing the latter. In order to do so, we will first have to introduce the concept of semi-algebraic functions.

Definition 3.32. *[Semi-algebraic sets and functions, [7, Definition 5]]*

i) A subset $S \subseteq \mathbb{R}^d$ *is called semi-algebraic, if there exist* $p, q \in \mathbb{N}$ *and polynomial* *functions* $g_{i,j}, h_{i,j} : \mathbb{R}^d \to \mathbb{R}$ *for* $i = 1, 2, ..., p,\ j = 1, 2, ..., q,$ *such that*

$$S = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \left\{ x \in \mathbb{R}^d \mid g_{i,j}(x) = 0, h_{i,j}(x) < 0 \right\}$$

ii) A function $\phi : \mathbb{R}^d \to (-\infty, \infty]$ *is called semi-algebraic, if its graph*

$$\left\{ (x, \phi(x)) \mid x \in \mathbb{R}^d \right\}$$

is a semi-algebraic subset of $\mathbb{R}^{d+1}$.

Example 3.33. *The function* $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(x) = \sqrt{|x|}$ *is semi-algebraic.*

Proof. We can write the graph of ϕ as follows:

$$\{(x, \phi(x)) \mid x \in \mathbb{R}\} = \left\{ (x, \sqrt{|x|}) \mid x \in \mathbb{R} \right\} =$$
$$= \{(x, \sqrt{x}) \mid x \in \mathbb{R}, x > 0\} \cup \left\{ (x, \sqrt{-x}) \mid x \in \mathbb{R}, x < 0 \right\} \cup \{(0, 0)\} =$$
$$= \left\{ (x, t) \mid x - t^2 = 0, -t < 0 \right\} \cup \left\{ (x, t) \mid x + t^2 = 0, -t < 0 \right\} \cup \{(x, t) \mid x = 0, t = 0\}.$$

□

The following theorem will enable us to show the desired result.

Theorem 3.34. *[7, Theorem 3] Let* $\phi : \mathbb{R}^d \to (-\infty, \infty]$ *be a proper and lower semi-continuous function. If* ϕ *is semi-algebraic, then* ϕ *is a KL function.*

Lemma 3.35. *The objective function*

$$F(u, \mu, \theta) = H(u, \mu, \theta) + f_1(u) + f_2(\mu) + f_3(\theta) = \lambda_{\mathrm{TV}} \mathrm{TV}_\epsilon(u - v) + \mathcal{I}_{\mathfrak{M}(u_0)}(u) +$$
$$+ \lambda_\mu \|\mu\|_1 + \lambda_\theta \sum_{i=2}^{n} \left(\left\| \theta^i \right\|_1 + \mathcal{I}_{\{. \geq 0\}}(\theta^i) \right) + \mathcal{I}_A(\theta^1_k)$$

is a KL function.

Proof. Since the objective function F is indeed proper and lower semi-continuous, it suffices to prove, that F is semi-algebraic in order to apply Theorem 3.34. According to [7, Example 2] the following functions $\mathbb{R}^d \to (-\infty, \infty]$ are semi-algebraic,

i) real polynomial functions

ii) indicator functions on semi-algebraic sets

iii) semi-algebraic functions multiplied with a real scalar

iv) finite sums and products of semi-algebraic functions

v) composition of semi-algebraic functions

vi) $x \mapsto \|x\|_p \coloneqq \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}}$ for $p \geq 1$ rational.

Now, let us apply this to the objective function F.

- by ii), f_1 is semi-algebraic
- by iii), and vi) f_2 is semi-algebraic
- by ii), iii), iv) and vi), f_3 is semi-algebraic
- Recall, that $H(u, \mu, \theta) = \lambda_{\text{TV}} \sum_{i=1}^{N} \sum_{j=1}^{M} \sqrt{((\nabla(u-v))_{i,j}^1)^2 + ((\nabla(u-v))_{i,j}^2)^2 + \epsilon}$. Since $v(\mu, \theta)$ is polynomial in (μ, θ) and the discrete gradient is linear, by i) we obtain, that the expression under the square root is semi-algebraic for any i, j. By Example 3.33 the square root is a semi-algebraic function as well, hence, by v), iv) and iii) H is semi-algebraic.

Altogether, by successive application of iv), the objective function F is semi-algebraic, and as a consequence also a KL function. □

We continue by investigating ∇H, which will conclude the verification of Assumption 3.8.

Lemma 3.36. *With*

$$\phi_\epsilon : \mathbb{R}^{N \times M \times 2} \to [0, \infty)$$
$$p = (p_{i,j}^k)_{\substack{i=1,2,\dots,N \\ j=1,2,\dots,M \\ k=1,2}} \mapsto \sum_{i=1}^{N} \sum_{j=1}^{M} \sqrt{(p_{i,j}^1)^2 + (p_{i,j}^2)^2 + \epsilon},$$

it holds true, that $\nabla \phi_\epsilon$ is Lipschitz continuous with a Lipschitz constant $L \leq \frac{3}{2\sqrt{\epsilon}}$ with respect to $\| \, . \, \|_2$, i.e., for all $p, q \in \mathbb{R}^{N \times M \times 2}$

$$\|\nabla \phi_\epsilon(p) - \nabla \phi_\epsilon(q)\|_2 \leq \frac{3}{2\sqrt{\epsilon}} \|p - q\|_2 \, .$$

Proof. We start by simply computing the gradient of ϕ_ϵ explicitly,

$$\partial_{p_{i,j}^k} \phi_\epsilon = \frac{p_{i,j}^k}{\sqrt{(p_{i,j}^1)^2 + (p_{i,j}^2)^2 + \epsilon}}.$$

In order to show Lipschitz continuity of $\nabla\phi_\epsilon$, it suffices to prove, that the the Jacobian of $\nabla\phi_\epsilon$ is bounded. Hence, we continue by also computing all second partial derivatives of ϕ_ϵ. Note that, $\partial_{p^k_{i,j}}\partial_{p^{k'}_{i',j'}}\phi_\epsilon = 0$, whenever $i \neq i'$ or $j \neq j'$. Therefore, and due to the symmetry in k of $\partial_{p^k_{i,j}}\phi_\epsilon$, it suffices to compute $\partial_{p^1_{i,j}}\partial_{p^1_{i,j}}\phi_\epsilon$ and $\partial_{p^2_{i,j}}\partial_{p^1_{i,j}}\phi_\epsilon$. We obtain

$$\left|\partial_{p^1_{i,j}}\partial_{p^1_{i,j}}\phi_\epsilon\right| = \left|\frac{\sqrt{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon} - p^1_{i,j}\frac{p^1_{i,j}}{\sqrt{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon}}}{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon}\right| =$$

$$= \left|\frac{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon-(p^1_{i,j})^2}{\left((p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon\right)^{\frac{3}{2}}}\right| = \frac{(p^2_{i,j})^2+\epsilon}{\left((p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon\right)^{\frac{3}{2}}} \leq \frac{(p^2_{i,j})^2+\epsilon}{\left((p^2_{i,j})^2+\epsilon\right)^{\frac{3}{2}}} =$$

$$= \frac{1}{\sqrt{(p^2_{i,j})^2+\epsilon}} \leq \frac{1}{\sqrt{\epsilon}}$$

and similarly

$$\left|\partial_{p^2_{i,j}}\partial_{p^1_{i,j}}\phi_\epsilon\right| = \left|p^1_{i,j}\left(-\frac{1}{2}\right)\left((p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon\right)^{-\frac{3}{2}}2p^2_{i,j}\right| = \left|\frac{p^1_{i,j}p^2_{i,j}}{\left((p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon\right)^{\frac{3}{2}}}\right| \leq$$

$$\underbrace{\leq}_{\text{Young's inequality}} \frac{\frac{(p^1_{i,j})^2}{2}+\frac{(p^2_{i,j})^2}{2}}{\left((p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon\right)^{\frac{3}{2}}} = \frac{1}{2}\underbrace{\frac{1}{\sqrt{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon}}}_{\leq\frac{1}{\sqrt{\epsilon}}}\underbrace{\frac{(p^1_{i,j})^2+(p^2_{i,j})^2}{(p^1_{i,j})^2+(p^2_{i,j})^2+\epsilon}}_{\leq 1} \leq \frac{1}{2\sqrt{\epsilon}}.$$

Therefore, we can already conclude, that $\nabla\phi_\epsilon$ is Lipschitz continuous, since every entry of its Jacobian is bounded by the constant $\frac{1}{\sqrt{\epsilon}}$. In order to get an estimate of the Lipschitz constant, we now have to compute a bound for the norm of the Jacobian matrix.

$$\|D\nabla\phi_\epsilon(p)\|_{2,2}^2 = \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \|D\nabla\phi_\epsilon(p)v\|_2^2 = \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M\\ k=1,2}} \left|(D\nabla\phi_\epsilon(p)v)_{i,j}^k\right|^2 =$$

$$= \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M\\ k=1,2}} \left| \sum_{\substack{i'=1,\dots,N\\ j'=1,\dots,M\\ k'=1,2}} \overbrace{\partial_{i',j'}^{k'}\partial_{i,j}^{k}\phi_\epsilon(p)}^{=0,\text{ if }(i,j)\neq(i',j')} \; v_{i',j'}^{k'} \right|^2 \leq$$

$$\leq \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} \left(\frac{1}{\sqrt{\epsilon}}\left|v_{i,j}^1\right| + \frac{1}{2\sqrt{\epsilon}}\left|v_{i,j}^2\right|\right)^2 + \left(\frac{1}{\sqrt{\epsilon}}\left|v_{i,j}^2\right| + \frac{1}{2\sqrt{\epsilon}}\left|v_{i,j}^1\right|\right)^2 =$$

$$= \frac{1}{\epsilon} \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} \left(\left|v_{i,j}^1\right|^2 + \frac{1}{4}\left|v_{i,j}^2\right|^2 + \left|v_{i,j}^1 v_{i,j}^2\right| + \left|v_{i,j}^2\right|^2 + \frac{1}{4}\left|v_{i,j}^1\right|^2 + \left|v_{i,j}^1 v_{i,j}^2\right|\right) =$$

$$= \frac{1}{\epsilon} \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} \left(\frac{5}{4}\left|v_{i,j}^1\right|^2 + \frac{5}{4}\left|v_{i,j}^2\right|^2 + \underbrace{2\left|v_{i,j}^1 v_{i,j}^2\right|}_{\leq \left|v_{i,j}^1\right|^2 + \left|v_{i,j}^2\right|^2}\right) \leq$$

$$\leq \frac{1}{\epsilon} \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \sum_{\substack{i=1,\dots,N\\ j=1,\dots,M}} \left(\frac{9}{4}\left|v_{i,j}^1\right|^2 + \frac{9}{4}\left|v_{i,j}^2\right|^2\right) =$$

$$= \frac{9}{4\epsilon} \max_{\substack{v\in\mathbb{R}^{N\times M\times 2}\\ \|v\|_2\leq 1}} \|v\|_2^2 = \frac{9}{4\epsilon}.$$

Consequently, $\|D\nabla\phi_\epsilon(p)\|_2 \leq \frac{3}{2\sqrt{\epsilon}}$, which is also a bound for the Lipschitz constant. □

Lemma 3.36 was already the most difficult part of proving the Lipschitz continuity of the (partial) gradient(s) of H. Now we can collect the results.

Lemma 3.37. *The following functions are Lipschitz continuous:*

i) $u \mapsto \nabla_u H(u,\mu,\theta)$ *for fixed* μ *and* θ*,*

ii) $\mu \mapsto \nabla_\mu H(u,\mu,\theta)$ *for fixed* u *and* μ*,*

iii) for $j = 1,2,\dots,n$*,* $\theta^j \mapsto \nabla_{\theta^j} H(u,\mu,\theta)$ *for fixed* u*,* μ *and* θ^i *for* $i \neq j$ *and* *iv) ∇H on bounded sets, i.e., for any bounded set $B_u \times B_\mu \times B_\theta \subset \mathbb{R}^{\text{size}(u)} \times \mathbb{R}^{\text{size}(\mu)} \times \mathbb{R}^{\text{size}(\theta)}$, there exists $M > 0$, such that for all $(u, \mu, \theta), (v, \eta, \rho) \in B_u \times B_\mu \times B_\theta$,*

$$\|\nabla H(u,\mu,\theta) - \nabla H(v,\eta,\rho)\|_2 \leq M \|(u-v, \mu-\eta, \theta-\rho)\|_2 .$$

Proof. The proof is done by computation and invoking previous results.

i) We start with $\nabla_u H$. In the following we will simply write $v = v(\mu, \theta)$.

$$\begin{aligned}
&\|\nabla_u H(u,\mu,\theta) - \nabla_u H(\tilde{u},\mu,\theta)\|_2 = \\
&= \|\lambda_{\text{TV}} \nabla^* \nabla \phi_\epsilon(\nabla(u-v)) - \lambda_{\text{TV}} \nabla^* \nabla \phi_\epsilon(\nabla(\tilde{u}-v))\|_2 \leq \\
&\leq \lambda_{\text{TV}} \|\nabla^*\|_2 \|\nabla \phi_\epsilon(\nabla(u-v) - \nabla \phi_\epsilon(\nabla(\tilde{u}-v)\|_2 \leq \\
&\underbrace{\leq}_{Lemma\ 3.36} \lambda_{\text{TV}} \|\nabla^*\|_2 \frac{3}{2\sqrt{\epsilon}} \|\nabla(u-\tilde{u})\|_2 \leq \lambda_{\text{TV}} \|\nabla^*\|_2 \frac{3}{2\sqrt{\epsilon}} \|\nabla\|_2 \|u-\tilde{u}\|_2 \leq \\
&\underbrace{\leq}_{\substack{Lemma\ 3.25 \text{ with} \\ \|\nabla^*\|_2 = \|\nabla\|_2}} \lambda_{\text{TV}} \frac{12}{\sqrt{\epsilon}} \|u-\tilde{u}\|_2 .
\end{aligned}$$

This shows not only that $u \mapsto \nabla_u H(u,\mu,\theta)$ is Lipschitz continuous, but also that the Lipschitz constant is uniform in μ and θ.

ii) We continue with $\nabla_\mu H$. Note, that, since $v(\mu,\theta)$ is linear in μ, $\nabla_\mu v(\mu,\theta) = \nabla_\mu v(\tilde{\mu},\theta)$ and we will simply denote it as $\nabla_\mu v(\theta)$.

$$\begin{aligned}
&\|\nabla_\mu H(u,\mu,\theta) - \nabla_\mu H(u,\tilde{\mu},\theta)\|_2 = \\
&= \|\lambda_{\text{TV}} \nabla_\mu v(\theta) \nabla^* \nabla \phi_\epsilon(\nabla(u - v(\mu,\theta))) - \lambda_{\text{TV}} \nabla_\mu v(\theta) \nabla^* \nabla \phi_\epsilon(\nabla(u - v(\tilde{\mu},\theta)))\|_2 \leq \\
&\leq \lambda_{\text{TV}} \|\nabla_\mu v(\theta)\|_2 \|\nabla^*\|_2 \|\nabla \phi_\epsilon(\nabla(u - v(\mu,\theta))) - \nabla \phi_\epsilon(\nabla(u - v(\tilde{\mu},\theta)))\|_2 \leq \\
&\underbrace{\leq}_{Lemma\ 3.36} \lambda_{\text{TV}} \|\nabla_\mu v(\theta)\|_2 \|\nabla^*\|_2 \frac{3}{2\sqrt{\epsilon}} \|\nabla(v(\mu,\theta) - v(\tilde{\mu},\theta))\|_2 \leq \\
&\leq \lambda_{\text{TV}} \|\nabla_\mu v(\theta)\|_2 \|\nabla^*\|_2 \frac{3}{2\sqrt{\epsilon}} \|\nabla\|_2 \|v(\mu,\theta) - v(\tilde{\mu},\theta)\|_2 \leq \\
&\underbrace{\leq}_{\|\nabla^*\|_2 = \|\nabla\|_2} \lambda_{\text{TV}} \|\nabla_\mu v(\theta)\|_2 \|\nabla\|_2^2 \frac{3}{2\sqrt{\epsilon}} \|D_\mu v(\theta)\|_2 \|\mu - \tilde{\mu}\|_2 \leq \\
&\underbrace{\leq}_{\|D_\mu v\|_2 = \|\nabla_\mu v\|_2} \lambda_{\text{TV}} \|\nabla\|_2^2 \frac{3}{2\sqrt{\epsilon}} \left\|K^\sigma_{\theta_1} K^\sigma_{\theta_2} ... K^\sigma_{\theta_n}\right\|_2^2 \|\mu - \tilde{\mu}\|_2 \leq \\
&\underbrace{\leq}_{\substack{Lemma\ 3.25 \\ Lemma\ 3.23}} \lambda_{\text{TV}} \frac{12}{\sqrt{\epsilon}} \sum_{k=1}^{n_{f,1}} \left\|\theta_k^1\right\|_1^2 \sum_{l=1}^{n_{f,1}} \sum_{k=1}^{n_{f,2}} \left\|(\theta_l^2)_k\right\|_1^2 \cdots \sum_{l=1}^{n_{f,n-1}} \sum_{k=1}^{n_{f,n}} \|(\theta_l^n)_k\|_1^2 \|\mu - \tilde{\mu}\| .
\end{aligned}$$

Therefore, $\mu \mapsto \nabla_\mu H(u,\mu,\theta)$ is Lipschitz continuous with a Lipschitz constant depending on θ.

iii) Let us now consider $\nabla_{\theta^j} H$. Let θ and $\tilde\theta$ be, such that $\theta^i = \tilde\theta^i$ for $i \neq j$, but it may be $\theta^j \neq \tilde\theta^j$. Again, $v(\mu,\theta)$ is linear in θ^j, therefore $\nabla_{\theta^j} v(\mu,\theta)$ does not depend on θ^j (but it does indeed depend on θ^i for $i < j$). We compute

$$
\begin{aligned}
&\left\|\nabla_{\theta^j} H(u,\mu,\theta) - \nabla_{\theta^j} H(u,\mu,\tilde\theta)\right\|_2 \leq \\
&\leq \lambda_{\mathrm{TV}} \left\|\nabla_{\theta^j} v(\mu,\theta)\right\|_2 \left\|\nabla^*\right\|_2 \frac{3}{2\sqrt{\epsilon}} \left\|\nabla\right\|_2 \left\|v(\mu,\theta) - v(\mu,\tilde\theta)\right\|_2 \leq \\
&\underbrace{\leq}_{\substack{\text{definition of}\\ v(\mu,\theta)}} \lambda_{\mathrm{TV}} \left\|\nabla_{\theta^j} v(\mu,\theta)\right\|_2 \left\|\nabla\right\|_2^2 \frac{3}{2\sqrt{\epsilon}} \left\| K^\sigma_{\theta^1} K^\sigma_{\theta^2} \dots K^\sigma_{\theta^{j-1}} \left[(C^\sigma_{\mu^j}(\theta^j_{i_{j-1}}))_{i_{j-1}} - (C^\sigma_{\mu^j}(\tilde\theta^j_{i_{j-1}}))_{i_{j-1}} \right] \right\|_2 \leq \\
&\leq \lambda_{\mathrm{TV}} \left\|\nabla_{\theta^j} v(\mu,\theta)\right\|_2 \left\|\nabla\right\|_2^2 \frac{3}{2\sqrt{\epsilon}} \left\| K^\sigma_{\theta^1} K^\sigma_{\theta^2} \dots K^\sigma_{\theta^{j-1}} \right\|_2 \sqrt{\sum_{i=1}^{n_{f,j-1}} \left\| C^\sigma_{\mu^j}(\theta^j_i) - C^\sigma_{\mu^j}(\tilde\theta^j_i) \right\|_2^2} \leq \\
&\leq \lambda_{\mathrm{TV}} \left\|\nabla_{\theta^j} v(\mu,\theta)\right\|_2 \left\|\nabla\right\|_2^2 \frac{3}{2\sqrt{\epsilon}} \left\| K^\sigma_{\theta^1} K^\sigma_{\theta^2} \dots K^\sigma_{\theta^{j-1}} \right\|_2 \sqrt{\sum_{i=1}^{n_{f,j-1}} \left\| C^\sigma_{\mu^j} \right\|_2^2 \left\| \theta^j_i - \tilde\theta^j_i \right\|_2^2} = \\
&= \lambda_{\mathrm{TV}} \left\|\nabla_{\theta^j} v(\mu,\theta)\right\|_2 \left\|\nabla\right\|_2^2 \frac{3}{2\sqrt{\epsilon}} \left\| K^\sigma_{\theta^1} K^\sigma_{\theta^2} \dots K^\sigma_{\theta^{j-1}} \right\|_2 \left\| C^\sigma_{\mu^j} \right\|_2 \left\| \theta^j - \tilde\theta^j \right\|_2 \leq \\
&\underbrace{\leq}_{\substack{Lemma\ 3.25\\ Lemma\ 3.23}} \lambda_{\mathrm{TV}} \frac{12}{\sqrt{\epsilon}} \sum_{k=1}^{n_{f,1}} \left\|\theta^1_k\right\|_1^2 \sum_{l=1}^{n_{f,1}} \sum_{k=1}^{n_{f,2}} \left\|(\theta^2_l)_k\right\|_1^2 \dots \sum_{l=1}^{n_{f,j-2}} \sum_{k=1}^{n_{f,j-1}} \left\|(\theta^{j-1}_l)_k\right\|_1^2 \sum_{k=1}^{n_{f,j}} \left\|\mu^j_k\right\|_1^2 \left\|\theta^j - \tilde\theta^j\right\|_2
\end{aligned}
$$

iv) Lastly, we have to show, that ∇H is Lipschitz continuous on bounded sets. To this aim it suffices to note, that $H(u,\mu,\theta) = \lambda_{\mathrm{TV}}\phi_\epsilon(\nabla(u - v(\mu,\theta)))$ is C^∞, since ϕ_ϵ is C^∞, the discrete gradient ∇ is linear and $v(\mu,\theta)$ is polynomial in (μ,θ). As a consequence, H is Lipschitz continuous on bounded sets.

□

Lemma 3.38. *Let $(u^m,\mu^m,\theta^m)_m$ be the sequence generated by Algorithm 3.28. If the sequence $(F(u^m,\mu^m,\theta^m))_m$ is bounded, then also $(u^m,\mu^m,\theta^m)_m$ is bounded.*

Proof. First of all, we should recall the definition of the objective functional.

$$
\begin{aligned}
F(u,\mu,\theta) = &\mathcal{I}_{\mathfrak{M}(u_0)}(u) + \lambda_{\mathrm{TV}} \mathrm{TV}_\epsilon(u - v(\mu,\theta)) + \\
&+ \lambda_\mu \left\|\mu\right\|_1 + \lambda_\theta \sum_{i=2}^{n} \left(\left\|\theta^i\right\|_1 + \mathcal{I}_{\{.\geq 0\}}(\theta^i) \right) + \mathcal{I}_A(\theta^1).
\end{aligned}
$$

Now assume, that $(F(u^m, \mu^m, \theta^m))_m$ is bounded. Since

$$\begin{gathered}\lambda_{\mathrm{TV}} \mathrm{TV}_\epsilon(u - v(\mu, \theta)) \leq F(u, \mu, \theta), \\ \lambda_\mu \left\|\mu\right\|_1 \leq F(u, \mu, \theta), \\ \lambda_\theta \left\|\theta^i\right\|_1 \leq F(u, \mu, \theta) \quad \text{for } i = 2, 3, ..., n \text{ and} \\ \mathcal{I}_A(\theta^1) \leq F(u, \mu, \theta),\end{gathered}$$

we can conclude, that also $(\mu^m)_m$ and $(\theta^m)_m$ are bounded. The remaining task is to show, that also $(u^m)_m$ is bounded. By definition of TV_ϵ, we have for $w \in \mathbb{R}^{N \times M}$

$$\begin{aligned}\mathrm{TV}_\epsilon(w) = \phi_\epsilon(\nabla w) = \sum_{i=1}^{N} \sum_{j=1}^{M} \sqrt{((\nabla w)^1_{i,j})^2 + ((\nabla w)^2_{i,j})^2 + \epsilon} \geq \\ \geq \sum_{i=1}^{N} \sum_{j=1}^{M} \sqrt{((\nabla w)^1_{i,j})^2 + ((\nabla w)^2_{i,j})^2}.\end{aligned}$$

Thus, we obtain, that $(\nabla(u^m - v(\mu^m, \theta^m)))_m$ is bounded. By Lemma 3.23, $v(\mu^m, \theta^m)_m$ is also bounded, since we have already shown, that $(\mu^m)_m$ and $(\theta^m)_m$ are. The discrete gradient is continuous, therefore we can deduce, that $(\nabla(v(\mu^m, \theta^m)))_m$ and consequently $(\nabla u^m)_m$ are bounded. The inpainting term of the objective functional $\mathcal{I}_{\mathfrak{M}(u_0)}(u)$ ensures, that for some (i, j), $(u^m_{i,j})_m$ is constant for all m. Together with the boundedness of $(\nabla u^m)_m$, this yields also boundedness of $(u^m)_m$. □

Ultimately, we are allowed to use Theorem 3.17, which guarantees, that the sequence generated by Algorithm 3.28 converges to a critical point of the objective functional F, provided that the generated sequence is bounded. Boundedness of the sequence is given, e.g., if the according sequence of objective functional values is bounded, as stated in Lemma 3.38.

3.3 Experiments

In this section, we present some visual results of the proposed method. The three different test images, that we use, can be found in Figure 3.2.

Figure 3.2: The test images we use for our experiments. From left to right, we will refer to the images as Barbara, Barbara crop and mix.

Comparison to Existing Methods

We compare our method to total variation inpainting, to the deep learning method [15] and to the single layer version of the proposed method from [9]. We implemented our method in Python and compute as much as possible parallelly on the GPU using PyOpenCL. For all results from [15], we use the code, that was published by the authors [16]. In particular, we use the version of their code, that was designed for inpainting with the same network architecture and parameters, that were used for [15, Figure 7 (top)]. The minimization problem considered for the task of inpainting in [15] is

$$\min_{\eta} \|\mathcal{M} \odot (f_\eta(z) - u_0)\|_2^2 ,$$

where $\mathcal{M} \in \{0,1\}^{\text{size}(u_0)}$ is an inpainting mask, as in our problem formulation, $\odot$ denotes the pointwise product and $f_\eta(z)$ denotes a convolutional neural network with parameters η and input z. In particular, we emphasize, that the data term used in [15] is different to ours. That is, the authors of [15] use the L2-norm on the set of pixels, where the original image is known, whereas we use an indicator function. As a result, the objective functional in [15] is of greater smoothness than the one we use.

The code from [9] is also publicly available [10]. We use the TGV (total generalized variation) version of their method for inpainting, resulting in the following formulation

$$\min_{u,v} \mathcal{I}_{\mathfrak{M}(u_0)}(u) + \lambda \text{TGV}(u - v) + \mu \mathcal{N}(v),$$

where TGV denotes the total generalized variation and $\mathcal{N}$ is a functional ensuring, that the texture part v is the output of a single layer convolutional neural network. $\mathcal{N}$ also includes a penalty term, which acts as a further regularization. For further details we refer to [9]. We use the parameters, that the authors of [9] provided to reproduce the results from [9].

For total variation inpainting, we solve

$$\min_u \mathcal{I}_{\mathfrak{M}(u_0)}(u) + \mathrm{TV}(u),$$

where no parameter is needed, due to the very nature of the functional. For given $p \in (0,1)$, the inpainting mask $\mathcal{M}$ is initialized, such that independently for all (i,j), $\mathcal{M}_{i,j} = 1$ with probability p and $\mathcal{M}_{i,j} = 0$ with probability $1-p$. Practically, this means, that p is approximately the rate of known pixels of the original image. The parameters of the objective functional in (DP), including the network dimensions, and the parameters of Algorithm 3.28 are set according to Table 3.1 if not stated differently. We use the same filter depth $n_{f,j} = n_f$ for all layers $j = 1, 2, ...n$. In order to show applicability of the method, we use the same parameters for all examples except for the parameter λ_{TV}, which varies.

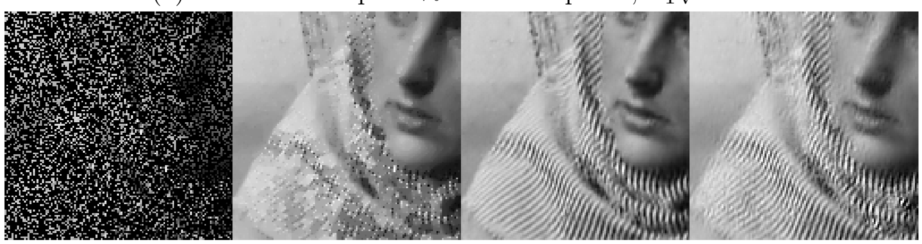
Parameter	Value
Parameters in (DP)	
ϵ	0.05
Network depth n	4
Filter depth $n_{f,j} = n_f$	12
Kernel size ksz	4
Stride σ	$\sigma = 1$ in 1st layer, $\sigma = 2$ in remaining layers
λ_{TV}	see examples
λ_μ	1.0
λ_θ	1.0
Parameters in Algorithm 3.28	
ϵ	0.03
$\beta_u^m, \overline{\beta_u}, \beta_\mu^m, \overline{\beta_\mu}, \beta_\theta^m, \overline{\beta_\theta}$	0.7
Number of iterations	5000

Table 3.1: Parameters of the objective functional and the algorithm

As a measure for the error of a result, we use the peak signal-to-noise ratio $\mathrm{PSNR} = -10\log_{10}\left(\frac{1}{NM}\|u-u_0\|_2^2\right)$, where $u \in \mathbb{R}^{N\times M}$ is the output of an algorithm and $u_0 \in \mathbb{R}^{N\times M}$ is the original image. This means, a high PSNR value implies, that the result u is close to the original image u_0. In Algorithm 3.28, we initialize u as $u_{i,j} = (u_0)_{i,j}$ for $\mathcal{M}_{i,j} = 1$ and $u_{i,j} = 0$ else. The coefficients μ and kernels θ from (DP) are initialized randomly. To be precise, each entry of θ is initialized independently according to a uniform distribution on $[0, n_f)$ and each entry of μ is initialized independently according to a uniform distribution on $[0,1)$ and then scaled, such that $\|v(\mu,\theta)\|_\infty = 1$. Unfortunately, the algorithm sometimes gets stuck in a local minimum, where the texture part is zero. In order to avoid this, we re-initialize μ and θ, when they become zero. Usually, after several iterations the re-initialization is not necessary anymore, which leads us to believe, that this issue could also be avoided by choosing appropriate initial values for the algorithm. Let us begin our practical investigation by comparing our results to existing methods. In Figure 3.3, we display examples of our method compared to TV inpainting and [15].

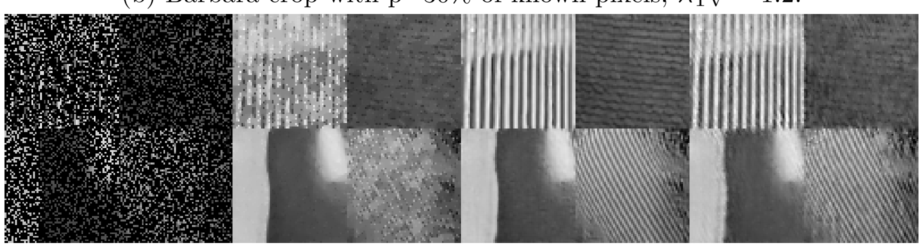
PSNR ≈ 8.89 PSNR ≈ 25.49 PSNR ≈ 29.32 PSNR ≈ 29.54

(a) Barbara with p=50% of known pixels, $\lambda_{\mathrm{TV}} = 1.0$.



PSNR ≈ 6.84 PSNR ≈ 19.26 PSNR ≈ 22.14 PSNR ≈ 23.83

(b) Barbara crop with p=30% of known pixels, $\lambda_{\mathrm{TV}} = 1.2$.



PSNR ≈ 7.10 PSNR ≈ 20.16 PSNR ≈ 28.10 PSNR ≈ 25.11

(c) Mix with p=30% of known pixels, $\lambda_{\mathrm{TV}} = 0.8$

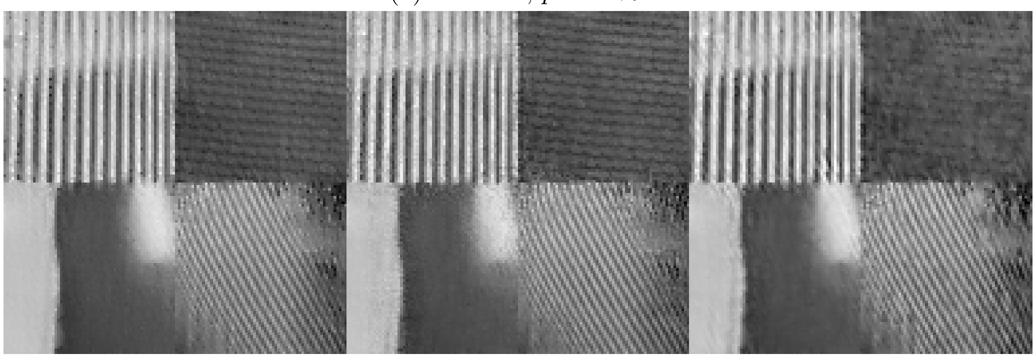
Figure 3.3: Comparison of our method to total variation inpainting and [15]. From left to right: Corrupted image, total variation inpainting, [15], proposed method

In all examples, our method outperforms total variation inpainting, in particular at areas of the image, where texture is located. Eventhough our method performs better than [15] regarding the PSNR value with the Barbara and Barbara crop image, the results from [15] appear visually superior to ours. One reason therefor could be the different data fidelity term. As mentioned above, our data term comprises a hard constraint to the known pixels, whereas [15] allows for deviation from the data also at known pixels. This might force the results of our method to be closer to the data. Another significant difference between the experiments in [15] and in the present thesis is the network architecture. In [15], authors use convolutional networks, that are substantially larger than the ones we use. For instance, when running the code from [15] with the mix image, there are 3002369 parameters to be fitted, while with our method, there are only 11436 network parameters (kernels θ and coefficients μ). Unfortunately, for comparably sized convolutional networks we experienced convergence problems when applying our algorithm to larger images such as the Barbara image. Moreover, in [15], there are also skip connections and activation functions included in the convolutional networks, which we did not use.

In Figure 3.4, we compare our method to the single layer version from [9]. We also added the result of our method with just one layer to the comparison. Note, that our method with only one convolutional layer still differs from [9], since there authors used lifting and relaxation techniques to obtain a convex problem, whereas we solved the original, non-convex problem (DP). The method from [9] outperforms the proposed method with the mix image, in particular in the upper right and lower left quadrants of the image. In the upper right area, the issue is, that our method does not generate the texture via the convolutional network, but treats it as cartoon part and therefore the TV penalty partly removes the texture (this effect can be observed in Figure 3.5 and Figure 3.7). In the lower left area, our method does not create as sharp edges as we would wish for. This could possibly be due to the fact, that we use a smooth approximation of TV, where we replaced the 1-norm by a smooth alternative (3.12), which lacks the sparsity enforcing property of the 1-norm. For a better comparison to [9], it would be interesting to adapt our method in the sense, that we also make use of a TGV-type cartoon prior.

PSNR ≈ 23.89 | $n = 1$, $n_f = 9$, $ksz = 15$, $\lambda_{\text{TV}} = 11$, PSNR ≈ 23.77 | $\lambda_{\text{TV}} = 1.2$, PSNR ≈ 23.83

(a) Barbara, $p = 30\%$.



PSNR ≈ 25.89 | $n = 1$, $n_f = 9$, $ksz = 15$, $\lambda_{\text{TV}} = 12.0$, PSNR ≈ 25.93 | $\lambda_{\text{TV}} = 1.2$, PSNR ≈ 25.11

(b) Mix, $p = 30\%$.

Figure 3.4: Comparison of the results to the single layer version. From left to right: [9] single layer, proposed method single layer, proposed method multilayer

Characteristics of the Proposed Method

In Figure 3.5 one can observe the implicit cartoon-texture decomposition of our method for the same examples and parameters as in Figure 3.3. With the Barbara image, as one would expect, the texture part contains the stripe pattern on the clothes of the pictured woman as well as the checked pattern on the table cloth. Also with the Barbara crop image, the stripe pattern is contained in the texture part. With the mix image, the texture is recognized as expected everywhere up to the upper right quadrant of the image, as already mentioned above. There, the convolutional network seems to fail to identify the texture correctly, at least for this specific parameter choice (see Figure 3.7).

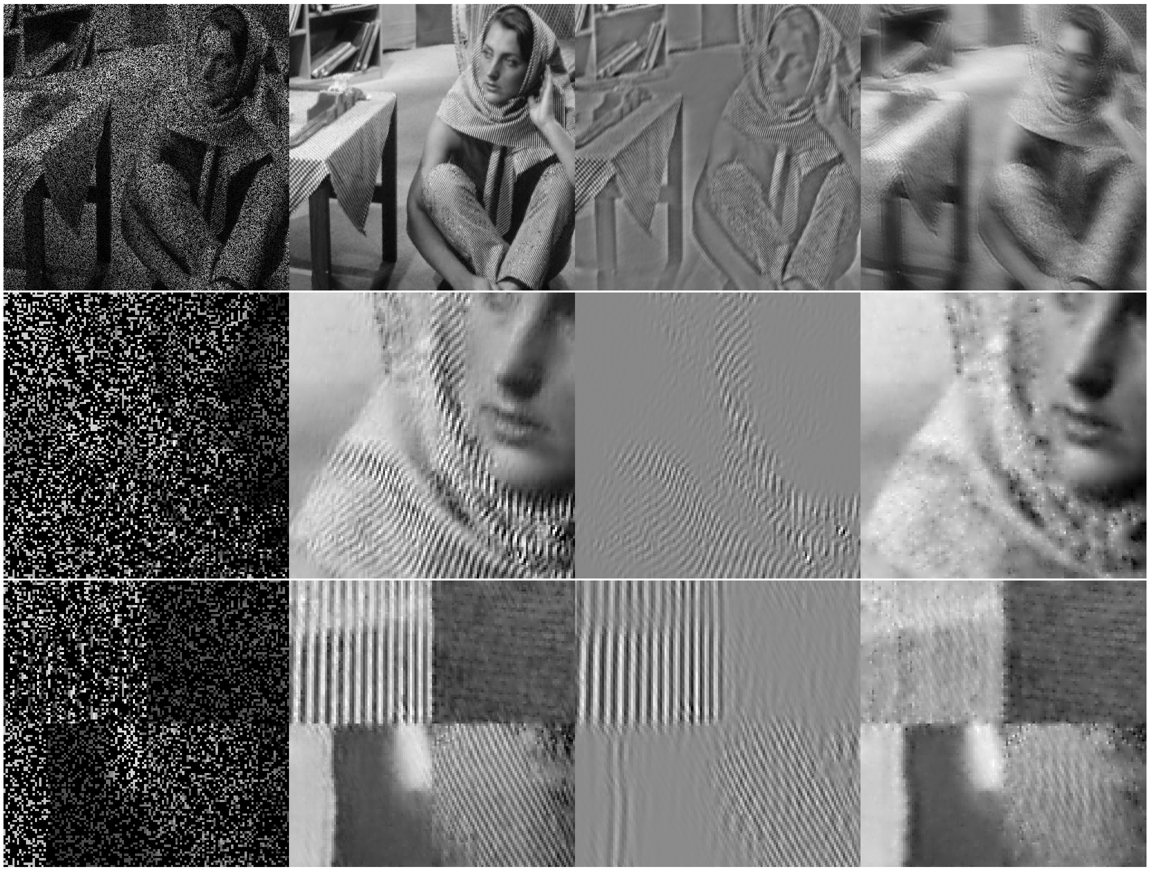
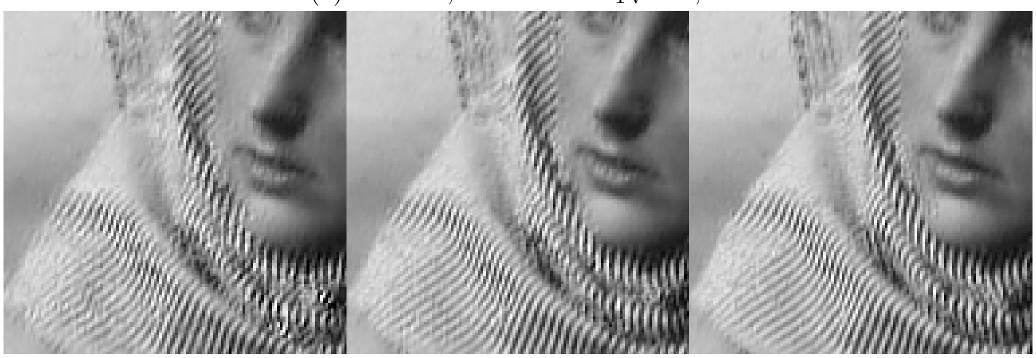


Figure 3.5: Implicit cartoon-texture decomposition of the proposed method. From left to right: Corrupted image, reconstructed image, texture part, cartoon part.

In Figure 3.6, the results of inpainting for different percentages of known pixels are displayed. As expected, as the value of p increases, the quality of the result improves as well. This can be observed among others at the right leg in the Barbara image, at the left bottom area in the Barbara crop image and at all four parts of the mix image.

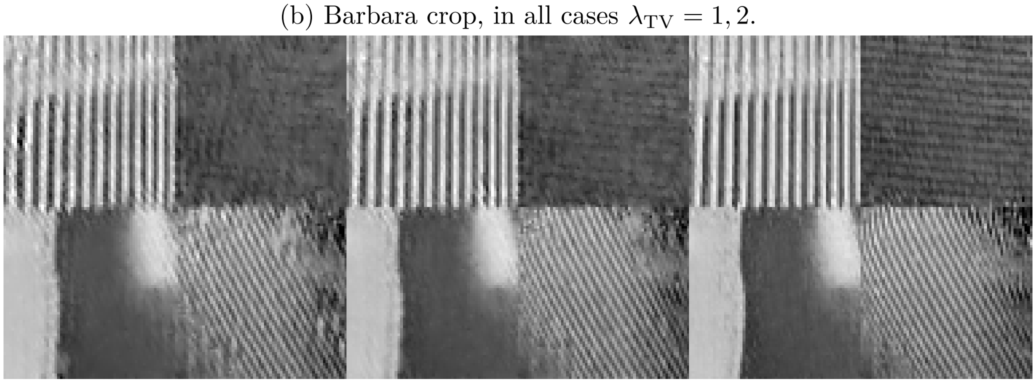
$p = 40\%$, PSNR ≈ 27.48 $p = 50\%$, PSNR ≈ 29.54 $p = 60\%$, PSNR ≈ 30.80

(a) Barbara, in all cases $\lambda_{\text{TV}} = 1,0$.



$p = 30\%$, PSNR ≈ 23.83 $p = 40\%$, PSNR ≈ 26.93 $p = 50\%$, PSNR ≈ 29.58

(b) Barbara crop, in all cases $\lambda_{\text{TV}} = 1,2$.

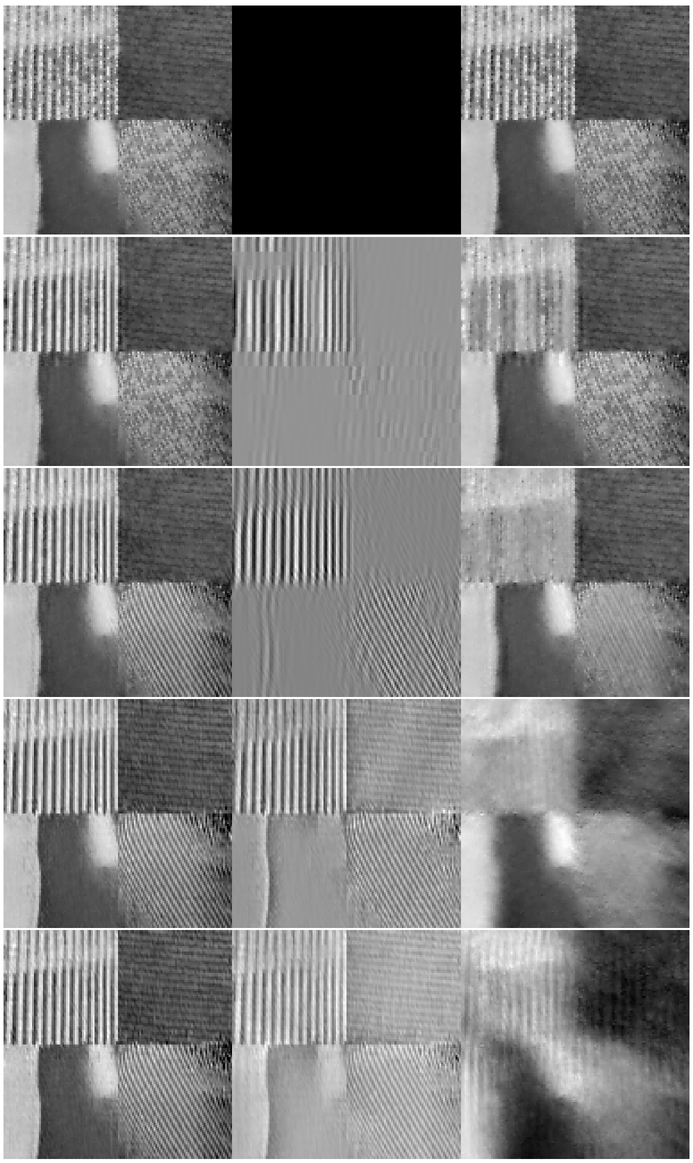


$p = 20\%$, $\lambda_{\text{TV}} = 1,2$ PSNR ≈ 22.83 $p = 30\%$, $\lambda_{\text{TV}} = 0,8$ PSNR ≈ 25.11 $p = 40\%$, $\lambda_{\text{TV}} = 1,8$ PSNR ≈ 27.14

(c) Mix

Figure 3.6: Comparison of the results of inpainting for different values of p

As a further experiment in order to gain more understanding of the proposed method, we illustrate the influence of the parameter λ_{TV} on the result of the method. In Figure 3.7, one can observe, that for small λ_{TV}, texture is penalized strongly enough, that the texture part of the image becomes zero. Therefore, we recover a similar result as for TV regularization. As λ_{TV} increases, the cartoon part becomes smoother and consequently the texture part has to be non-trivial.



$\lambda_{\mathrm{TV}} = 0.2$
PSNR ≈ 21.95

$\lambda_{\mathrm{TV}} = 0.4$
PSNR ≈ 23.84

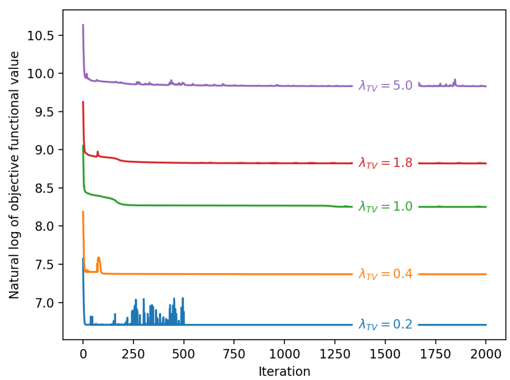
$\lambda_{\mathrm{TV}} = 1.0$
PSNR ≈ 26.56

$\lambda_{\mathrm{TV}} = 1.8$
PSNR ≈ 27.14

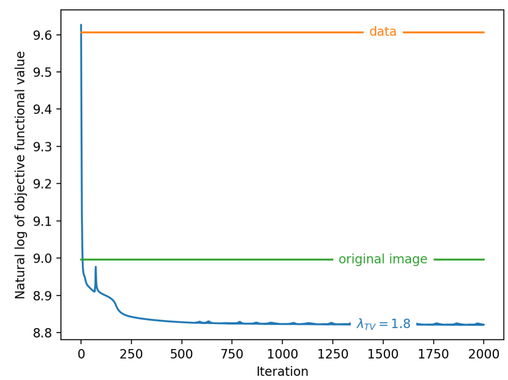
$\lambda_{\mathrm{TV}} = 5.0$
PSNR ≈ 26.60

Figure 3.7: Mix with $p = 40\%$ of known pixels and different values of λ_{TV}. From left to right: Result of the method, texture, cartoon

In Figure 3.8, we plot the objective functional over the number of iterations for the examples of Figure 3.7. Please keep in mind, that for different values of λ_{TV} the objective functional itself is different, therefore the different curves are not to be compared in absolute values, but rather qualitatively. The curves denoted 'original image' and 'data' show the objective functional value for u being the original image and data, respectively, and the texture being zero. One should not be surprised, that the original image has a larger objective functional value than the result of our method, since the original image is not decomposed into texture and cartoon properly.



Objective functional value over number of iterations for different λ_{TV}



Objective functional value for $\lambda_{\mathrm{TV}} = 1.8$ vs. objective functional for the data and the original image

Figure 3.8: Objective functional for different λ_{TV}

We will now have a closer look on the learned features, i.e., the convolution kernels and coefficients. The convolution kernels are difficult to interpret. Due to the penalty of the 1-norm in our objective functional, in the deeper layers, the kernels are highly sparse and some of them vanish entirely. But also the convolution kernels in the first layer seemed to be rather unstructured as can be observed in Figure 3.9. This might be due to the fact, that the kernels are rather small ($4x4$) and therefore not capable of containing a lot of structural information individually.

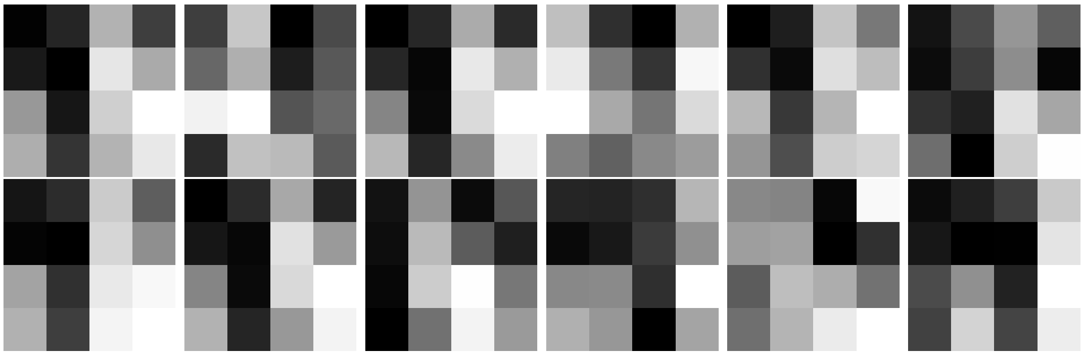
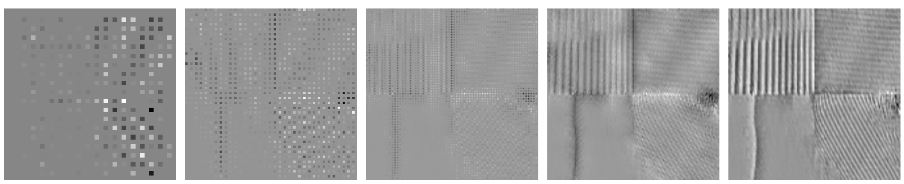
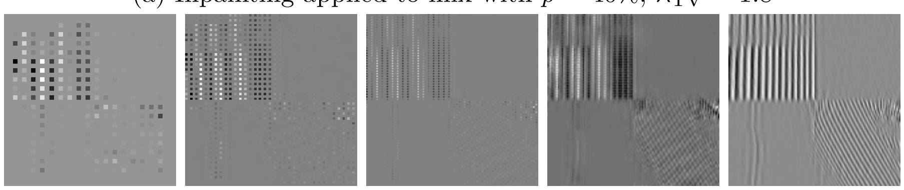


Figure 3.9: Learned kernels $(\theta_k^1)_{k=1}^{12}$ for inpainting applied to mix with $p = 40\%$, $\lambda_{\text{TV}} = 1.8$

In Figure 3.10, we show the coefficient of maximal norm of each layer, that is, for all j we display the coefficient $\mu_{i_j}^j$, where $i_j = \underset{k}{\operatorname{argmax}} \left\| \mu_k^j \right\|_1$.



(a) Inpainting applied to mix with $p = 40\%$, $\lambda_{\text{TV}} = 1.8$



(b) Inpainting applied to mix with $p = 100\%$, $\lambda_{\text{TV}} = 0.8$

Figure 3.10: Coefficient of the greatest norm of each layer. From left to right: Deepest layer to top layer

It seems in Figure 3.10, the higher the layer, the more details of the texture are developed. In the deeper layers, there are only non-zero entries at pixels, where texture is located and in the higher layers the texture itself is developed at those locations. In an effort to further visualize the information contained in the convolution kernels θ, in Figure 3.11 we show an exemplary delta response of the neural network. That is, we replace the input μ of the convolutional network resulting from our algorithm with $\delta \in \mathbb{R}^{\text{size}(\mu)}$, containing only zeros up to a single entry, that is one. Then, in Figure 3.11 for each layer j, we display the coefficient of maximal norm as before in Figure 3.10. The network used for this experiment is obtained from solving the inpainting problem with the mix image and $p = 40\%$.

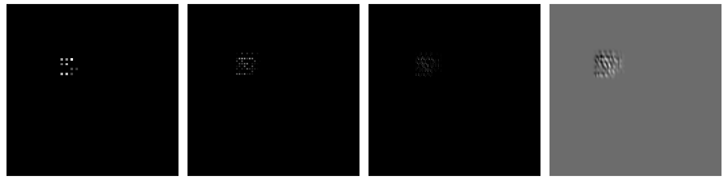


Figure 3.11: Delta response of the mix network with $p = 40\%$, $\lambda_{\mathrm{TV}} = 1.8$. From left to right: Deepest layer to top layer

Note, that in Figure 3.11 and Figure 3.10, we show the zero-interpolations of the coefficients on each layer, i.e., using the notation of Definition 3.20, we show $\tilde{\mu}$ instead of μ. Unfortunately, the delta response does not allow for a lot of interpretation. It seems, that only the superposition of several delta responses builds a meaningful texture.

Timing

Lastly, we would like to show some timing results of our method. As already mentioned in the beginning of this section, we implemented the method in Python using parallel computation with PyOpenCL. In Table 3.2, there are the computation times for the mix image with $p = 40\%$ and $\lambda_{\mathrm{TV}} = 1.8$ for different numbers of iterations. Moreover, in Figure 3.12 one can observe the corresponding results of the method. After 500 iterations, visually there is hardly any further change in the result.

Number of iterations	100	500	1000	2500	5000
Computation time in sec	8	53	127	360	698

Table 3.2: Computation times for different numbers of iterations for the mix image with $p = 40\%$ and $\lambda_{\mathrm{TV}} = 1.8$.

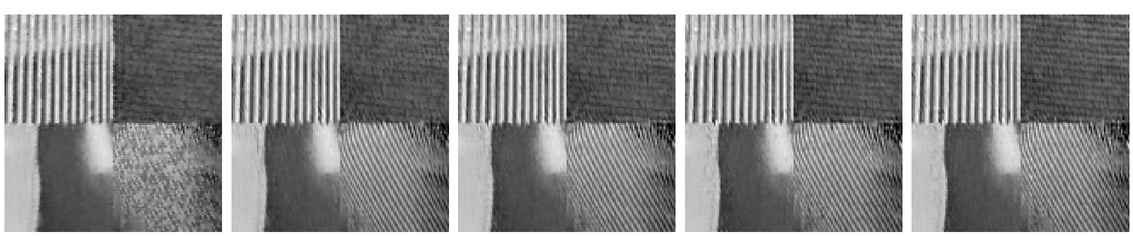


Figure 3.12: Results for different numbers of iterations for the mix image with $p = 40\%$ and $\lambda_{\mathrm{TV}} = 1.8$. From left to right, number of iterations is: $100, 500, 1000, 2500, 5000$.

Chapter 4

Discussion

In the present thesis, we have developed and analysed a novel method for solving inverse problems in image processing. Our method aims to combine the strengths of conventional methods, such as total variation based image reconstruction, and more recent deep learning methods. The main contribution of the present work compared to most publications about deep learning in image processing is the introduction of the model in a continuous setting. Employing functional analytic results, we could prove well posedness of the proposed method. We have also introduced our method in a discrete setting and applied it to test images using a state of the art optimization algorithm. Although the results are promising and outperform simple total variation based methods, there is still potential for improvement. First of all, the numerical implementation suffered from instability in some cases. Especially in the case of larger sized neural networks, it sometimes happened, that the used algorithm did not converge to a meaningful result. Unfortunately, sometimes this happened despite the previous iterates already looking very promising. As possible remedies to this issue we propose either to find better initial values for the algorithm, or to change the data fidelity term. Using appropriate initial values for the algorithm is a reasonable idea, since the considered objective functional is non-convex and may therefore have several local minima. Hence, in order to find a good solution, we should start the algorithm with an initial value, that is somehow close to this solution. Changing the data fidelity term to e.g. a L2-norm type functional is inspired by the results presented in [15] and by our stability result Theorem 2.84 and could improve the convergence of the algorithm due to increased smoothness of the objective functional. Further, the results of the proposed method are still inferior compared to some recent deep learning methods, such as [15]. However, we would like to point out, that the slightly decreased performance can be seen as a trade off for the gain of an underlying mathematical foundation of our method and we aim to decrease this discrepancy in performance in the future. For further research we propose extending our method to more complex convolutional networks, possibly employing also activation functions and skip connections etc., in order to improve the visual results of the method. This extension should be performed, without losing the solid underlying mathematical foundation in the form of an understood continuous model. Further, it would be interesting, to try to replace the functional J from $(P(y_0))$ with a functional like the total generalized variation, in order to obtain an even richer model for the cartoon part as done in [9]. Finally, an even more extensive parameter determination could possibly yield better results.

Appendix A

Auxiliary Results

Lemma A.1. *Let* $1 \leq p < \infty$ *and* $u \in L^p(\mathbb{R}^d)$*. Then*

$$\|u(.+h) - u\|_p \to 0$$

as $|h| \to 0$*.*

Proof. We will prove the desired result, by showing it for smooth functions first and extending the result via density of smooth functions in $L^p(\mathbb{R}^d)$. So let $\phi \in C_c^\infty(\mathbb{R}^d)$ and $\Omega \subset \mathbb{R}^d$ bounded and open, such that $\operatorname{supp}(\phi) \Subset \Omega$. As a continuous function on compact support, ϕ is bounded. Denoting $\Omega_\delta = \{x + v \mid x \in \Omega,\ |v| < \delta\}$ and χ_{Ω_δ} the characteristic function on Ω_δ, we find for $x \in \mathbb{R}^d$ and $|h| < \delta$

$$|\phi(x+h) - \phi(x)|^p \leq (2\,\|\phi\|_\infty)^p \chi_{\Omega_\delta}(x) \qquad \in L^1(\mathbb{R}^d).$$

Moreover, $|\phi(x+h) - \phi(x)|^p \to 0$ as $|h| \to 0$ for all $x \in \mathbb{R}^d$ by continuity. By Lebesgue's dominated convergence theorem, we find

$$\lim_{|h|\to 0} \int_{\mathbb{R}^d} |\phi(x+h) - \phi(x)|^p \; dx = 0,$$

i.e., $\|\phi(.+h) - \phi\|_p \to 0$ as $|h| \to 0$. Now for general $u \in L^p(\mathbb{R}^d)$, by density we can find a sequence $(\phi_n)_n \subset C_c^\infty(\mathbb{R}^d)$, converging to u in $L^p(\mathbb{R}^d)$. Let $\epsilon > 0$ arbitrary. Choose $n_0 \in \mathbb{N}$, such that $\|u - \phi_{n_0}\|_p < \frac{\epsilon}{4}$. For n_0 fixed, choose $\delta > 0$, such that for $|h| < \delta$, $\|\phi_{n_0}(.+h) - \phi_{n_0}\|_p < \frac{\epsilon}{2}$, which is possible, as we have shown above. Then for $|h| < \delta$, we find

$$\begin{aligned}\|u(.+h) - u\|_p &\leq \|u(.+h) - \phi_{n_0}(.+h)\|_p + \|\phi_{n_0}(.+h) - \phi_{n_0}\|_p + \|\phi_{n_0} - u\|_p \leq \\ &\leq 2\,\|u - \phi_{n_0}\|_p + \|\phi_{n_0}(.+h) - \phi_{n_0}\|_p < \epsilon.\end{aligned}$$

□

Lemma A.2. *Let X be a topological space, $F : X \to [0, \infty)$ (weakly) lsc, and $\phi : [0, \infty) \to [0, \infty)$ continuous and monotonically increasing. Then the composition $\phi \circ F : X \to [0, \infty)$ is (weakly) lsc.*

Proof. Let $(x_n)_n \subset X$ be a sequence converging (weakly) to $x \in X$. First we note, that for all $n \in \mathbb{N}$ and all $l \geq n$,

$$\phi(\inf_{k \geq n} F(x_k)) \leq \phi(F(x_l)),$$

due to monotonicity of ϕ. Therefore, by taking the infimum over l on the right-hand side, we obtain

$$\phi(\inf_{k \geq n} F(x_k)) \leq \inf_{k \geq n} \phi(F(x_k)). \tag{A.1}$$

We distinguish two cases. Assume first, that $\liminf_{n\to\infty} F(x_n) = \infty$, i.e., $(\inf_{k \geq n} F(x_k))_n$ is an increasing sequence, converging two ∞. Hence, for n sufficiently large, it will be true, that $\inf_{k \geq n} F(x_k) \geq F(x)$. For such n, due to monotonicity of ϕ, we see, that

$$\phi(F(x)) \leq \phi(\inf_{k \geq n} F(x_k)) \underbrace{\leq}_{\text{(A.1)}} \inf_{k \geq n} \phi(F(x_k))$$

and as a consequence

$$\phi(F(x)) \leq \lim_{n\to\infty} \inf_{k \geq n} \phi(F(x_k)) = \liminf_{n\to\infty} \phi(F(x_k)).$$

In the case, that $\liminf_{n\to\infty} F(x_n) < \infty$, we can compute

$$\begin{aligned}\phi \circ F(x) = \phi(F(x)) \quad &\underbrace{\leq}_{\substack{F \text{ lsc,} \\ \phi \text{ increasing}}} \quad \phi(\liminf_{n\to\infty} F(x_n)) = \phi(\lim_{n\to\infty} \inf_{k \geq n} F(x_k)) = \\ &\underbrace{=}_{\phi \text{ cont}} \lim_{n\to\infty} \phi(\inf_{k \geq n} F(x_k)) \underbrace{\leq}_{\text{(A.1)}} \lim_{n\to\infty} \inf_{k \geq n} \phi(F(x_k)) = \liminf_{n\to\infty} \phi(F(x_n)).\end{aligned}$$

□

Lemma A.3. *[Composition of projections] Let $d \in \mathbb{N}$ and consider $\mathbb{R}^d$ equipped with the standard scalar product and the induced Euclidean norm. Let $W \subseteq \mathbb{R}^d$ be a linear subspace of $\mathbb{R}^d$ and $B \subseteq \mathbb{R}^d$ the closed unit ball in $\mathbb{R}^d$. Let further $P_W : \mathbb{R}^d \to W$ and $P_B : \mathbb{R}^d \to B$ be the projections onto W and B respectively and $P_{W\cap B} : \mathbb{R}^d \to W \cap B$ the projection onto $W \cap B$. Then it holds true, that*

$$P_{W\cap B} = P_B P_W.$$

Proof. Firstly, we note that all the mentioned projections exist, since B is a closed and convex set and W is a vector space, so both sets as well as their intersection admit a projection. For $v \in \mathbb{R}^d$

$$P_B v = \frac{v}{\max\{\|v\|_2, 1\}}.$$

Hence, P_B is a multiplication with a scalar function. Therefore, for any $v \in \mathbb{R}^d$, $P_B P_W v \in B$, since P_B maps to B and $P_B P_W v \in W$, since $P_W v \in W$ and W is a linear vector space, ergo multiplication with a scalar does not map out of W. So $P_B P_W v \in W \cap B$. Now let $v \in \mathbb{R}^d$ and $w \in W \cap B$ arbitrary. We compute

$$\begin{aligned} \|w - v\|_2^2 &= \|w - P_B P_W v + P_B P_W v - v\|_2^2 = \\ &= \|w - P_B P_W v\|_2^2 + \|P_B P_W v - v\|_2^2 + 2\langle w - P_B P_W v, P_B P_W v - v\rangle = \\ &= \|w - P_B P_W v\|_2^2 + \|P_B P_W v - v\|_2^2 \\ &\quad -2\underbrace{\langle w - P_B P_W v, v - P_W v\rangle}_{\text{(i)}} -2\underbrace{\langle w - P_B P_W v, P_W v - P_B P_W v\rangle}_{\text{(ii)}} \geq \\ &\geq \|w - P_B P_W v\|_2^2 + \|P_B P_W v - v\|_2^2. \end{aligned}$$

The presented inequality may need some justification. Expression (i) is equal to zero since $w - P_B P_W v \in W$ and $v - P_W v \perp W$ since P_W is an orthogonal projection and (ii) is less than or equal to zero due to the basic properties of projections onto convex, closed sets. As a result, we see that

$$\|P_B P_W v - v\|_2^2 = \min\left\{\|w - v\|_2^2 \mid w \in B \cap W\right\},$$

proving, that $P_B P_W v = P_{W \cap B} v$. □

Bibliography

[1] Aviad Aberdam, Jeremias Sulam, and Michael Elad. Multi-layer sparse coding: The holistic way. SIAM Journal on Mathematics of Data Science, 1(1):46–77, 2019.

[2] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on signal processing, 54(11):4311–4322, 2006.

[3] Hans Wilhelm Alt. Lineare Funktionalanalysis. Springer, 2006.

[4] Micol Amar, Virginias De Cicco, and Nicola Fusco. Lower semicontinuity and relaxation results in bv for integral functionals with bv integrands. ESAIM: Control, Optimisation and Calculus of Variations, 14:456–477, 21 2007.

[5] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. Functions of Bounded Variation and Free Discontinuity Problems. Oxford Mathematical Monographs, 2000.

[6] Daniel Otero Baguer, Johannes Leuschner, and Maximilian Schmidt. Computed tomography reconstruction using deep image prior and learned reconstruction methods, 2020.

[7] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Mathematical Programming, 146(1):459–494, 2014.

[8] Kristian Bredies and Dirk Lorenz. Mathematische Bildverarbeitung. Vieweg Teubner, 2011.

[9] A. Chambolle, M. Holler, and T. Pock. A convex variational model for learning convolutional image atoms from incomplete data. Journal of Mathematical Imaging and Vision, 62(3):417–444, 2020.

[10] A. Chambolle, M. Holler, and T. Pock. A convex variational model for learning convolutional image atoms from incomplete data. `https://github.com/hollerm/convex_learning`, 2020.

[11] Mysid Dake. A simplified view of an artifical neural network. https://commons.wikimedia.org/w/index.php?curid=1412126, 2006. Accessed: 2020-07-28.

[12] Sören Dittmer, Tobias Kluth, Peter Maass, and Daniel Otero Baguer. Regularization by architecture: A deep prior approach for inverse problems. Journal of Mathematical Imaging and Vision, pages 1–15, 2019.

[13] Ivar Ekeland, Roger Temam, and Diego Pallara. Convex Analysis and Variational Problems. Siam, 1987.

[14] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image processing, 15(12):3736–3745, 2006.

[15] V. Lempitsky, A. Vedaldi, and D. Ulyanov. Deep image prior. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9446–9454, 2018.

[16] V. Lempitsky, A. Vedaldi, and D. Ulyanov. Deep image prior. https://github.com/DmitryUlyanov/deep-image-prior, 2018.

[17] Boris S. Mordukhovich. Variational Analysis and Generalized Differentiation I. Springer, 2006.

[18] Daniel Obmann, Johannes Schwab, and Markus Haltmeier. Deep synthesis regularization of inverse problems. arXiv preprint arXiv:2002.00155, 2020.

[19] Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad. Convolutional dictionary learning via local processing. In Proceedings of the IEEE International Conference on Computer Vision, pages 5296–5304, 2017.

[20] Thomas Pock and Shoham Sabach. Inertial proximal alternating linearized minimization (ipalm) for nonconvex and nonsmooth problems. SIAM Journal on Imaging Sciences, 9(4):1756–1787, Jan 2016.

[21] Jeremias Sulam, Aviad Aberdam, Amir Beck, and Michael Elad. On multi-layer basis pursuit, efficient algorithms and convolutional neural networks. IEEE transactions on pattern analysis and machine intelligence, 2019.

[22] Jeremias Sulam, Vardan Papyan, Yaniv Romano, and Michael Elad. Multilayer convolutional sparse modeling: Pursuit and dictionary learning. IEEE Transactions on Signal Processing, 66(15):4090–4104, 2018.

[23] Dave Van Veen, Ajil Jalal, Mahdi Soltanolkotabi, Eric Price, Sriram Vishwanath, and Alexandros G. Dimakis. Compressed sensing with deep image prior and learned regularization, 2018.

[24] Dirk Werner. Funktionalanalysis. Springer, 2011.

[25] Brendt Wohlberg. Efficient algorithms for convolutional sparse representations. IEEE Transactions on Image Processing, 25(1):301–315, 2015.

[26] He Zhang and Vishal M Patel. Convolutional sparse coding-based image decomposition. In BMVC, 2016.