Convergence rates and structure of solutions of inverse problems with imperfect forward models

Martin Burger¹, Yury Korolev²,⁴ and Julian Rasch³

¹ Department Mathematik, University of Erlangen-Nürnberg, Cauerstrasse 11, 91058 Erlangen, Germany
² Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom
³ Institute for Analysis and Numerics, University of Münster, Einsteinstr. 62, 48149 Münster, Germany

E-mail: martin.burger@fau.de, y.korolev@damtp.cam.ac.uk and julian.rasch@wwu.de

Received 26 June 2018, revised 19 November 2018
Accepted for publication 7 December 2018
Published 3 January 2019

Abstract
The goal of this paper is to further develop an approach to inverse problems with imperfect forward operators that is based on partially ordered spaces. Studying the dual problem yields useful insights into the convergence of the regularised solutions and allow us to obtain convergence rates in terms of Bregman distances—as usual in inverse problems, under an additional assumption on the exact solution called the source condition. These results are obtained for general absolutely one-homogeneous functionals. In the special case of TV-based regularisation we also study the structure of regularised solutions and prove convergence of their level sets to those of an exact solution. Finally, using the developed theory, we adapt the concept of debiasing to inverse problems with imperfect operators and propose an approach to pointwise error estimation in TV-based regularisation.

Keywords: operator uncertainty, total variation, extended support, Bregman distances, source condition, error estimation, debiasing

(Some figures may appear in colour only in the online journal)

1. Introduction

Inverse problems are typically concerned with the interpretation of indirect measurements. The measurable data \( f \) are typically connected to the quantities of interest \( u \) through some

⁴ Author to whom any correspondence should be addressed.
forward operator or forward model $A$ that models the data acquisition process. To obtain the quantities of interest $u$ from the data $f$, we need to invert this forward model. Since the inverse of $A$ is typically not continuous, the inversion is ill-posed and one needs to employ regularisation to obtain a stable approximation to $u$. Variational regularisation is a common approach to solving ill-posed problems and consists in minimising a weighted sum of a data fidelity term enforcing closeness to the measured data and a regularisation term enforcing some regularity of the reconstructed solution.

In this paper we consider inverse problems in form of an ill-posed operator equation

$$ Au = f, $$

(1.1)

where $A : L^1(\Omega) \rightarrow L^\infty(\Omega)$ is a linear operator and $\Omega \subset \mathbb{R}^m$ is a bounded domain. We assume that there exists a non-negative solution of (1.1).

For an appropriate functional $J(\cdot) : L^1 \rightarrow \mathbb{R}_+ \cup \{\infty\}$ we consider non-negative $J$-minimising solutions, which solve the following problem:

$$ \min_{u \in L^1 : u \geq 0} J(u) \quad \text{s.t.} \quad Au = f. $$

(1.2)

We assume that the feasible set in (1.2) has at least one point with a finite value of $J$ and denote a (possibly non-unique) solution of (1.2) by $\bar{u}_J$. Throughout this paper it is assumed that the regularisation functional $J(\cdot)$ is convex, proper and absolutely one-homogeneous.

In practice the data $f$ are not known precisely and only their perturbed version $\tilde{f}$ is available. In this case, we cannot simply replace the constraint $Au = f$ in (1.2) with $Au = \tilde{f}$, since the solutions of the original problem (1.1) would no longer be feasible in this case. Therefore, we need to relax the equality in (1.2) to guarantee the feasibility of solutions of the original problem (1.1). This is the idea of the residual method [20, 23]. If the error in the data is bounded by some known constant $\delta$, the residual method accounts to solving the following constrained problem:

$$ \min_{u \in L^1} J(u) \quad \text{s.t.} \quad ||Au - \tilde{f}|| \leq \delta. $$

(1.3)

The fidelity function becomes in this case the characteristic function of the convex set $\{u : ||Au - \tilde{f}|| \leq \delta\}$. In the linear case, the residual method is equivalent to Tikhonov regularisation

$$ \min_{u \in L^1} ||Au - \tilde{f}||^2 + \alpha J(u) $$

(1.4)

with the regularisation parameter $\alpha = \alpha(\tilde{f}, \delta)$ chosen according to Morozov’s discrepancy principle [23].

In many practical situations not only the data contain errors, but also the forward operator, that generated the data, are not perfectly known. In order to guarantee the feasibility of solutions of the original problem (1.1) in the constrained problem (1.3), one needs to account for the errors in the operator in the feasible set. If the errors in the operator are bounded by a known constant $h$ (in the operator norm), the feasible set can be amended as follows in order to guarantee feasibility of the solutions of the original problem (1.1):

$$ \min_{u \in L^1} J(u) \quad \text{s.t.} \quad ||\tilde{A}u - \tilde{f}|| \leq \delta + h||u||. $$

(1.5)

where $\tilde{A}$ is the noisy operator. This optimisation problem is non-convex and therefore presents considerably more computational challenges than its counterpart with the exact operator (1.3). Thus, in the context of the residual method, uncertainty in the operator results in a qualitative change in the optimisation problem to be solved, which, in general, requires using different
numerical approaches from those in (1.3). The reason for non-convexity is the fact that we used the operator norm to quantify the error in the operator. An alternative approach was proposed in [26]. Instead of the operator norm, it uses intervals in an appropriate partial order to quantify the error in the operator. It assumes that, instead of only one instance of approximate data \( \tilde{f} \) and approximate operator \( \tilde{A} \), lower and upper bounds for them are available, i.e. \( f^l, f^u \in L^\infty \) and \( A^l, A^u : L^1 \to L^\infty \) such that

\[
 f^l \leq f \leq f^u, \quad A^l \leq A \leq A^u. \tag{1.6}
\]

The first two inequalities are understood in the sense of partial order in \( L^\infty \) and the last two in the sense of partial order for linear operators \( L^1 \to L^\infty \) (more details on how partial order is defined for linear operators will be given in section 2.1).

Using the bounds (1.6), the residual method can be reformulated as the following optimisation problem:

\[
 \min_{u \in L^1 : u \geq 0} J(u) \quad \text{s.t.} \quad A^l u \leq f^u, \quad A^u u \geq f^l. \tag{1.7}
\]

This optimisation problem is convex and has the same structure in the case of errors in the operator as in the error-free case. The fidelity term in this case is the characteristic function of a convex polyhedron. It can be easily verified that any solution of the original problem (1.1) is a feasible solution of (1.7).

In this paper, we study the dual problem of (1.7), which can be written as follows

\[
 \max_{\lambda, \mu \geq 0} (\mu, -\varphi) \quad \text{s.t.} \quad \lambda - B^* \mu \in \partial J(0), \tag{1.8}
\]

where \((\cdot, \cdot)\) denotes the duality pairing between \( L^1 \) and \( L^\infty \), \( \varphi = \begin{pmatrix} f^u \\ -f^l \end{pmatrix} \) is the adjoint of \( B \) and \( \partial J(0) \subset L^\infty(\Omega) \) is the subdifferential of the regularisation functional at zero. We shall see that, under certain assumptions, \( \lambda \) and \( \mu \) in (1.8) are Lagrange multipliers corresponding to the positivity constraint and the constraints \( A^l u \leq f^u \) and \( A^u u \geq f^l \) in (1.7), respectively, and \( p = \lambda - B^* \mu \) is a subgradient of the regulariser \( J \) at the optimal solution of (1.7).

To study the convergence of the minimisers of (1.7) to a solution of (1.1), we need some notion of convergence of the bounds \( f^l, f^u \) and \( A^l, A^u \) to the exact data \( f \) and operator \( A \), respectively. For this purpose, we consider sequences of lower and upper bounds \( f^l_n, f^u_n \) and \( A^l_n, A^u_n \) such that

\[
 f^l_n \leq f \leq f^u_n, \quad f^l_{n+1} \geq f^l_n, \quad f^u_{n+1} \leq f^u_n, \quad A^l_n \leq A \leq A^u_n, \quad A^l_{n+1} \geq A^l_n, \quad A^u_{n+1} \leq A^u_n \tag{1.9}
\]

and

\[
 \|f^u_n - f^l_n\| \to 0, \quad \|A^u_n - A^l_n\| \to 0. \tag{1.11}
\]

With these sequences of bounds, we obtain a sequence of optimisation problems

\[
 \min_{u \in L^1 : u \geq 0} J(u) \quad \text{s.t.} \quad A^l_n u \leq f^u_n, \quad A^u_n u \geq f^l_n. \tag{1.12}
\]

It was shown in [26] (see also theorem 5 in section 3) that the minimisers \( u_n \) of (1.12) converge to \( u_J \) as \( n \to \infty \). In this paper we study this convergence in more detail, ultimately aiming at obtaining convergence rates.

It is well-known [3] that solutions of the dual problem play an important role in establishing convergence rates, therefore we study the behaviour of the dual problem as \( n \to \infty \) in...
more detail. Uncertainty in the operator results in a perturbation of the feasible set of the dual problem (1.8). In order to ensure the convergence of its solutions, we would like to know that the solution of the dual problem is stable with respect to such perturbations. Stability theory for optimisation problems with perturbations [4] emphasises the role of the so-called Robinson regularity [30, 31] in the stability of the solution under perturbations of the feasible set. In our particular case a condition on the interior of $\partial J(0)$ (see assumption 4) plays a crucial role in the stability of the dual problem.

Establishing the stability of the dual problem allows us to relate its solutions to solutions of the dual problem in the limit case of exact data and operator, which has a very similar form to (1.8):

$$\max_{\lambda, \mu \geq 0} \left( \mu, -\varphi \right) \quad \text{s.t.} \quad \lambda - \bar{B}^* \mu \in \partial J(0),$$

(1.13)

where $\varphi = \left( f - f_0 \right)$ and $\bar{B} = \left( \begin{array}{c} A \\ -A \end{array} \right)$. If the original problem (1.1) is ill-posed, existence of such limit solutions of the dual problem (1.13) cannot be guaranteed, unless additional assumptions on the exact solution are made, known as the source condition [10], which in our case takes the form (3.10). Under the source condition we are able to prove uniform boundedness of the Lagrange multipliers and convergence of the subgradient, which allows us to establish convergence rates (section 3.8). For the symmetric Bregman distance [25] between the minimisers $J(u_n)$ of (1.12) and any $J$-minimising solution $\bar{u}$ we obtain the following estimate

$$D_{\bar{J}}(u_n, \bar{u}) \leq C \cdot \varepsilon_n,$$

(1.14)

where $\|f_n - f_0\| = O(\varepsilon_n)$ and $\|A_n^u - A_n^l\| = O(\varepsilon_n)$. These convergence rates coincide with those from [3] for problems with exact operators, providing an interface with existing theory.

We further investigate the solutions of problem (1.7) by studying their geometric properties in the spirit of [14]. In particular, we prove Hausdorff convergence of the level sets of TV-regularised solutions to those of the exact solution. However, unlike the original paper [14], we cannot use $J(\cdot) = TV(\cdot)$, since it does not guarantee stability of the dual problem and convergence of the subgradient. Instead, we use the full (weighted) BV norm, choosing $J(u) = TV(u) + \gamma \|\cdot\|_1$, $\gamma > 0$.

Our numerical experiments with deblurring demonstrate that reconstructions obtained with $J(\cdot) = TV(\cdot) + \gamma \|\cdot\|_1$, $\gamma > 0$, are, indeed, piecewise-constant (if so is the ground truth), while $J(\cdot) = TV(\cdot)$ misses some jumps and results in smoother reconstructions. This is surprising, since it contradicts the typical behaviour of TV in ROF-type models [32], which is known as staircasing [24, 29]. The reason for this is the additional freedom provided by our constraint-based approach. While in classical ROF-denosing zero is in the subgradient of TV only if the minimiser is equal to the data (which rarely happens for noisy data), the constraint-based approach allows the subgradient to be zero whenever the noise is small enough and contained within a prescribed corridor around the true data. However, with $J(\cdot) = TV(\cdot) + \gamma \|\cdot\|_1$, $\gamma > 0$, whenever the subgradient of $J$ is zero, the subgradient of TV is equal to $-\gamma$, forcing the reconstruction to be piecewise-constant.

Finally, we use the developed theory to adopt the concept of two-step debiasing [9, 15, 16], which allows to reduce the systematic bias in the reconstruction, such as loss of contrast, to our framework. We also propose a method of obtaining asymptotic pointwise lower and upper bounds of TV-regularised solutions in areas, where the exact solution is piecewise-constant.

The paper is organised as follows. In section 2 we introduce the primal and the dual problems for fixed bounds $f^l, f^u, A^l, A^u$ and study their properties. In section 3 we present the convergence analysis and establish convergence rates. In section 4 we study geometric properties.
of TV-regularised solutions and prove Hausdorff convergence of the level sets. In section 5 we describe our approach to debiasing and pointwise error estimation and in section 6 we present the results of our numerical experiments. The appendices contain some results of more technical nature that we need for the proofs.

2. Primal and dual problems

In order to accurately formulate the primal problem (1.7), we briefly recall some definitions from the theory of functional spaces with partial order.

2.1. Banach lattices

$L_p$ spaces, endowed with a partial order relation

$$f \leq g \text{ iff } f(\cdot) \leq g(\cdot) \text{ a.e.},$$

become Banach lattices, i.e. partially ordered Banach spaces with well-defined suprema and infima of each pair of elements and a monotone norm [35]. The set \( \{ f \in L^p : f \geq 0 \} \) is called the positive cone. It can be shown that the interior of the positive cone in an \( L^p \) space is empty, unless \( p = \infty \) [34, 35].

Partial orders in \( L^p \) and \( L^q \) induce a partial order in a subspace of the space of linear operators acting from \( L^p \) to \( L^q \), namely in the space of regular operators. A linear operator \( A : L^p \to L^q \) is called regular, if it can be represented as a difference of two positive operators. An operator \( A \) is called positive and we write \( A \geq 0 \) iff \( \forall x \in L^p, \ x \geq 0 \Rightarrow Ax \geq 0 \). Partial order in the space of regular operators is introduced as follows: \( A \geq B \) iff \( A - B \) is a positive operator. Every regular operator acting between two Banach lattices is continuous [35].

2.2. Primal and dual problems

In this section we study in more detail the optimisation problem (1.7) and its dual (1.8). For convenience, we repeat problem (1.7) here:

$$\min_{u \in L^1 : u \geq 0} \mathcal{J}(u) \quad \text{s.t. } A'u \leq f^u, \ A^u u \geq f^l.$$

(2.1)

In order to simplify notation, we introduce

$$B = \begin{pmatrix} A' \\ -A^u \end{pmatrix} \quad \text{and} \quad \bar{B} = \begin{pmatrix} A \\ -A \end{pmatrix}$$

as well as

$$\varphi = \begin{pmatrix} f^u \\ -f_l \end{pmatrix} \quad \text{and} \quad \bar{\varphi} = \begin{pmatrix} f \\ -f \end{pmatrix}.$$
Denote a (possibly non-unique) minimiser of (2.1) by \( \tilde{u} \).

Now let us turn to the dual problem of (2.1). The Lagrange function is given by the following expression

\[
L(u, \lambda, \mu) = J(u) + (\mu, Bu - \varphi) - (\lambda, u),
\]

where \( \lambda \in L^\infty, \mu \in (L^\infty)^* \times (L^\infty)^*, \lambda, \mu \geq 0 \). Taking the minimum in \( u \), we obtain the following expression for the dual objective:

\[
\min_{u \in L^1} L(u, \lambda, \mu) = -J^*(\lambda - B^* \mu) - (\mu, \varphi),
\]

where \( J^*(\cdot) \) is the convex conjugate of \( J(\cdot) \). Since we assumed that \( J(\cdot) \) is absolutely one-homogeneous, we have that \( J^*(\cdot) \) is the characteristic function of \( \partial J(0) \). We discuss the properties of absolutely one-homogeneous regularisation functionals in more detail in appendix A. Hence we obtain the following formulation of the dual problem:

\[
\max_{\lambda, \mu \geq 0} (\mu, -\varphi) \quad \text{s.t.} \quad \lambda - B^* \mu \in \partial J(0). \tag{2.2}
\]

We will mainly consider regularisation functionals as functionals in \( L^1 \) (and not, for example, in \( BV \)), \( \partial J(0) \) will therefore be understood as a subset of \( L^\infty \) (and not \( BV^* \)), with exceptions denoted by a subscript \( \partial J_0(0) \), where \( \mathcal{V} \) is the corresponding subspace. Properties of \( \partial J(0) \) for regularisation functionals \( J : L^1 \to \mathbb{R}_+ \cup \{ \infty \} \) of the type \( J(\cdot) = TV(\cdot) \) and \( J(\cdot) = TV(\cdot) + \gamma \| \cdot \|_1, \gamma > 0 \), (where \( TV \) may be replaced with a similar regularisation functional, such as \( TGV \)) will be discussed in appendix B.

The following characterisation [13] of the subdifferential of an absolutely one-homogeneous functional will be useful for us later:

\[
\partial J(u) = \{ p \in L^\infty : J(v) \geq (p, v) \ \forall v \in L^1, \ J(u) = (p, u) \}. \tag{2.3}
\]

In particular, for \( u = 0 \) we get

\[
\partial J(0) = \{ p \in L^\infty : J(v) \geq (p, v) \ \forall v \in L^1 \}. \tag{2.4}
\]

Clearly, the set \( \partial J(0) \) is nonempty, convex and closed, although it may be unbounded.

### 2.3. Robinson regularity

We would like to establish strong duality between (2.1) and (2.2). To do this, we need to recall a concept from optimisation theory called **Robinson regularity**.

Consider an optimisation problem

\[
\min_{x \in C} f(x) \quad \text{s.t.} \ G(x) \in K, \tag{2.5}
\]

where \( C \subset X \) is a closed and convex set, \( G : X \to Y \) is continuously Fréchet differentiable, \( X \) and \( Y \) are Banach spaces and \( K \) is a closed convex subset of \( Y \). We say that the Robinson regularity condition [21] is satisfied at \( x_0 \) in problem (2.5) if

\[
0 \in \text{int}(G(x_0) + G(x_0)(C - x_0) - K). \tag{2.6}
\]

The next result [4, theorem 4.2] demonstrates the role that Robinson regularity plays in the existence of the Lagrange multipliers associated with the constraint \( G(x) \in K \).

**Proposition 1.** Suppose that

- problem (2.5) is convex;
- its optimal value is finite;
• \( G(x) \) is continuously differentiable and
• Robinson regularity condition is satisfied in (2.5).

Then

• strong duality holds between problem (2.5) and its dual;
• the set of optimal solutions of the dual problem is non-empty and bounded;
• if the set of Lagrange multipliers is nonempty for an optimal solution of (2.5), then it is the same for any other optimal solution of (2.5) and coincides with the set of optimal solutions of the dual problem.

Robinson regularity also plays an important role in the stability of problem (2.5) under small perturbations in \( G(\cdot) \). Consider a perturbation of the form \( G(x) + u \in K \). Denote by \( S(u) \) the feasible set in the perturbed problem. The following result holds [4, proposition 3.3].

**Proposition 2.** Suppose that Robinson condition (2.6) holds in (2.5) at \( x_0 \). Then for every \((x, u)\) in a neighbourhood of \((x_0, 0)\) one has

\[
\text{dist}(x, S(u)) = O(\text{dist}(G(x) + u, K)),
\]

where \( \text{dist}(x, S(u)) := \inf_{\xi \in S(u)} \| x - \xi \| \) is the distance from \( x \in X \) to the set \( S(u) \subset X \) and \( \text{dist}(G(x) + u, K) := \inf_{\eta \in K} \| G(x) + u - \eta \| \) is the distance from \( G(x) + u \in Y \) to \( K \subset Y \).

This stability result will play an important role in establishing the boundedness of solutions of the dual problem in sections 3.4–3.6.

2.4. Relationship between the primal and the dual problems

Our aim in this section is to show that Robinson condition (2.6) holds for the primal problem (2.1). This will ensure existence of the Lagrange multipliers and strong duality between (2.1) and (2.2).

In our case, the function \( G \) from (2.5) is linear, \( G(u) = Bu - \varphi \), the set \( K \) is the non-positive cone \( K \leq 0 := \{ \psi \in L^\infty : \psi \leq 0 \} \) and \( C \) is the non-negative cone \( C = \{ u \in L^1 : u \geq 0 \} \). Since the constraint is linear, the Robinson condition (2.6) can be written as follows:

\[
0 \in \text{int}\{ BC - \varphi - K \leq 0 \}. \tag{2.7}
\]

To prove Robinson regularity in problem (2.1), we need to assume that \( f^l \) and \( f^u \) are uniformly bounded away from the true data:

\[
| f^u - f^l | \geq \varepsilon \mathbb{1} \quad \text{for some } \varepsilon > 0. \tag{2.8}
\]

This assumption will be extended in assumption 1 to cover the case of sequences \( f^l_n \) and \( f^u_n \).

Now we can proceed with the Robinson condition.

**Lemma 3.** If (2.8) holds then the Robinson condition (2.7) is fulfilled at any minimiser \( \bar{u} \) of (2.1).

**Proof.** Fix \( \varepsilon > 0 \) and take an arbitrary \( \omega \in L^\infty \) with \( \| \omega \|_{\infty} < \varepsilon \). Our aim is to find \( u \geq 0 \) and \( \varphi \leq 0 \) such that \( \omega = Bu - \varphi - \psi \).

Choose \( u = \bar{u} \) \( \geq 0 \). Then we have that

\[
\psi = B\bar{u} - \varphi - \omega \leq B\bar{u} - \varphi - \omega = \varphi - \varphi - \omega \leq -\varepsilon \mathbb{1} - \omega.
\]
We see that $v \leq 0$ holds if $\|\omega\|_\infty$ is small enough, therefore, Robinson condition (2.7) holds at $\tilde{u}$ in the primal problem (2.1).

Now we are ready to study the relationship between the primal problem (2.1) and its dual (2.2).

Proposition 4. Under the assumptions of lemma 3, strong duality between (2.1) and (2.2) holds, the complementarity conditions
\[
\begin{cases}
(\tilde{\mu}, B\tilde{u} - \varphi) = 0, \\
(\tilde{\lambda}, \tilde{u}) = 0
\end{cases}
\]
are satisfied and $\tilde{\lambda} - B^*\tilde{\mu} \in \partial J(\tilde{u})$, where $(\tilde{\lambda}, \tilde{\mu})$ denotes the solution of the dual problem (2.2).

Proof. Strong duality between the primal problem (2.1) and its dual (2.2) follows from proposition 1, since the primal problem (2.1) is convex, its optimal value is bounded (by $J(\bar{u})$) and Robinson condition (2.7) is satisfied. Therefore, we have that
\[
J(\tilde{u}) = (\tilde{\mu}, -\varphi).
\]
Consider the element $\tilde{\lambda} - B^*\tilde{\mu} \in \partial J(0)$. Since $\tilde{\lambda} - B^*\tilde{\mu}$ is a subgradient, we get that
\[
J(\tilde{u}) - (\tilde{\lambda} - B^*\tilde{\mu}, \tilde{u}) \geq 0
\]
and, since $J(\tilde{u}) = (\tilde{\mu}, -\varphi)$, also that
\[
0 \leq (\tilde{\lambda}, \tilde{u}) \leq (\tilde{\mu}, B\tilde{u} - \varphi) \leq 0
\]
(the latter inequality holds since $\tilde{\mu} \geq 0$ and $B\tilde{u} \leq \varphi$). Therefore, the complementarity conditions (2.9) are satisfied.

Since $(\tilde{\lambda} - B^*\tilde{\mu}, \tilde{u}) = (\tilde{\mu}, -B\tilde{u}) = (\tilde{\mu}, -\varphi) = J(\tilde{u})$ and $\tilde{\lambda} - B^*\tilde{\mu} \in \partial J(0)$, we conclude that $\tilde{\lambda} - B^*\tilde{\mu} \in \partial J(\tilde{u})$ by proposition A.4.

3. Convergence analysis

In this section we turn our attention to sequences of primal and dual problems defined using sequences of bounds (1.9):
\[
\min_{u \in L^1: u \geq 0} J(u) \quad \text{s.t.} \quad B_u u = \varphi_u
\]
and
\[
\max_{\lambda, \mu \geq 0} (\mu, -\varphi_u) \quad \text{s.t.} \quad \lambda - B_u^*\mu \in \partial J(0).
\]
We will be particularly interested in the convergence of their solutions to those of the limit problems with exact data and operator (note that (3.3) is just another way of writing (1.2)).

---

5 Note that since the interior of the positive cone is empty in all $L^p$ spaces except for $L^\infty$, a bound on any other $L^p$ norm of $\omega$ would be insufficient.
\[
\min_{u \in L^1: u \geq 0} J(u) \quad \text{s.t. } \bar{B}u = \varphi
\]

and

\[
\max_{\lambda, \mu \geq 0} \{ \mu, -\varphi \} \quad \text{s.t. } \lambda - B^*\mu \in \partial J(0).
\]

We start with the convergence of primal variables \(u_n\) — solutions of (3.1).

### 3.1. Convergence of primal solutions

It can be easily verified that any \(J\)-minimising solution \(\bar{u}_J\) satisfies \(B_n\bar{u}_J \leq \varphi_n\) for all \(n\), which implies \(J(u_n) \leq J(\bar{u}_J)\). It has been shown in [26] that under standard assumptions on \(J\) the minimisers of (3.1) converge to a \(J\)-minimising solution \(\bar{u}_J\) strongly in \(L^1\):

**Theorem 5.** If the regulariser \(J(\cdot): L^1 \rightarrow \mathbb{R}^+ \cup \{\infty\}\)

- is strongly lower-semicontinuous in \(L^1\),
- its non-empty sub-levelsets \(\{u: J(u) \leq C\}\) are strongly sequentially compact,

then there exists a minimiser \(u_n\) of (2.1), \(u_n \rightarrow \bar{u}_J\) strongly in \(L^1\) (possibly, along a subsequence) and \(J(u_n) \rightarrow J(\bar{u}_J)\).

The proof is similar to that in [26, theorem 2].

Assumptions of theorem 5 are satisfied, for example, for the (weighted) \(BV\)-norm \(J(u) = TV(u) + \gamma \|u\|_1\), \(\gamma > 0\), or its topological equivalents with TV replaced with, e.g. TGV [6] or TVL\(^p\) [12]. The term \(\gamma \|u\|_1\) can be dropped if its boundedness is implied by the condition \(B_nu \leq \varphi_n\) (we will see an example of this in section 3.2).

In order to make sure the Robinson condition is satisfied in (3.1) for all \(n\), we need to extend the assumption that we already made in (2.8) to sequences of bounds \(f^n_l\) and \(f^n_u\). In order to have all assumptions on convergence in one place, we also include our assumptions on the convergence of the operator, which we will need later, in the following

**Assumption 1.** Suppose that there exists a sequence \(\varepsilon_n \downarrow 0\) and a constant \(C_0 \geq 1\) as well as a sequence \(\eta_n \downarrow 0\) and a constant \(D_0 > 0\) such that

\[
\varepsilon_n \mathbb{1} \leq \varphi_n - \bar{\varphi} \leq C_0 \cdot \varepsilon_n \mathbb{1},
\]

\[
\|B - B_n\|_{L^1 \rightarrow L^\infty} \leq \eta_n,
\]

\[
\limsup_{n \rightarrow \infty} \frac{\eta_n}{\varepsilon_n} \leq D_0.
\]

The meaning of (3.5) is that \(\varphi_n\) converges to \(\bar{\varphi}\) uniformly, but not too fast; the difference is always uniformly bounded away from zero. The second inequality in (3.5) obviously implies that \(\|\varphi_n - \bar{\varphi}\|_\infty = O(\varepsilon_n)\). The meaning of (3.7) is that the data do not converge faster than the operator.

### 3.2. Boundedness of feasible solutions of the primal problem

In this section we will show that under some assumptions about the exact forward operator \(A\) all elements of the feasible set \(\{u \geq 0: B_nu \leq \varphi_n\}\) are uniformly bounded in \(L^1\). Assumptions
from this section will not be used in the rest of the paper, unless specifically stated, and the results of other sections will be also valid for more general forward operators.

Since all elements $u$ of the feasible set $\{ u \geq 0 : B_n u \leq \varphi_n \}$ are positive, we have that $\|u\|_1 = (u, 1)$. Consider the following optimisation problem:

$$\max_{u \geq 0} (u, 1) \quad \text{s.t.} \quad B_n u \leq \varphi_n. \quad (3.8)$$

It is a linear programming problem and its dual is as follows [2]

$$\min_{\mu \geq 0} (\mu, \varphi_n) \quad \text{s.t.} \quad B^*_n \mu \geq 1. \quad (3.9)$$

We make the following assumption about the exact forward operator $A$:

**Assumption 2.** Assume that the adjoint operator $A^*: (L^\infty)^* \to L^\infty$ satisfies the following condition:

$$A^* 1 \geq c 1$$

for some constant $c > 0$.

This assumption is satisfied in many imaging inverse problems, such as deconvolution [11] and PET [33]. It also trivially satisfied for denoising and inpainting.

3.2.1. The case $B_n \equiv \bar{B}$. In order to get some intuition, let us first consider the case $B_n \equiv \bar{B}$.

**Theorem 6.** Suppose that $B_n \equiv \bar{B}$ for all $n$ and assumption 2 is satisfied. Then all elements of the feasible set in (3.1) are uniformly bounded in $L^1$.

**Proof.** It is easy to verify that $\mu = \frac{1}{c} (1, 0)$ is a feasible solution of (3.9) (here $c$ is the constant from assumption 2). Indeed, we have that $\bar{B}^* \mu = \frac{1}{c} A^* 1 - 0 \geq 1$. By weak duality we have that problem (3.8) is bounded and $\|u\|_1 \leq \frac{1}{c} (f^*_n, 1) \leq C$, since $f^*_n \to f$ strongly in $L^\infty$. \qed

3.2.2. The general case. In the general case we obtain a similar result using assumption 1.

**Theorem 7.** Suppose that (3.6) holds and assumption 2 is satisfied. Then all elements of the feasible set in (2.1) are uniformly bounded in $L^1$.

**Proof.** (3.6) implies that $\|(A^n_n)^* - (A^n_l)^*\|_{(L^\infty)^* \to L^\infty} \leq \eta_n$ and $\|(A^n_n)^* \mu - (A^n_l)^* \mu\|_{L^\infty} \leq \eta_n \|\mu\|_1$ for any $\mu \in (L^\infty)^*$. Therefore, we have that $-\eta_n \|\mu\|_1 1 \leq (A^n_n)^* \mu - (A^n_l)^* \mu \leq \eta_n \|\mu\|_1 1$ and

$$(A^n_n)^* \mu \geq (A^n_l)^* \mu - \eta_n \|\mu\|_1 1.$$ Taking $\mu = 1$, we get that

$$(A^n_n)^* 1 \geq (A^n_l)^* 1 - \eta_n \|1\|_1 1 \geq A^* 1 - \eta_n |\Omega| 1 \geq (c - |\Omega| \eta_n) 1.$$
Now consider \( \mu = \frac{2}{\epsilon}(\mathbb{1}, 0) \). It is a feasible solution of problem (3.9), since
\[
B^\star_n \mu = \frac{2}{\epsilon}(A^\star_n) \mathbb{1} - 0 \geq (2 - \frac{2}{\epsilon}|\Omega|\eta_n) \mathbb{1} \geq \mathbb{1}
\]
if \( n \) is large enough, since \( \|\mathbb{1}\|_1 = |\Omega| < \infty \). Therefore, problem (3.8) is bounded and \( \|u\|_1 \leq \frac{2}{\epsilon}(f, \mathbb{1}) \leq C \).

\[\square\]

3.3. Strong duality in the limit case

Since the exact operator \( B \) is ill-posed, we cannot expect Robinson regularity to hold in the primal limit problem (3.3) and, therefore, we cannot guarantee strong duality between (3.3) and (3.4) or even the existence of solutions of the dual limit problem (3.4), let alone its stability and convergence of the solutions of (3.2). As usual in ill-posed problems, we will need to make an additional assumption about the dual limit problem, called the source condition [10], which in our case is can be written as follows:

**Assumption 3 (Source condition).** Assume that \( \exists \overline{\mu} \geq 0 \) such that
\[
- B^\star \overline{\mu} \in \partial J(\overline{u}),
\]
(3.10)

Let us note that since \( \overline{\mu} = (\overline{\mu}_1, \overline{\mu}_2) \), where \( \overline{\mu}_{1,2} \in (L^\infty)^* \), \( - B^\star \overline{\mu} = A^\star \overline{\mu}_2 - A^\star \overline{\mu}_1 = A^\star(\overline{\mu}_2 - \overline{\mu}_1) = A^\star \nu \) with \( \nu \in (L^\infty)^* \). Therefore, (3.10) implies the source condition from [10]. On the other hand, since every element \( \nu \in (L^\infty)^* \) can be represented as a difference of two positive elements \( \nu = \nu_+ - \nu_- \) [35], the source condition from [10] also implies (3.10).

**Proposition 8.** Under assumption 3, the pair \((0, \overline{\mu})\) solves the dual problem (3.4) and strong duality between (3.3) and (3.4) holds.

**Proof.** Since \( J \) is absolute one-homogeneous, the source condition (3.10) implies that \( (\overline{\mu}, -\overline{\varphi}) = (\overline{\mu}, -B \overline{\mu}, \overline{u}) = (-B^\star \overline{\mu}, \overline{u}) = J(\overline{u}) \) (see proposition A.1 in appendix A). On the other hand, for any feasible \( \mu, (\mu, -\varphi) \leq J(\overline{u}) \) by weak duality. Therefore, the pair \((0, \overline{\mu})\) solves (3.4) and strong duality holds. \(\square\)

3.4. Stability of the dual problem

The goal of this section is to show that the feasible set in the dual limit problem (3.4) is stable under perturbations of the following form:
\[
\lambda - B^\star \mu + r \in \partial J(0)
\]
for some small \( r \in L^\infty \). Denote by \( S(r) \) the feasible set of the perturbed problem. From proposition 2 we know that if Robinson condition holds at some point \( (\lambda_0, \mu_0) \) at \( r = 0 \) then \( \text{dist}((\lambda_0, \mu_0), S(r)) = O(\text{dist}(\lambda_0 - B^\star \mu_0 + r, \partial J(0))) \).

In order to show that Robinson condition (2.6) is satisfied in (3.4) at \((0, \overline{\mu})\), we need to make the following

**Assumption 4.** Assume that
\[
0 \in \text{int}_{L^\infty} \partial J(0).
\]
We emphasise that, since we consider the regularisation functional in $L^1$, its subdifferential at zero should be considered in $L^\infty$, rather than, for instance, $BV^1$. Assumption 4 holds, for example, for the (weighted) $BV$ norm $\mathcal{J}(\cdot) = TV(\cdot) + \gamma \| \cdot \|_1$, $\gamma > 0$, also in the case when it is considered as a functional from $L^1$ to $\mathbb{R} \cup \{\infty\}$ and not from $BV$ to $\mathbb{R}$, as shown in appendix B. Assumption 4 fails, however, for $\mathcal{J}(\cdot) = TV(\cdot)$ (see appendix B as well).

**Lemma 9.** Under assumption 4 Robinson condition (2.6) holds in (3.4).

**Proof.** We need to show that

$$0 \in \text{int}\{(\lambda - B^* \mu - p \mid \lambda, \mu \geq 0, p \in \partial \mathcal{J}(0)\}.$$

For this, we need to show that for an arbitrary $q \in L^\infty$, $\|q\| < \varepsilon$, we have that $q \in \{(\lambda - B^* \mu - p \mid \lambda, \mu \geq 0, p \in \partial \mathcal{J}(0)\}$, if $\varepsilon$ is small enough. Fix some $q \in L^\infty$, $\|q\| < \varepsilon$. We need to find $\lambda, \mu \geq 0$ and $p \in \partial \mathcal{J}(0)$ such that

$$q = \lambda - B^* \mu - p.$$

The required condition is satisfied if we take $\lambda = \mu = 0$ and $p = -q$. Since $|p| = \|q\| < \varepsilon$ and $0 \in \text{int}\partial \mathcal{J}(0)$, we see that $p \in \partial \mathcal{J}(0)$ and Robinson condition is satisfied. □

### 3.5. Boundedness of Lagrange multipliers $\mu_n$

Now we want to investigate the convergence of the Lagrange multipliers $\mu_n$ and $\lambda_n$, which by the results of section 2.4 are related to the subgradient of $\mathcal{J}$ at $u_n$. As noted earlier, convergence of the subgradient plays an important role in establishing convergence rates.

#### 3.5.1. The case $B_n \equiv \tilde{B}$

Again, we will first consider the case $B_n \equiv \tilde{B}$. We will see that it significantly differs from the general case, because it does not require assumption 4.

**Theorem 10.** Suppose that $B_n \equiv \tilde{B}$ for all $n$ and assumptions 1 (convergence) and 3 (source condition) hold. Then

$$\|\mu_n\|_1 \leq C_0 \|\tilde{\mu}\|_1. \quad (3.11)$$

**Proof.** Consider two problems (3.2) and (3.4). Since $B_n \equiv \tilde{B}$, their feasible sets coincide and $(0, \tilde{\mu})$ is a feasible solution of (3.2). Therefore, $(\tilde{\mu}, -\tilde{\varphi}_n) \leq (\mu_n, -\varphi_n)$. Similarly, $(\lambda_n, \mu_n)$ is a feasible solution of (3.4) and $(\mu_n, -\tilde{\varphi}) \leq (\tilde{\mu}, -\tilde{\varphi})$. Combining these two estimates, we conclude that $(\mu_n, \varphi_n - \tilde{\varphi}) \leq (\tilde{\mu}, \varphi_n - \tilde{\varphi})$. Assumption 1 implies that

$$\varepsilon_n(\mu_n, 1) \leq (\mu_n, \varphi_n - \tilde{\varphi}) \leq C_0 \cdot \varepsilon_n(\tilde{\mu}, 1).$$

Since $\mu_n \geq 0$ and $\tilde{\mu} \geq 0$, estimate (3.11) follows. □

#### 3.5.2. The general case

In the general case the optimal solution of one problem is no longer a feasible solution of the other one, but due to the stability of the feasible set, there are feasible points ‘not too far away’.

**Theorem 11.** Suppose that assumptions 1 (convergence), 3 (source condition) and 4 (non-empty interior) hold. Then there exists a constant $C \geq C_0$ such that
\[ \| \mu_n \|_1 \leq \tilde{C} \| \tilde{\mu} \|_1. \tag{3.12} \]

**Proof.** Consider first the limit problem (3.4). Since Robinson condition holds in the dual problem (lemma 9), by proposition 2 we have that
\[ \text{dist}(0, \tilde{\mu}), \{ \lambda, \mu \geq 0 : \lambda - B_n^* \mu \in \partial J(0) \} \leq C_1 \text{dist}(-B_n^* \tilde{\mu}, \partial J(0)) \leq C_1 \| B - B_n \| \| \tilde{\mu} \|_1, \]
where the last inequality holds because \(-B_n^* \tilde{\mu} = -B^* \tilde{\mu} + (B^* \tilde{\mu} - B_n^* \tilde{\mu}) \) and \(-B^* \tilde{\mu} \in \partial J(0)\).

Therefore, there exist \( \tilde{\lambda}, \tilde{\mu} \geq 0 \) such that \( \tilde{\lambda} - B_n^* \tilde{\mu} \in \partial J(0) \) and \( \| (\tilde{\lambda}, \tilde{\mu}) - (0, \tilde{\mu}) \| \leq C_1 \| B - B_n \| \| \tilde{\mu} \|_1 \) (and, therefore, \( \| \tilde{\mu} - (0, \tilde{\mu}) \| \leq C_1 \| B - B_n \| \| \tilde{\mu} \|_1 \)). Since \( (\tilde{\lambda}, \tilde{\mu}) \) is feasible, we get that \( (\tilde{\mu}, -\varphi_\infty) \leq (\mu_n, -\varphi_\infty) \). Furthermore, since \( (\tilde{\mu} - \tilde{\mu}, \varphi_\infty) \leq \| \tilde{\mu} - \tilde{\mu} \| \| \varphi_\infty \| \leq C_1 \| B - B_n \| \| \tilde{\mu} \|_1 \)
where \( C_1 = C_1 \| \varphi_\infty \| \) and can be chosen arbitrary close to \( C_1 \| \varphi \| \), we get that
\[ (\tilde{\mu}, -\varphi_\infty) \leq (\mu_n, -\varphi_\infty) + \tilde{C}_1 \| B - B_n \| \| \tilde{\mu} \|_1. \tag{3.13} \]

Similarly, in the dual problem for finite \( n \) (3.4) we get that there exist \( \tilde{\lambda}_n, \tilde{\mu}_n \geq 0 \) such that \( \tilde{\lambda}_n - B_n^* \tilde{\mu}_n \in \partial J(0) \) and \( \| \tilde{\mu}_n - \mu_n \| \leq \tilde{C}_2 \| B - B_n \| \| \mu_n \|_1 \). Since \( (\tilde{\lambda}_n, \tilde{\mu}_n) \) is feasible, we get that \( (\tilde{\mu}_n, -\varphi) \leq (\mu_n, -\varphi) \) and
\[ (\mu_n, -\varphi) \leq (\tilde{\mu}_n, -\varphi) + \tilde{C}_2 \| B - B_n \| \| \mu_n \|_1. \tag{3.14} \]

Combining (3.13) and (3.14), we get the following estimate
\[ \varepsilon_n \| \mu_n \|_1 \leq (\mu_n, \varphi_\infty - \varphi) \leq (\tilde{\mu}_n, \varphi_\infty - \varphi) + \tilde{C}_1 \| B - B_n \| \| \tilde{\mu}_n \|_1 + \tilde{C}_2 \| B - B_n \| \| \mu_n \|_1 \]
\[ \leq C_0 \varepsilon_n \| \tilde{\mu} \|_1 + \tilde{C}_1 \eta_n \| \tilde{\mu} \|_1 + \tilde{C}_2 \eta_n \| \mu_n \|_1, \]
or, equivalently,
\[ \| \mu_n \|_1 \left( 1 - \frac{\tilde{C}_2 \eta_n}{\varepsilon_n} \right) \leq C_0 \| \tilde{\mu} \|_1 \left( 1 + \frac{\tilde{C}_1 \eta_n}{C_0 \varepsilon_n} \right). \]

If the constant \( D_0 \) in (3.7) is small enough, this implies (3.12) with
\[ \tilde{C} = C_0 \left( 1 + \frac{\tilde{C}_1 \eta_n}{C_0 \varepsilon_n} \right). \tag{3.15} \]

**Corollary 12.** Since the sequence \( \{ \mu_n \} \) is bounded in \( (L^\infty)^* \), the sequence \( \{ B_n^* \mu_n \} \) is bounded in \( L^\infty \) if the operators \( B_n^* \) are bounded from \( (L^\infty)^* \) to \( L^\infty \).

**Remark 13.** Note that in the case \( B_n \equiv \bar{B} \) we did not use the stability of the dual problem and, therefore, did not need assumption 4 to show that the Lagrange multipliers \( \mu_n \) are bounded. This demonstrates that assumption 4 plays an important role specifically in problems with an imperfect operator.
3.6. Boundedness of Lagrange multipliers $\lambda_n$

**Proposition 14.** Suppose that $\mathcal{J}(\mathbb{I}) < +\infty$. Then under the assumptions of theorem 11 we have that $\|\lambda_n\|_1 \leq C$.

**Proof.** Since $\lambda_n - B_n^* \mu_n \in \partial \mathcal{J}(0)$, we have that $\forall n \mathcal{J}(u) - (\lambda_n - B_n^* \mu_n, u) \geq 0$, or, equivalently,

$$(\lambda_n, u) \leq \mathcal{J}(u) + (\mu_n, B_n u).$$

Choosing $u = \mathbb{I}$, we obtain an estimate of the $L^1$ norm of $\lambda_n$ (since $\lambda_n \geq 0$):

$$\|\lambda_n\|_1 = (\lambda_n, \mathbb{I}) \leq \mathcal{J}(\mathbb{I}) + (\mu_n, B \mathbb{I}) \leq \mathcal{J}(\mathbb{I}) + \|\mu_n\|_1 \|\mathbb{I}\|_\infty \leq C. \quad (3.16)$$

\[\square\]

It is worth noting that, although $\lambda_n \in L^\infty$, we only get a bound on the $L^1$ norm of $\lambda_n$ here.

3.7. Convergence of the subgradient

Now we are ready to study the convergence of the subgradient of $\mathcal{J}$ at the optimal solution of the primal problem $u_n$.

**Proposition 15.** Under the assumptions of theorem 11 the sequence $\mu_n$ has a weakly-$^*$ convergent subsequence (in $(L^\infty)^*$), which we still denote by $\mu_n$, $\mu_n \rightharpoonup^* \hat{\mu}$, and $B_n^* \mu_n \rightharpoonup^* B^* \hat{\mu}$ in $L^\infty$.

**Proof.** Since $\mu_n$'s are bounded in $(L^\infty)^*$, $\mu_n \rightharpoonup^* \hat{\mu}$ (along a subsequence) in $(L^\infty)^*$ by the Banach–Alaoglu theorem [11], i.e. for any $\varphi \in L^\infty$ we have that $(\mu_n, \varphi) \to (\hat{\mu}, \varphi)$. Considering, for an arbitrary $u \in L^1$, the scalar product $(B_n^* \mu_n, u)$, we note that $(B_n^* \mu_n, u) = (\mu_n, B_n u) \to (\hat{\mu}, B u) = (B^* \hat{\mu}, u)$, since $\mu_n \rightharpoonup^* \hat{\mu}$ in $(L^\infty)^*$ and $B_n u \to B u$ in $L^\infty$. Therefore, $B_n^* \mu_n \rightharpoonup^* B^* \hat{\mu}$ in $L^\infty$. \[\square\]

**Remark 16.** Since (3.12) holds for any $\hat{\mu}$ delivering the source condition, i.e. every $\hat{\mu} \geq 0$ such that $\exists \lambda \geq 0$, $\lambda - B^* \hat{\mu} \in \partial \mathcal{J}(\hat{\mu})$, it also holds for the (possibly non-unique) minimum-norm certificate that solves the following problem:

$$(\hat{\lambda}_{\min}, \hat{\mu}_{\min}) \in \mathcal{arg \ min}_{\lambda \geq 0} \|\mu\|_1 \quad \text{s.t.} \quad \lambda - B^* \mu \in \partial \mathcal{J}(u_{\mathcal{J}}). \quad (3.17)$$

Since $\mu_n \rightharpoonup^* \hat{\mu}$ in $L^\infty$, we have that $\|\hat{\mu}\|_1 \leq \tilde{C}\|\hat{\mu}_{\min}\|_1$, where $\tilde{C}$ is given by (3.15). If the operator converges faster than the data, i.e. $\|\hat{\mu}\|_1 \leq \tilde{C}\|\hat{\mu}_{\min}\|_1$, then, taking the limit in (3.15) and letting $D_0 \to 0$, we get that $\|\hat{\mu}\|_1 \leq C_0\|\hat{\mu}_{\min}\|_1$. On the other hand, since $\lambda - B^* \hat{\mu} \in \partial \mathcal{J}(\hat{\mu})$, we have that $\|\hat{\mu}\|_1 \leq |\hat{\mu}|_1$. If $C_0 = 1$, we have that $\|\hat{\mu}\|_1 = \|\hat{\mu}_{\min}\|_1$, i.e. $\hat{\mu}$ is a minimum-norm certificate. In the general case (3.6) we can only say that the norm of $\hat{\mu}$ is bounded by that of $\hat{\mu}$ times a constant.

We would like to have that the whole subgradient $p_n = \lambda_n - B_n^* \mu_n$ is bounded in $L^{\infty}$, however, we have only a bound in $L^1$ for the first summand $\lambda_n$. However, we know that $(\lambda_n, u_n) = 0$ for all $n$ (see (2.9)) and, as we shall see later, the same holds for the characteristic functions of the level sets of $u_n$ (see section 4). Therefore, boundedness (in $L^\infty$) of $B_n^* \mu_n$ will suffice in most cases.
To study the convergence of the subgradient $p_n$, let us consider the subspace $\mathcal{V} = \{u \in L^1 : J(u) < +\infty\}$.

**Theorem 17.** Suppose that $\mathcal{J}(\cdot)$ is a norm on $\mathcal{V} = \{u \in L^1 : J(u) < +\infty\}$. Then, under the assumptions of theorem 11, we have that

$$p_n = \lambda_n - B^*_n \mu_n \rightarrow^* \hat{p} \quad \text{in } \mathcal{V}^*$$

and $\hat{p} \in \partial \mathcal{J}_{\mathcal{V}^*}(\bar{u}_\mathcal{J})$ for all $\mathcal{J}$-minimising solutions $\bar{u}_\mathcal{J}$. We also have that $\lambda_n \rightarrow^* \lambda$ in $\mathcal{V}^*$, $(\lambda, u) \geq 0$ for any $u \in \mathcal{V}$ such that $u \geq 0$ and $(\lambda, \bar{u}_\mathcal{J}) = 0$. All convergences are along a subsequence, which we do not relabel.

**Proof.** Since the dual of a norm is the characteristic function of the unit ball in the dual norm [5], we have that

$$\partial \mathcal{J}_{\mathcal{V}^*}(0) = \{p \in \mathcal{V}^* : \|p\|_{\mathcal{V}^*} \leq 1\}.$$

By the Banach–Alaoglu theorem we get weak-* convergence of a subsequence $p_n \rightarrow^* \hat{p}$ in $\mathcal{V}^*$. Weak-* convergence $B^*_n \mu_n \rightarrow^* \hat{B}^* \bar{\mu}$ in $L^\infty$ (and, therefore, in $\mathcal{V}^*$) implies that $\lambda_n \rightarrow^* \lambda = \hat{p} = \hat{B}^* \bar{\mu}$ in $\mathcal{V}^*$.

To study the properties of $\hat{p}$, we make the following observation:

$$(p_n, u_n) = (\lambda_n, u_n) - (B^*_n \mu_n, u_n) = (\mu_n, -\mu_n),$$

$$(p_n, \bar{u}_\mathcal{J}) = (\lambda_n, \bar{u}_\mathcal{J}) - (B^*_n \mu_n, \bar{u}_\mathcal{J}) \geq (\mu_n, -\bar{u}_\mathcal{J}) = (\mu_n, -\bar{u}_\mathcal{J}) = (p_n, u_n)$$

for any $\mathcal{J}$-minimising solution $\bar{u}_\mathcal{J}$. (Note that the term $(\lambda_n, u_n)$ in the first line vanishes by proposition 4). Therefore, we have that

$$0 \leq \mathcal{J}(\bar{u}_\mathcal{J}) - (p_n, \bar{u}_\mathcal{J}) \leq \mathcal{J}(\bar{u}_\mathcal{J}) - (p_n, u_n) = \mathcal{J}(\bar{u}_\mathcal{J}) - \mathcal{J}(u_n) \rightarrow 0.$$

Hence, we get that $(p_n, u_n) \rightarrow \mathcal{J}(\bar{u}_\mathcal{J})$. Combining this with $(p_n, \bar{u}_\mathcal{J}) \rightarrow (\hat{p}, \bar{u}_\mathcal{J})$ (since $\bar{u}_\mathcal{J} \in \mathcal{V}$), we get that $(\hat{p}, \bar{u}_\mathcal{J}) = \mathcal{J}(\bar{u}_\mathcal{J})$ and $\hat{p} \in \partial \mathcal{J}_{\mathcal{V}^*}(\bar{u}_\mathcal{J})$ (the condition $\hat{p} \in \partial \mathcal{J}_{\mathcal{V}^*}(0)$ follows from weak-* closedness of the unit ball in $\mathcal{V}^*$).

Clearly, $(\lambda, u) \geq 0$ for all $u \in \mathcal{V}$, $u \geq 0$. Noting that

$$\mathcal{J}(u_n) = (p_n, u_n) = (\lambda_n, u_n) - (B^*_n \mu_n, u_n)$$

$$= (-B^*_n \mu_n, u_n) \rightarrow (-\hat{B}^* \bar{\mu}, \bar{u}_\mathcal{J}),$$

we conclude that $(-\hat{B}^* \bar{\mu}, \bar{u}_\mathcal{J}) = \mathcal{J}(\bar{u}_\mathcal{J})$. Combining this with $(\hat{p}, \bar{u}_\mathcal{J}) = \mathcal{J}(\bar{u}_\mathcal{J})$, we get that $(\lambda, \bar{u}_\mathcal{J}) = 0$. \hfill \Box

3.8. Convergence rates

The results of the previous sections allow us to obtain convergence rates of $u_n \rightarrow \bar{u}_\mathcal{J}$ in terms of the (generalised) Bregman distance [10]. Indeed, consider the symmetric Bregman distance, which for absolutely one-homogeneous functionals can be written as follows

$$D_{\mathcal{J}}^{\text{sym}}(\bar{u}_\mathcal{J}, u_n) = D_{\mathcal{J}}^S(\bar{u}_\mathcal{J}, u_n) + D_{\mathcal{J}}^S(u_n, \bar{u}_\mathcal{J}) = (p_n - \hat{p}, u_n - \bar{u}_\mathcal{J}).$$

**Theorem 18.** Under the assumptions of theorem 11 the following estimate holds for any
\[
\tilde{J}\text{-minimising solution } \tilde{u}_\tilde{J}:
\]
\[
D_{\tilde{J}}^{\text{symm}}(\tilde{u}_\tilde{J}, u_n) \leq C\|\tilde{\mu}\|_{1} \cdot \varepsilon_n.
\]  

**Proof.** We obtain the following estimate for the symmetric Bregman distance:
\[
(p_n - \hat{p}, u_n - \bar{u}_\tilde{J}) = (\lambda_n, u_n) - (\lambda_n, \tilde{u}_\tilde{J}) - (\hat{\lambda}, u_n) + (\hat{\lambda}, \tilde{u}_\tilde{J}) - (\mu_n, B_n u_n) + (\mu_n, B_n \bar{u}_\tilde{J}) - (\mu_n, B_n \tilde{u}_\tilde{J}) - (\mu_n, B_n \tilde{u}_\tilde{J}),
\]

since \((\lambda_n, u_n) = 0 \text{ (proposition 4)}, (\lambda_n, \bar{u}_\tilde{J}) \geq 0, (\hat{\lambda}, u_n) \geq 0, (\hat{\lambda}, \tilde{u}_\tilde{J}) = 0 \text{ (theorem 17)}. Using the fact that \((\mu_n, B_n u_n) = (\mu_n, \varphi_n)\) and \(B_n \bar{u}_\tilde{J} \leq \varphi_n\), we note that
\[
-(\mu_n, B_n u_n) + (\mu_n, B_n \bar{u}_\tilde{J}) = (\mu_n, B_n (\bar{u}_\tilde{J} - \varphi_n)) \leq 0
\]

and, therefore,
\[
(p_n - \hat{p}, u_n - \bar{u}_\tilde{J}) \leq (\mu_n, B_n u_n - \varphi) + (\mu_n, B_n \varphi - \varphi) + (\hat{\mu}, (B - B_n)u_n) \\
\leq (\mu_n, \varphi_n - \varphi) + (\hat{\mu}, (B - B_n)u_n).
\]

The last inequality is due to the fact that \(B_n u_n \leq \varphi_n\). Using assumption 1 and the fact that \(\|\mu\|_{1} \leq C\|\mu\|_{1}\) (theorem 11 and proposition 15), we finally obtain the required estimate
\[
(p_n - \hat{p}, u_n - \bar{u}_\tilde{J}) \leq (\hat{\mu}, \varphi_n - \varphi) + (\hat{\mu}, (B - B_n)u_n) \leq C \cdot \varepsilon_n \cdot (\hat{\mu}, 1) + \|\mu\|_{1} \cdot \|B - B_n\| \leq C\|\tilde{\mu}\|_{1} \cdot \varepsilon_n.
\]

Not surprisingly, the convergence rate only depends on the convergence of the data, since we assumed that the operator converges at least at the same rate (assumption 1).

**Remark 19.** The estimate (3.19) is consistent with existing theory for inverse problems with exact forward operators. If \(B_n \equiv B\), the constraint \(Bu \leq \varphi_n\) is essentially a bound on the (perhaps, weighted) \(L^\infty\) norm of \(Au - f\) for \(f = \frac{\|\mu\|_{1}}{2}\). The case when the fidelity function is a characteristic function of the set \(\{u: \|Au - f\| \leq \delta\}\) was studied in [3, theorem 5.1], where the authors obtained the same convergence rate as (3.19).

4. Convergence of the level sets of \(u_n\)

Our goal in this section is to understand the structure of the minimisers \(u_n\) in the case of TV-based regularisation. In particular, we want to know whether the level sets of \(u_n\) converge to those of \(\bar{u}_\tilde{J}\), where \(\bar{u}_\tilde{J}\) is the \(\tilde{J}\)-minimal solution of (1.1), to which \(u_n\) converges. In this section we consider \(J(\cdot) = TV(\cdot) + \gamma \|\cdot\|_{1}\), where \(\gamma\) is a small constant (recall that \(J(\cdot) = TV(\cdot)\) does not satisfy assumption 4). We follow [14] and [22], where the authors proved Hausdorff convergence of the level sets of solutions of the ROF model [32] (for denoising in [14] and for general linear inverse problems in [22]) to those of \(\bar{u}_\tilde{J}\). In particular, if \(\bar{u}_\tilde{J}\) is piecewise-constant, the authors of [14] conclude that the reconstructions are piecewise-constant outside the so-called extended support of the gradient of \(\bar{u}_\tilde{J}\) in the low noise regime.

Our case requires several adjustments of the proofs in [14]. First, [14] considers \(J(\cdot) = TV(\cdot)\), while we need to consider \(J(\cdot) = TV(\cdot) + \gamma \|\cdot\|_{1}\). Therefore,
instead of considering sets satisfying $P(E) = \int_E p_n$ we need to consider sets satisfying $P(E) + \gamma|E| = \int_E p_n$. Therefore, the level sets of $u_n$ (as defined in [14]) solve the following optimisation problem (instead of the prescribed mean curvature problem):

\[
\min_{X \subset \Omega} P(X) + \gamma|X| - \int_X p_n. \tag{4.1}
\]

(Note that the case with the opposite sign of the integral does not occur since $u_n \geq 0$.)

Denote by $E_n^{(1)}$ (or by $E_n$, where this will cause no confusion) the level sets of $u_n$. To prove that $E_n$ indeed solves problem (4.1) we note that, since

\[
\int g \varphi \, dx = \int \left( \int_0^{\infty} \mathbb{1}_{g(x) \geq \varphi(x)} \, dt \right) \, dx
\]

for any functions $g, \varphi$ such that $g \geq 0$, we get that

\[
(p_n, u_n) = \int_0^{\infty} \left( \int_{E_n^{(1)}} p_n \right) \, dt,
\]

\[
\|u_n\|_1 = (u_n, \mathbb{1}) = \int_0^{\infty} \left( \int_{E_n^{(1)}} \mathbb{1} \right) \, dt
\]

and $TV(u_n) = \int_0^{\infty} \left( \int_{E_n^{(1)}} TV(\mathbb{1}_{E_n^{(1)}}) \right) \, dt$ by the coarea formula [14]. Combining this with $\mathcal{J}(u_n) = (p_n, u_n)$, we get that

\[
\int_0^{\infty} \left( P(E_n^{(1)}) + \gamma|E_n^{(1)}| - \int_{E_n^{(1)}} p_n \right) \, dt = 0.
\]

Since $p_n \in \partial \mathcal{J}(0)$, the expression in the outer integral is non-negative and we get the desired equality

\[
P(E_n) + \gamma|E_n| = \int_{E_n} p_n. \tag{4.2}
\]

Since the objective in (4.1) is non-negative, $E_n$ indeed solves (4.1).

To prove Hausdorff convergence of the level sets of $u_n$ to those of $\bar{u}_\mathcal{J}$ along the lines of [14], we need to prove lemma 2 and proposition 8 (following the notation of the arXiv version of the paper). The proofs in [14] rely on strong $L^2$ convergence of the subgradients, which we do not have in our case. However, weak$^*$ convergence of $B_n^* \mu_n$ in $L^\infty$ along with some orthogonality properties of $\lambda_n$ will be enough to obtain similar results, as we shall see.

Before we proceed with the proofs, let us note that for any level set $E_n^{(1)}$, $t > 0$, the following inequality holds: $\mathbb{1}_{E_n^{(1)}} \leq \frac{1}{t} u_n$. Therefore, $0 \leq (\lambda_n, \mathbb{1}_{E_n^{(1)}}) \leq \frac{1}{t}(\lambda_n, u_n) = 0$.

**Lemma 20 (Lemma 2 in [14]).** The level sets $E_n$ have a finite perimeter and area.

**Proof.** Since $E_n \subset \Omega$ and we assumed that $\Omega$ is bounded, finiteness of the area of $E_n$ is trivial. For the perimeter $P(E_n) = TV(\mathbb{1}_{E_n})$ we obtain the following estimate using (4.2)

\[
P(E_n) \leq \mathcal{J}(\mathbb{1}_{E_n}) = (p_n, \mathbb{1}_{E_n}) = (\lambda_n, \mathbb{1}_{E_n}) - (B_n^* \mu_n, \mathbb{1}_{E_n}) = (-B_n^* \mu_n, \mathbb{1}_{E_n}) \leq \|B_n^* \mu_n\|_\infty |E_n| \leq C. \tag*{□}
\]

**Proposition 21 (Proposition 8 in [14]).** $\exists r_0 > 0$ such that $\forall r \in (0, r_0), \forall E_n$ and $\forall x \in \partial E_n$ the following estimates hold:
where $B(x, r)$ denotes a ball of radius $r$ centered at $x$.

**Proof.** Due to the optimality of $E_n$ in problem (4.1), we get that

$$P(E_n) + \gamma|E_n| - \int_{E_n} p_n \leq P(E_n \setminus B(x, r)) + \gamma|E_n \setminus B(x, r)| - \int_{E_n \setminus B(x, r)} p_n,$$

which implies the following estimate:

$$P(E_n) - \int_{E_n} p_n \leq P(E_n \setminus B(x, r)) - \int_{E_n \setminus B(x, r)} p_n - \gamma|E_n \cap B(x, r)|.$$

Geometric considerations yield:

$$P(E_n \cap B(x, r)) \leq \int_{E_n \cap B(x, r)} p_n - \gamma|E_n \cap B(x, r)| + 2\mathcal{H}^1(\partial B(x, r) \cap E_n).$$

For the first term on the left hand side we get the following estimate:

$$\int_{E_n \cap B(x, r)} p_n = (\lambda_n, \lambda_{E_n \cap B(x, r)}) - (B_n^*, \lambda_{E_n \cap B(x, r)}) \leq \|B_n^*\|_{\infty} \cdot \|\lambda_{E_n \cap B(x, r)}\| \leq C|E_n \cap B(x, r)|.$$

The isoperimetric inequality [14] yields:

$$\sqrt{4\pi} |E_n \cap B(x, r)|^{1/2} \leq P(E_n \cap B(x, r)).$$

Denote $g(r) := |E_n \cap B(x, r)|$. Then $g'(r) = \mathcal{H}^1(\partial B(x, r) \cap E_n)$ we get the following inequality:

$$\sqrt{4\pi} \sqrt{g(r)} \leq 2g'(r) + (\tilde{C} - \gamma)g(r).$$

Since $g(r) \to 0$ as $r \to 0$, for small $r$ we have that $g(r) \leq \sqrt{g(r)}$ and, therefore,

$$\left(\sqrt{4\pi} - \tilde{C} + \gamma\right) \leq \frac{d}{dr} \sqrt{g(r)}.$$

If the constants $C_0$ and $D_0$ from assumption 1 are small enough, the constant on the left hand side is positive and, integrating, we get that

$$r(\sqrt{4\pi} - \tilde{C} + \gamma) \leq \sqrt{g(r)}$$

and

$$\frac{|B(x, r) \cup E_n|}{|B(x, r)|} \geq \frac{\left(\sqrt{4\pi} - \tilde{C} + \gamma\right)^2}{\pi}.$$
Comparing $E_n$ with $E_n \cup B(x, r)$ in problem (4.1), we get in a similar way the estimate

$$\frac{|B(x, r) \setminus E_n|}{|B(x, r)|} \geq \frac{\left(\sqrt{4\pi} - \bar{C} - \gamma\right)^2}{\pi}. \square$$

These results are sufficient to show Hausdorff convergence of the level sets of $u_n$ to those of $\bar{u}_J$ [14], theorem 1, [22], theorem 2. Similarly to theorem 1 in [14], one can also show that $E_n \rightarrow E$ in the sense that $\lim_{n \rightarrow \infty} |E \triangle E_n| = 0$ and $P(E) + \gamma|E| = \int_E \hat{p}$. Indeed, passing to the limit in $J(\mathbb{1}_E) = (p_n, \mathbb{1}_{E_n})$, we get that

$$P(E) + \gamma|E| \leq \liminf_{n \rightarrow \infty} P(E_n) + \gamma|E_n| = \lim_{n \rightarrow \infty} \left(\lambda_n - B_n^* \mu_n, \mathbb{1}_{E_n}\right)$$

$$= \lim_{n \rightarrow \infty} \left(-B_n^* \mu_n, \mathbb{1}_{E_n}\right) = (-B^* \hat{\mu}, \mathbb{1}_E) = (\bar{\lambda} - B^* \hat{\mu}, \mathbb{1}_E) = (\hat{p}, \mathbb{1}_E)$$

due to $L^1$ convergence of $\mathbb{1}_{E_n} \rightarrow \mathbb{1}_E$ and weak-* convergence of $B_n^* \mu_n \rightharpoonup B^* \hat{\mu}$ in $L^\infty$. Therefore, $TV(\mathbb{1}_E) < +\infty$ and $\mathbb{1}_E \in BV$. Since $\hat{p} \in \partial J(0)$, we get that $J(\mathbb{1}_E) \geq (\hat{p}, \mathbb{1}_E)$ and, therefore, $(\hat{p}, \mathbb{1}_E) = P(E) + \gamma|E|$. 

**Remark 22.** From $(\hat{p}, \mathbb{1}_E) = J(\mathbb{1}_E)$ the authors of [14] conclude that $\partial E$ is in the extended support of the gradient of $\bar{u}_J$. We can make a similar connection in the case $\frac{\mu}{\gamma} \rightarrow 0$, when $\mu_n$ converges to a minimum-norm certificate (see remark 16). However, due to non-uniqueness of the minimum norm certificate in our case, the definition of the extended support needs to be amended. We consider all $u$, whose subgradient contains a minimum-norm certificate $(\bar{\lambda}_\text{min}, \bar{\mu}_\text{min})$ solving (3.17):

$$\operatorname{Ext}(\bar{u}_J) = \bigcup \{\supp(Du) \mid u: \exists (\lambda_n, \mu_n) \text{ s.t. } \lambda_n - B^* \mu_n \in \partial J(u)\}. \quad (4.3)$$

**5. Debiasing and error estimation**

Two-step debiasing [9, 15, 16] aims at removing systematic bias in variational regularisation (such as loss of contrast with TV) by solving an additional optimisation problem on the so-called model manifold defined as follows:

$$\mathcal{M} = \{u \in L^1 : J(u) = (p_n, u)\},$$

where $p_n \in \partial J(u_n)$. The model manifold is the set of all elements of the solution space with zero Bregman distance to $u_n$. In other words, it is the set of all elements sharing the subgradient with the approximate solution $u_n$. Informally, the idea of two-step debiasing is that the approximate solution captures well the structure of the exact solution, such as the jump set in TV-based regularisation, but is not perfect quantitatively due to a systematic bias introduced by the regulariser. This systematic bias is (partially) removed by optimising the fidelity term on the model manifold.

**5.1. Debiasing and model manifolds**

Our goal is to adapt the idea of debiasing to our specific setting. We assume that the first step, i.e. the solution of problem (3.1), is already done and an approximate solution $u_n$ is available along with the corresponding subgradient $p_n = \lambda_n - B_n^* \mu_n$. 


We slightly amend the definition of the model manifold for our specific setting. Fix some positive constants $\varepsilon$ and $C$ and consider the following set:

$$\mathcal{M}_n^{\varepsilon,C} = \{ u \geq 0 : B_n u \leq \varphi_n, \quad J(u) - (p_n, u) \leq \varepsilon, \quad (p_n, u - u_n) \leq C \}. \quad (5.1)$$

We introduced two novel constraints as compared to the original feasible set in (2.1). The inequality $J(u) - (p_n, u) \leq \varepsilon$ is an upper bound on the Bregman distance between $u_0$ and $u$. The condition $(p_n, u - u_n) \leq C$ has a more technical nature and will be discussed in more detail in later (see remark 25).

Next we examine some properties of the sets $\mathcal{M}_n^{\varepsilon,C}$.

**Proposition 23.** For sufficiently large $n$ any $J$-minimising solution $\bar{u}_\mathcal{J}$ is an element of $\mathcal{M}_n^{\varepsilon,C}$.

**Proof.** Since $(p_n, u_\mathcal{J}) \to J(\bar{u}_\mathcal{J})$ and $(p_n, u_n) \to J(\bar{u}_\mathcal{J})$, we conclude that $(p_n, \bar{u}_\mathcal{J} - u_n) \leq C$ for sufficiently large $n$. Similarly, $J(\bar{u}_\mathcal{J}) - (p_n, \bar{u}_\mathcal{J}) = (\hat{p} - p_n, \bar{u}_\mathcal{J}) \to 0$ and therefore $J(\bar{u}_\mathcal{J}) - (p_n, \bar{u}_\mathcal{J}) \leq \varepsilon$ for sufficiently large $n$. \hfill $\Box$

**Proposition 24.** $J(\cdot)$ is uniformly bounded on $\mathcal{M}_n^{\varepsilon,C}$ and any sequence $v_n \in \mathcal{M}_n^{\varepsilon,C}$ contains a subsequence (which we do not relabel) that strongly converges to a solution of (1.1) (not necessarily a $J$-minimising solution).

**Proof.** Indeed, $\forall u \in \mathcal{M}_n^{\varepsilon,C}$ we have that

$$J(u) \leq (p_n, u) + \varepsilon \leq (p_n, u_n) + C + \varepsilon \leq J(u_n) + C + \varepsilon.$$

Since the sub-level sets of $J(\cdot)$ are sequentially compact, so are the sets $\mathcal{M}_n^{\varepsilon,C}$ as closed subsets of a compact set. Therefore, we conclude that any sequence $v_n \in \mathcal{M}_n^{\varepsilon,C}$ has a strongly convergent subsequence (that we do not relabel). Since only those $u$ that solve (1.1) belong to all sets $\mathcal{M}_n^{\varepsilon,C}$ (for all $n$), we conclude that $v_n$ converges to a solution of (1.1). \hfill $\Box$

**Remark 25.** Consider the expression $(p_n, u - u_n)$ for some $u \geq 0$ such that $B_n u \leq \varphi_n$. This expression is supposed to be bounded by a ‘user-defined’ constant $C$. Since $(p_n, u_n) = J(u_n) \leq J(\bar{u}_\mathcal{J})$, it is effectively an upper bound on $(p_n, u)$ on the feasible set. Consider the following estimate:

$$(p_n, u) = (\lambda_n, u) - (B_n^* \mu_n, u) \leq (\lambda_n, u) + \|B_n^* \mu_n\|_{\infty} \|u\|_1. \quad (5.2)$$

If assumption 2 is satisfied, $\|u\|$ is bounded and we get a bound $(p_n, u) \leq (\lambda_n, u) + C$, i.e., effectively, we only need an upper bound on $(\lambda_n, u)$ for all feasible $u$. Since $(\lambda_n, u_n) = 0$, we can drop the constant $C$ from the definition of $\mathcal{M}_n^{\varepsilon,C}$ whenever $u_n > 0$ a.e.

Following [9], to correct for the systematic bias of $u_0$ we would need to optimise the data term on $\mathcal{M}_n^{\varepsilon,C}$. However, since in our case the data term is the characteristic function of the set $\{ u \geq 0 : B_n u \leq \varphi_n \}$, optimising it on $\mathcal{M}_n^{\varepsilon,C}$ does not make any sense (any element of $\mathcal{M}_n^{\varepsilon,C}$ is a minimiser). A possible way around this would be to choose an operator $\hat{A}_n \in [A_n, A_n^2]$ and a right-hand side $\hat{f}_n \in \{ f_n, f_n^2 \}$, for example, $\hat{A}_n = \frac{A_n + A_n^2}{2}$ and $\hat{f}_n = \frac{f_n + f_n^2}{2}$, and optimise the discrepancy $\|A_n u - \hat{f}_n\|$ on $\mathcal{M}_n^{\varepsilon,C}$. Convergence of the minimisers is guaranteed by proposition 24.

---

6 In the setting of [9] the Bregman distance is assumed to be zero, although the proposed numerical scheme allows some deviation.
(since the sets $\mathcal{M}^{\epsilon,C}_n$ are closed, the minimisers also belong to $\mathcal{M}^{\epsilon,C}_n$ and, therefore, converge to a solution of (1.1), possibly along a subsequence).

The choice of $\tilde{A}_n$ and $\tilde{f}_n$ depends on our additional assumptions about the nature of the errors in the operator and the data. For example, in the case of symmetric noise, the choice $\tilde{A}_n = \frac{A_n + A^*_n}{2}$ and $\tilde{f}_n = \frac{f_n + f^*_n}{2}$ is quite intuitive.

5.2. Pointwise error estimates in constant regions

Proposition 24 paves way for pointwise error estimates of TV-regularised solutions in areas where the minimiser $u_n$ is constant (by the results of section 4, these areas converge to the areas where $u_\mathcal{F}$ is constant in the sense of Hausdorff convergence). To obtain a meaningful result on the convergence of the pointwise bounds, we assume that the operator $A$ is injective and therefore the exact solution is unique (and will be denoted by $\bar{u}$).

We will make use of the following important property of the model manifold in case of TV-based regularisation: as pointed out in [9], the model manifold in the case $J(u) = TV(u) + \gamma \|u\|_1$, $\gamma = \text{const} > 0$ contains all solutions that share the jump set with $u_n$ (more precisely, they do not jump where $u_n$ does, but do not have to jump where $u_n$ does). This is still valid in the case $J(u) = TV(u) + \gamma ||u||_1$, as shown in appendix C.

Theorem 26 (Pointwise error bars). Suppose that

- $A$ is injective;
- the exact solution $\bar{u}$ is piecewise-constant;
- $J(u) = TV(u) + \gamma ||u||_1$, $\gamma = \text{const} > 0$.

Denote any region where $u_n$ is constant by $\omega_n$. Define $u^l_{\omega_n}$ and $u^u_{\omega_n}$ as follows

$$u^l_{\omega_n} = \min_{u \in \mathcal{M}^{\epsilon,C}_n} (u, \mathbb{1}_{\omega_n}) \quad \text{and} \quad u^u_{\omega_n} = \max_{u \in \mathcal{M}^{\epsilon,C}_n} (u, \mathbb{1}_{\omega_n}).$$

Then $(u^u_{\omega_n} - u^l_{\omega_n}) \to 0$ and for sufficiently large $n$ we have that

$$u^l_{\omega_n} \leq \bar{u}|_{\omega_n} \leq u^u_{\omega_n}.$$  

Proof. Consider an arbitrary $u \in \mathcal{M}^{\epsilon,C}_n$ and denote its value $\omega_n$ by $u|_{\omega_n}$. Consider the following linear functional:

$$(u, \mathbb{1}_{\omega_n}) = u|_{\omega_n} \cdot |\omega_n|.$$  

(5.2)

Since by proposition 23 $\bar{u} \in \mathcal{M}^{\epsilon,C}_n$ if $n$ is sufficiently large, the jump sets of $u_n$ and $\bar{u}$ coincide and we have that

$$(u, \mathbb{1}_{\omega_n}) = \bar{u}|_{\omega_n} \cdot |\omega_n|.$$  

Therefore, minimising and maximising this functional on $\mathcal{M}^{\epsilon,C}_n$ gives us a lower and an upper bound for $\bar{u}$. By proposition 24, the arg min and arg max converge to a solution of (1.1), which is unique by the injectivity of $A$. Therefore, by the continuity of the linear functional (5.2) we get that

$$u^u_{\omega_n} := \max_{u \in \mathcal{M}^{\epsilon,C}_n} (u, \mathbb{1}_{\omega_n}) \to \bar{u}|_{\omega_n} \quad \text{and} \quad u^l_{\omega_n} := \min_{u \in \mathcal{M}^{\epsilon,C}_n} (u, \mathbb{1}_{\omega_n}) \to \bar{u}|_{\omega_n},$$  

21
proving the conjecture.

Remark 27. Note that due to the fact that for a fixed \( n \bar{u} \in M_n^{\infty,C} \) only for sufficiently small \( \varepsilon \), we can only guarantee that \( u_n \) captures the jump set of \( \bar{u} \) in the limit. If we had the inclusion \( \bar{u} \in M_n^{\infty,C} \) with \( \varepsilon = 0 \) for a fixed (but sufficiently large) \( n \), the jump sets of \( u_n \) and \( \bar{u} \) would coincide by the results of Appendix C. It is not clear, whether under any suitable assumptions \( \bar{u} \in M_n^{\infty,C} \) with \( \varepsilon = 0 \) already for a fixed \( n \), and can be an interesting direction of future research.

6. Numerical experiments

In this section we present numerical experiments illustrating the results of the previous sections. We concentrate on 1D examples in order to see the effects of different settings more clearly. We use CVX [18, 19] in all our experiments.

We consider deblurring with uncertainty in the blurring kernel, which has been studied in the partial-order based setting in [27]. Consider the signal shown in figure 1(a) in blue (dashed line). This signal is convolved with a Gaussian blurring kernel with standard deviation 0.5 and Dirichlet boundary conditions and then 2.5\% uniform noise is added to it. The blurred and noisy signal is shown in figure 1(a) in green (solid line). Knowing the amount of noise in this signal, we can obtain lower and upper data bounds \( f_l \) and \( f_u \) as explained in [27].

Being a convolution with a Gaussian kernel, the forward operator \( A \) is injective and therefore the exact solution is unique. Assumption 2 is also satisfied for a convolution operator, which implies that the \( L^1 \) norm of \( u \) is bounded on the feasible set \( \{ u \geq 0 : B_n u \leq \varphi_n \} \) and the regulariser \( J(\cdot) = TV(\cdot) \) satisfies the conditions of theorem 5. However, assumption 4 is not satisfied for \( J(\cdot) = TV(\cdot) \) (see appendix B) and we cannot expect Hausdorff convergence of the level sets in this case (convergence rates (3.19) do not apply either).

In our experiments we are going solve the following problem

\[
\begin{align*}
\min_{\substack{u \in L^1 \cap u \geq 0}} & \; J(u) \\
\text{s.t.} & \; A^l u \leq f^l, \; A^u u \geq f^u
\end{align*}
\]

with different choices of \( A^l \) and \( A^u \). We will use both \( J(\cdot) = TV(\cdot) \) and \( J(\cdot) = TV(\cdot) + \gamma \| u \|_1 \), where \( \gamma \) is a small constant.

6.1. Reconstruction quality

Let us assume that only a slightly perturbed version \( \tilde{A} \) of the blurring operator \( A \) is available:

\[
\tilde{a}_{ij} = \max \{ a_{ij} + r_{ij} \cdot d, 0 \},
\]

where \( d = 0.05 \cdot \max_{i,j} a_{ij} \) and \( r_{ij} \) are i.i.d. uniform random numbers with support \([-1, 1]\) (i.e. the error in the operator is 5\%). Let us use the incorrect operator as if it were exact and solve (6.1) with \( A^l = A^u = A^l \) and \( J(\cdot) = TV(\cdot) \) (the results for \( J(\cdot) = TV(\cdot) + \gamma \| u \|_1 \) are similar). As demonstrated in [27], this yields highly oscillatory solutions (figure 1(b)).

Knowing the amount of noise in the operator \( A \), we can obtain lower and upper bounds \( A^l \) and \( A^u \) for the unknown exact operator \( A \) (as also explained in [27]). Using these bounds, let us now reconstruct the signal by solving problem (6.1) with \( J(u) = TV(u) \) and \( J(u) = TV(u) + \gamma \| u \|_1 \), where \( \gamma = 10^{-4} \). The results are shown in figures 1(c) and (d).

As expected, in both reconstructions the oscillations disappear and we obtain a stable reconstruction. But it is striking how much difference the small addition \( \gamma \| \cdot \|_1 \) makes on the
While the reconstruction with
\[ J(u) = TV(u) + \gamma \| u \|_1 \]
has a structure very similar to that of the exact solution, the reconstruction based on plain
\( TV(u) \)
is smooth in regions where the exact solution has jumps. We will discuss the structural
properties of the reconstruction in both cases later on in this section. Let us note that for this
signal the value of
\( TV(u) \)
is about 180, while
\( \gamma \| u \|_1 \)
is less than 0.5.

It is worth noting that the value of
\( TV(u) \)
in both cases is identical up to machine precision. Therefore, the solution in figure 1(d) also solves problem (6.1) with
\[ J(u) = TV(u) + \gamma \| u \|_1 \]
\( \gamma = 10^{-4} \)
(red dash-dotted line). This demonstrates that the choice
\[ J(u) = TV(u) + \gamma \| u \|_1 \]
can produce piecewise-constant reconstructions, while the choice
\[ J(u) = TV(u) + \gamma \| u \|_1 \]
produces them with a guarantee.

Figure 1. Reconstruction using a noisy operator yields a highly oscillatory solution (b). The interval-based approach yields stable reconstructions ((c) and (d)). However, the geometric structure of the reconstructions is different. Reconstructions with
\[ J(u) = TV(u) + \gamma \| u \|_1 \]
are piecewise-constant (d), whilst those with
\[ J(u) = TV(u) \]
are not.
6.2. Structure of solutions

The behaviour that TV demonstrates in figure 1(c) is surprising, since TV is known for introducing new jumps, referred to as staircasing [24, 29], and not for overlooking existing ones. To better understand what happened in figure 1(c), let us consider the simplest scenario, \( B_n \equiv B \equiv E, \varepsilon_n \leq u - f \leq \varepsilon_n \), and solve the following problem:

\[
\min_{u: -\varepsilon_n \leq u - f \leq \varepsilon_n} TV(u).
\]

We follow the analysis in [29]. In the one-dimensional case, the optimality condition reads as follows

\[
q' + \nu_n^u - \nu_n^l = 0,
\]

where \( q' \in \partial TV(u), |q| \leq 1 \), and \( \nu_n^u \neq 0 \) when \( u = f + \varepsilon_n \), \( \nu_n^l \neq 0 \) when \( u = f - \varepsilon_n \). If neither of the bounds is active, we get that \( q' = 0 \) and \( q \) can stay equal to 1 until either the lower or the upper bound becomes active, resulting in piecewise-monotone reconstructions between the constant regions.

If we replace \( TV(\cdot) \) with \( TV(\cdot) + \gamma \| \cdot \|_1, \gamma > 0 \), we get the following optimality condition (we assume that \( f - \varepsilon_n \geq 0 \) and omit \( \text{sign}(u) \)):

\[
\gamma + q' + \nu_n^u - \nu_n^l = 0.
\]

Therefore, if neither of the bounds are active, we get that \( q' = -\gamma < 0 \) and \( q \) cannot stay equal to 1, resulting in piecewise-constant reconstructions.

In multiple dimensions the situation is different. The optimality condition in this case is as follows

\[
\text{div} q + \nu_n^u - \nu_n^l = 0
\]

for a smooth vector-field \( q \) with \( \|q\|_{\infty} \leq 1 \) and with inactive bounds we merely get that \( \text{div} q = 0 \), which does not imply that \( q \) is constant, in contrast to the one-dimensional case.

An appropriate generalisation to multiple dimensions would be

\[
J(u) = \|\text{div} u\|_1
\]

for vector-valued images \( u \in L^1(\Omega, \mathbb{R}^n) \) [8]. In this case one indeed has

\[
\nabla q + \nu_n^u - \nu_n^l = 0
\]

for a scalar field \( q \) and \( \nabla q = 0 \) when neither of the bounds is active. Numerical experiments with vector-valued images are beyond the scope of this paper.

6.3. Comparison with Tikhonov-type regularisation

For comparison, we solve the deblurring problem using a Tikhonov-type approach combined with Morozov’s discrepancy principle (e.g. [17]). Denote the exact signal by \( f \), the noise level in the signal by \( \delta \) and the noisy signal by \( f_\delta \), so that we get that \( \|f - f_\delta\|_\infty \leq \delta \). We solve the following problem

\[
\min_{u \in L^1: a \geq 0} \|A u - f_\delta\|_\infty + \alpha J(u), \quad (6.2)
\]

where \( A \) is the noisy operator and \( \alpha \) is chosen such that

\[
\|A u_0^\alpha - f_\delta\|_\infty = C\delta, \quad (6.3)
\]
holds for the reconstructed signal $u_0^\delta$ with a constant $C$ slightly greater than 1 (we chose $C = 1.01$). This approach is equivalent to (6.1) with $A^t = A^w = \tilde{A}$, since in the absence of the operator error the constraints in (6.1) are equivalent to a constraint on $\|Au - f_0\|_{\infty}$ and $\alpha$ chosen according to (6.3) is just the Lagrange multiplier for this constraint. The result is shown in figure 2(a). Not surprisingly, we get the same kind of oscillations as in figure 1(b).
The reason for such oscillations is that the ground truth does not belong to the feasible set in (6.1) with \( A' = A'' = \tilde{A} \). A possible solution to this is modifying the feasible set so that the ground truth would become feasible. This could be achieved by replacing the constraint

\[
\|\tilde{A}u - f_\delta\|_\infty \leq \delta
\]  

(6.4)

with

\[
\|\tilde{A}u - f_\delta\|_\infty \leq \delta + h\|u\|_1, 
\]  

(6.5)

where \( h \) is the noise level in the operator, i.e. \( h \) is such that \( \|\tilde{A} - A\|_{L^1 \rightarrow L^\infty} \leq h \). Using this constraint in the context of the residual method would result in a non-convex optimisation problem, however, in the context of Tikhonov-type regularisation it can be implemented in a convex manner using the following modification of the discrepancy principle (see [35] for the theory in Hilbert spaces)

\[
\|\tilde{A}u_\alpha^\delta - f_\delta\|_\infty = C(\delta + h\|u\|_1). 
\]  

(6.6)

Since the constraint (6.5) is rather conservative (it comes from the triangle inequality) and the feasible set is large, we could expect the regulariser to have a significant impact on the reconstruction. The results obtained using this approach are shown in figures 2(b)–(d). With 5\% noise in the operator the regulariser (TV in this case) almost completely flattens out the reconstruction (figure 2(b)). With less operator noise (2.5\%) the reconstructions retain more structure, but we observe a significant loss of contrast (figures 2(c) and (d)). We notice again the same difference in the structure of the solutions produced by \( J(u) = TV(u) \) (figure 2(c)) and \( J(u) = TV(u) + \gamma\|u\|_1 \) (figure 2(d)) as in figures 1(c) and (d).

The reason of the superior performance of the approach (6.1) as compared to (6.2) with parameter choice (6.6) is that the feasible set in (6.1) much smaller than that in (6.5) (in fact, one can show that the feasible set in (6.1) is a subset of (6.5) if \( \tilde{A} = \frac{A' + A''}{2} \) and \( f_\delta = \frac{f' + f''}{2} \), see [26, theorem 2]).

6.4. Debiasing

Although the reconstruction in figure 1(d) does well at capturing the qualitative structure of the solution, it still demonstrates a systematic bias in form of a loss of contrast. The same applies to the reconstruction in figure 1(c). We will attempt to restore the contrast by optimising on the set \( M_{\varepsilon}^{\mu} \) (see (5.1)) the discrepancy \( \|\tilde{A}u - \tilde{f}\|_2 \), where \( \tilde{A} \) and \( \tilde{f} \) are the noisy operator and noisy data, respectively.

To define the the set \( M_{\varepsilon}^{\mu} \), we need to fix two constants, \( \varepsilon \) and \( C \). \( \varepsilon \) defines how close we want to stay to the model manifold; we choose to stay close and set \( \varepsilon = 10^{-6} \). Since in our case \( \mu_n > 0 \) a.e., we can drop the constant \( C \) from the definition of \( M_{\varepsilon}^{\mu} \) (see remark 25).

The results of debiasing applied to solutions in figures 1(c) and (d) are shown in figures 3(a) and (b), respectively. We see that debiasing was able to almost perfectly recover the ground truth in both cases, although the qualitative nature of the reconstruction with \( J(\cdot) = TV(\cdot) \) is quite different from that of the ground truth. Note also that naive reconstruction with the noisy operator produced oscillatory results shown in figure 1(b), whilst the two-step approach involving solving problem (2.1) and debiasing nearly perfectly recovered the ground truth.
6.5. Error bars

The results of section 5.2 allow us to provide pointwise error estimates in regions where the minimiser \( u_n \) is constant. Therefore, we need to guarantee that \( u_n \) is piecewise-constant if the exact solution \( \bar{u} \) is. We can only guarantee this for the case \( J(u) = TV(u) + \gamma \|u\|_1 \), therefore, we will only consider this case.

In order to provide a pointwise error estimate for a piecewise constant solution \( u_n \), we need to automatically determine the regions where it is constant. We proceed as follows. First observe that since \( \| \cdot \|_1 \) is continuous at \( 0 \in \text{dom}(TV) \), we have that \( \partial J(u) = \partial TV(u) + \gamma \partial \|u\|_1 \) [5]. Any \( p \in \partial TV(u) \) can be written as a divergence of some function \( q \in L^\infty \), \( \|q\|_\infty \leq 1 \), such that \( \nabla \cdot q = TV(u) \) [11]. The latter equality can be rewritten as \( -\nabla \cdot q = TV(u) \), since the gradient is the adjoint of the negative divergence. Taking into account that \( \partial \|u\|_1 = \{y \in L^\infty : \|y\|_\infty \leq 1, \|u\|_1 = (y,u)\} \), we get the following expression:

\[
p_n = y_n + \nabla \cdot q_n, \quad \|y_n\|_\infty \leq 1, \|u_n\|_1 = (y_n, u_n), \|q_n\|_\infty \leq 1, (-q_n, \nabla u_n) = TV(u_n). \tag{6.7}
\]

This function \( q_n \) contains the information about jumps of \( u_n \): whenever \( |q_n| < 1 \), \( u_n \) has to be constant [9]. Therefore, we can locate jumps of \( u_n \) by finding points where \( q_n = \pm 1 \). In general, \( q_n \) (as well as \( y_n \)) will be non-unique, but we can pick one solving the following optimisation problems:

\[
\min_{y : \|y\|_\infty \leq 1} \|y\|_1 \quad \text{s.t.} \quad \|u_n\|_1 = (y, u_n) \tag{6.8}
\]

and

\[
\min_{q : \|q\|_\infty \leq 1} \|q\|_1 \quad \text{s.t.} \quad p_n = \nabla \cdot q + \gamma y_n. \tag{6.9}
\]

Finding where \( |q_n| > 1 - \nu \) for some small constant \( \nu \) (we took \( \nu = 10^{-6} \) in our experiments), we can locate the jumps of \( u_n \). Alternatively, instead of solving (6.8) and (6.9), we can solve the following problem:
\[
\min_{q: \|q\|_\infty \leq 1} \|q\|_1 \quad \text{s.t.} \quad \nabla u_n = (-q, \nabla u_n).
\]  

(6.10)  

Both methods gave the same results in our experiments, although the method based on (6.10) was much less sensitive to the cut-off constant \(\nu\).

Having identified regions where \(u_n\) is constant, we can proceed with finding pointwise error bounds as described in section 5.2. We present results for 10\% noise in the operator (figure 4(a)) and 5\% noise (figure 4(b)). First of all, we see that the exact solution \(\tilde{u}\) is indeed contained within the bounds, together with the approximate solution \(u_n\) and the debiased solution. As expected, the error bars get tighter as the error in the operator gets smaller. One can also notice that the minimiser \(u_n\) often lies ‘on the boundary’ of the feasible set, its values coinciding with either the lower of the upper bound in almost all intervals. We also see that with 10\% operator noise the reconstruction \(u_n\) has a small additional jump at the value of the argument of around 8 that also becomes clearly visible in the lower bound (see figure 4(a)).

7. Conclusions

The paper presents a theoretical analysis of inverse problems with imperfect forward models in the setting of variational regularisation using one-homogeneous functionals. Convergence rates in terms of Bregman distances are obtained that coincide with existing results on inverse problems with exact operators, providing a natural generalisation of the existing theory.

An important aspect of the paper is the study of the interplay between the errors in the data and the operator. It turned out that errors in the data should not converge faster than the errors in the operator for the theory to work. This result is rather intuitive: there is no need to measure something more precisely than we can predict it. Therefore, it might be useful in practice to artificially decrease the quality of the data in order to match that of the operator.
the same lines goes the observation that the data should converge in such a way that there is always a uniform gap between the upper and the lower bound, see assumption 1. In the special case of TV-based regularisation we obtained results on the convergence of the level sets of the approximate solutions to those of the ground truth, building on recent work by Chambolle et al on problems with exact operators. It turned out that, unlike the classical case with an $L^2$ fidelity, TV does not guarantee the convergence of the level sets, while TV + $\gamma \| \cdot \|_1$, $\gamma > 0$, does. The deciding property in this respect is having a subdifferential at 0 with non-empty interior, which holds for TV + $\gamma \| \cdot \|_1$, $\gamma > 0$, but fails for plain TV.

Using our theoretical results we generalised the concept of two-step debiasing to problems with imperfect operators and proposed a method of obtaining asymptotic pointwise lower and upper bounds for the ground truth if it is piecewise constant, demonstrating the performance of both techniques in numerical experiments.

Acknowledgments

MB acknowledges the support of ERC via Grant EU FP 7—ERC Consolidator Grant 615216 LifeInverse. A significant portion of the work presented in this paper was done while YK was a Humboldt Fellow at the University of Münster. YK acknowledges the support of the Humboldt Foundation in that period. Currently YK holds a Newton International Fellowship sponsored by the Royal Society, whose support he also acknowledges.

Appendix A. Absolutely one-homogeneous regularisation functionals

A functional $J(\cdot)$ is called absolutely one-homogeneous if

$$J(su) = |s|J(u) \quad \forall s \in \mathbb{R}, \forall u \in L^1.$$ 

Absolutely one-homogeneous functionals are widely used in regularisation and play a crucial role, for instance, in non-linear spectral theory [13].

Absolutely one-homogeneous convex functionals have some useful properties, for example, it is obvious that $J(0) = 0$. Some further properties are listed below.

**Proposition A.1.** Let $J(\cdot)$ be a convex absolutely one-homogeneous functional and let $p \in \partial J(u)$. Then the following equality holds:

$$J(u) = (p, u).$$

**Proof.** Indeed, consider the (generalised) Bregman distance [7]

$$D_p^J(v, u) = J(v) - J(u) - (p, v - u) \geq 0 \forall v.$$

Taking $v = 0$, we get that $J(u) \leq (p, u)$, while taking $v = 2u$ and noting that $J(v) = 2J(u)$, we get that $J(u) \geq (p, u)$, hence $J(u) = (p, u)$. \qed

**Remark A.2.** The Bregman distance $D_p^J(v, u)$ in this case can be written as follows:

$$D_p^J(v, u) = J(v) - (p, v).$$

**Proposition A.3.** Let $J(\cdot)$ be a convex absolutely one-homogeneous functional. Then the convex conjugate $J^*(\cdot)$ is the characteristic function of the convex set $\partial J(0)$. 


Proof. By the definition of the convex conjugate, we have that
\[ J^*(p) = \sup_u ((p, u) - J(u)). \]

Since 0 is a feasible element, the supremum is \( \geq 0 \) for all \( p \). If \( (p, u_0) - J(u_0) > 0 \) for some \( u_0 \), then, choosing \( u = C \cdot u_0 \) with an arbitrary \( C > 0 \), we get that the supremum is unbounded and \( J^*(p) = +\infty \). Therefore, for all \( p \) s.t. \( J^*(p) < +\infty \) we have that \( (p, u) - J(u) \leq 0 \) \( \forall u \). By the definition of a subgradient we get that \( p \in \partial J(0) \). Since \( (p, u) - J(u) \leq 0 \ \forall u \) and \( \sup_u ((p, u) - J(u)) \geq 0 \), we conclude that \( J^*(p) = 0 \) whenever \( J^*(p) < +\infty \), hence the assertion. \( \square \)

An obvious consequence of the above results is the following

Proposition A.4. For any \( u \in \mathcal{U} \), \( p \in \partial J(u) \) if and only if \( p \in \partial J(0) \) and \( J(u) = (p, u) \).

Appendix B. Properties of \( \partial J(0) \)

In this section we discuss two different classes of regularisation functionals for which assumption 4 is satisfied (or not). First let us start with functionals of the form
\[ J(u) = \|u\|_1 + g(u), \]
where \( g \geq 0 \) is an absolutely one-homogeneous functional. Since \( J(u) \geq \|u\|_1 \) and convex conjugation is order reversing \([5]\), we get that
\[ \chi_{\partial J(0)}(\cdot) = \partial J^*(\cdot) \leq (\|\cdot\|_1)^* = \chi_{\|\cdot\|_\infty \leq 1}, \quad (B.1) \]
where \( \chi_X(\cdot) \) is the characteristic function of the set \( X \). Therefore, the inclusion \( \{ p : \|p\|_\infty \leq 1 \} \subset \partial J(0) \) holds and hence the condition \( 0 \in \text{int}(\partial J(0)) \) (understood in \( L^\infty \)). This proves that the regulariser \( J(\cdot) = TV(\cdot) + \gamma \|\cdot\|_1 \), \( \gamma > 0 \), satisfies assumption 4.

Now consider an absolutely one-homogeneous regularisation functional \( J(\cdot) \) such that \( J(u) = J(u + C) \) for any constant \( C \) (for example, \( J(\cdot) = TV(\cdot) \)). By the definition of a convex conjugate, we get the following equality
\[ J^*(p) = \sup_u ((p, u) - J(u)) = \sup_u ((p, u + C) - J(u + C)) = C(p, \mathbb{1}) \]
\[ = J^*(p) - C(p, \mathbb{1}). \quad (B.2) \]

Equality (B.2) implies that either \( J^*(p) = \infty \) or \( (p, \mathbb{1}) = 0 \). On the other hand, since \( J(\cdot) \) is absolutely one-homogeneous, \( J^*(\cdot) = \chi_{\partial J(0)}(\cdot) \). Therefore, \( \forall p \in \partial J(0) \) we have that \( (p, \mathbb{1}) = 0 \), which implies that \( \inf(dJ(0)) = 0 \).

Appendix C. Model manifolds

Here we derive model manifolds in \( L^1 \) related to the \( L^1 \) norm and the TV-seminorm, as well as their combination, i.e. the BV norm.

C.1. The \( L^1 \) norm

By the absolute one-homogeneity of the \( L^1 \) norm we can express its subdifferential at \( u \in L^1 \) by the following expression

\[ \partial J^* \text{, where } J^* = \|\cdot\|_1 \standalone\]
\[ \partial \|u\|_1 = \{ r \in L^\infty \mid \|u\|_1 = (r, u), \|v\|_1 \geq (r, v) \forall v \in L^1 \}. \]  

(C.1)

From this one can easily derive that

\[ \|r\|_\infty \leq 1 \quad \text{and} \quad \int_\Omega |u| - ru \, dx = 0. \]  

(C.2)

Obviously, for any two numbers \( \eta, \nu \in \mathbb{R} \) such that \( |\nu| \leq 1 \) we have that \( |\nu| - \nu \eta \geq 0 \), with equality holding if and only if either of the numbers is zero or \( \nu = \text{sign}(\eta) \). Hence the integrand in (C.2) has to be nonnegative for a.e. \( x \in \Omega \), and the vanishing integral implies that \( |u| = ru \) a.e. Since \( |r| \leq 1 \), we get that \( r = \text{sign}(u) \) on every set of nonzero measure where \( u \neq 0 \).

Consider now \( r \in \partial \|u\|_1 \), and assume that \( |r| < 1 \) whenever \( u \) vanishes (in the a.e. sense). We need to compute the related model manifold

\[ M = \{ v \in L^1 \mid \|v\|_1 - (r, v) = \int_\Omega |v| - rv = 0 \}, \]

where \( D^*_\|\|_1(v, u) = \|v\|_1 - (r, v) \) is the (generalised) Bregman distance. By the same argument as before, the integrand has to be positive a.e. and the integral may only vanish if the integrand vanishes for a.e. \( v \in \Omega \). Under the above assumptions on \( r \), whenever \( u \) vanishes we have that \( |r| < 1 \) and therefore \( v \) also has to vanish, and whenever \( u \neq 0 \), \( v \) can be any positive number with the same sign as \( u \). In sum, we get the following expression for the model manifold

\[ M = \{ v \in L^1 \mid \text{supp}(v) \subset \text{supp}(u), v = \lambda \text{sign}(u) \text{ whenever } u \neq 0, \lambda \geq 0 \}, \]

where \text{supp} has to be interpreted in the a.e. sense.

C.2. The TV seminorm

The argumentation for TV is almost identical. A function \( u \in L^1 \) lies in \( BV(\Omega) \) if and only if its distributional derivative \( Du \) is a finite Radon measure [1], and

\[ \|Du\|_{\text{TV}} = TV(u), \]

where \( \| \cdot \|_{\text{TV}} \) denotes the Radon norm. Furthermore, \( Du \) possesses a polar decomposition \( Du = f_{Du}|Du| \), where \( |Du| \) denotes the total variation of \( u \) and \( \|f_{Du}\|_1 = 1 \) for the related density function \( f_{Du} \). By the chain rule, \( p \in \partial TV(u) \) if and only if \( p = D^*q \) for some \( q \) in the subdifferential of the \( L^1 \) norm. Assuming that \( q \in C_0(\Omega, \mathbb{R}^d) \), i.e. \( q \) lies in the predual instead of only in the dual space of the space of finite Radon measures, we know by the computations in [28] that \( q = f_{Du}|Du|\)-a.e. In other words, the ‘vector field’ \( q \) of the subgradient may be decomposed into direction and magnitude, indicating the direction and the hight of jumps across the edges.

We may now rewrite

\[ D^*_{TV}(v, u) = D^*_{\|\|_1}(Dv, Du) = \int_\Omega 1 - q \cdot f_{Du} \, d|Dv|. \]

This, by the same argument as for \( L^1 \), vanishes if and only if either \( |Dv| = 0 \) or \( f_{Dv} = q = f_{Du} \), implying that the jump set and its direction of \( v \) has to be contained in the jump set of \( u \). In other words, \( v \) may only jump where \( u \) jumps as well. Note that in this case the magnitude of the jump, i.e. \( |Dv| \), can be arbitrary.

It should be mentioned that the (technical) assumption of a continuous subgradient is crucial for this illustration, but not necessarily for the result, meaning that a zero Bregman
distance with respect to $\text{TV}$ is still well-defined if $q$ is not continuous. However, then it is hard to say anything about the behavior, and we refer to [28] for further information. Moreover, it is possible that $|q| = 1$ on a $\{Du\}$-zero set, i.e. the vector field of the subgradient might be saturated where $u$ is constant. In this case, indeed it is possible for $v$ to jump even though $u$ is flat. In practice, however, this situation is rarely observed, or has to be enforced by assumption as in the above $L^1$ case.

C.3. The BV norm

Let $J(u) = \text{TV}(u) + \gamma \|u\|_1$ for $\gamma > 0$. Since $\|\cdot\|_1$ is continuous everywhere on $L^1$ we have that [5, theorem 4.4.3 and lemma 4.3.1]

$$\text{cont}(\|\cdot\|_1) \cap \text{dom}(\text{TV}) \neq \emptyset,$$

and

$$\partial J(v) = \partial \text{TV}(v) + \gamma \partial \|u\|_1.$$

Hence for $p \in \partial J(v)$ we have that there exists $q \in \partial \text{TV}(v)$ and $r \in \partial \|u\|_1$ such that $p = q + \gamma r$. Considering the Bregman distance $D^p_J(u, v)$, we get the following expression

$$D^p_J(u, v) = \text{TV}(u) + \gamma \|u\|_1 - (q + \gamma r, u)$$

$$= \text{TV}(u) - (q, u) + \gamma (\|u\|_1 - (r, u))$$

$$= D^q_{\text{TV}}(u, v) + \gamma D^r_{\|\cdot\|_1}(u, v).$$

Since both Bregman distances are non-negative we have that

$$D^p_J(u, v) = 0 \iff D^q_{\text{TV}}(u, v) = 0 \quad \text{and} \quad D^r_{\|\cdot\|_1}(u, v) = 0,$$

implying that the manifold with respect to $J$ and $p$ contains all elements sharing the same jump set (including its direction) and the same (signed) support.

ORCID iDs

Yury Korolev https://orcid.org/0000-0002-6339-652X
Julian Rasch https://orcid.org/0000-0002-0732-8805

References

[1] Ambrosio L, Fusco N and Pallara D 2000 Functions of Bounded Variation and Free Discontinuity Problems (Oxford: Clarendon)
[2] Anderson E and Nash P 1987 Linear Programming in Infinite-Dimensional Spaces: Theory and Applications (New York: Wiley)
[3] Benning M and Burger M 2011 Error estimates for general fidelities Electron. Trans. Numer. Anal. 38 44–68
[4] Bonnans J F and Shapiro A 1998 Optimization problems with perturbations: a guided tour SIAM Rev. 40 228–64
[5] Borwein J and Zhu Q 2005 Techniques of Variational Analysis (Berlin: Springer)
[6] Bredies K, Kunisch K and Pock T 2011 Total generalized variation SIAM J. Imaging Sci. 3 492–526
[7] Bregman L 1967 The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming USSR Comput. Math. Math. Phys. 7 200–17
[8] Briani A, Chambolle A, Novaga M and Orlandi G 2011 On the gradient flow of a one-homogeneous functional Confluentes Math. 03 617–35
[9] Brinkmann E-M, Burger M, Rasch J and Sutour C 2017 Bias reduction in variational regularization J. Math. Imaging and Vis. 59 534–66
[10] Burger M and Osher S 2004 Convergence rates of convex variational regularization Inverse Problems 20 1411
[11] Burger M and Osher S 2013 A guide to the tv zoo Level-Set and PDE-Based Reconstruction Methods ed M Burger and S Osher (Berlin: Springer)
[12] Burger M, Papafitsoros K, Papoutsellis E and Schönlieb C-B 2016 Inifmal convolution regularisation functionals of bv and P spaces J. Math. Imaging Vis. 55 343–69
[13] Burger M et al 2016 Spectral decompositions using one-homogeneous functionals J. Math. Imaging Vis. 55 343–69
[14] Burger M et al 2016 Spectral decompositions using one-homogeneous functionals SIAM J. Imaging Sci. 9 1374–408
[15] Chambolle A, Duval V, Peyré G and Poon C 2017 Geometric properties of solutions to the total variation denoising problem Inverse Problems 33 015002
[16] Deledalle C-A, Papadakis N and Salmon J 2015 On debiasing restoration algorithms: applications to total-variation and nonlocal-means Scale Space and Variational Methods in Computer Vision ed M Burger and S Osher (Berlin: Springer) pp 129–41
[17] Deledalle C-A, Papadakis N, Salmon J and Vaiter S 2017 Clear: covariant least-square refitting with applications to image restoration SIAM J. Imaging Sci. 10 243–84
[18] Engl H, Hanke M and Neubauer A 1996 Regularization of Inverse Problems (Berlin: Springer)
[19] Grant M and Boyd S 2008 Graph implementations for nonsmooth convex programs Recent Advances in Learning, Control (Lecture Notes in Control and Information Sciences) ed V Blondel et al (Berlin: Springer) pp 95–110
[20] Grant M and Boyd S 2014 CVX: matlab software for disciplined convex programming, version 2.1 http://cvxr.com/cvx
[21] Grasmair M, Haltmeier M and Scherzer O 2011 The residual method for regularizing ill-posed problems Appl. Math. Comput. 218 2693–710
[22] Hintermüller M, Hinze M and Rautenberg N 2012 Optimal control of total variation regularization with applications to image denoising SIAM J. Imaging Sci. 5 1130–50
[23] Ivanov V K, Vasin V V and Tanana V P 2002 Theory of Linear Ill-Posed Problems and its Applications (Berlin: de Gruyter & Co)
[24] Jalalzai K 2016 Some remarks on the staircasing phenomenon in total variation-based image denoising J. Math. Imaging Vis. 54 256–68
[25] Kiwiel K 1997 Proximal minimization methods with generalized bregman functions SIAM J. Control Optim. 35 1142–68
[26] Korolev Y 2014 Making use of a partial order in solving inverse problems: II Inverse Problems 30 085003
[27] Korolev Y and Lehmkuhl J 2018 Image reconstruction with imperfect forward models and applications in deblurring SIAM J. Imaging Sci. 11 197–218
[28] Rasch J, Brinkmann E-M and Burger M 2017 Joint reconstruction via coupled bregman iterations with applications to pet-mr imaging Inverse Problems 34 014001
[29] Ring W 2000 Structural properties of solutions to total variation regularization problems ESAIM: M2AN 34 799–810
[30] Robinson S M 1975 Stability theory for systems of inequalities. Part i: linear systems SIAM J. Numer. Anal. 12 754–69
[31] Robinson S M 1976 Stability theory for systems of inequalities. Part ii: differentiable nonlinear systems SIAM J. Numer. Anal. 13 497–513
[32] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms Physica D 60 259–68
[33] Sawatzky A et al 2013 Em-tv methods for inverse problems with poission noise Level Set and PDE Based Reconstruction Methods in Imaging (Lecture Notes in Mathematics vol 2090) (Berlin: Springer)
[34] Schaefer H 1958 Halbgeordnete lokalkonvexe vektorräume Math. Ann. 135 115–41
[35] Tikhonov A N, Goncharsky A V, Stepanov V V and Yagola A G 1995 Numerical Methods for the Solution of Ill-Posed Problems (Dordrecht: Kluwer)