Subgradient-based Lavrentiev regularisation of monotone ill-posed problems

Markus Grasmair
markus.grasmair@ntnu.no

Fredrik Hildrum
arxiv@fredrik.hildrum.net

Department of Mathematical Sciences,
NTNU – Norwegian University of Science and Technology,
7491 Trondheim, Norway

October 6, 2024

Abstract. We introduce subgradient-based Lavrentiev regularisation of the form

$$\mathcal{A}(u) + \alpha \partial \mathcal{R}(u) \ni f^\delta$$

for linear and nonlinear ill-posed problems with monotone operators $\mathcal{A}$ and general regularisation functionals $\mathcal{R}$. In contrast to Tikhonov regularisation, this approach perturbs the equation itself and avoids the use of the adjoint of the derivative of $\mathcal{A}$. It is therefore especially suitable for time-causal problems that only depend on information in the past and allows for real-time computation of regularised solutions. We establish a general well-posedness theory in Banach spaces and prove convergence-rate results with variational source conditions. Furthermore, we demonstrate its application in total-variation denoising in linear Volterra integral operators of the first kind and parameter-identification problems in semilinear parabolic PDEs.

Key words and phrases: nonlinear ill-posed problems; Lavrentiev regularisation; monotone operators; convergence rates; variational source conditions; integral equations; parameter-identification problems.

Mathematics Subject Classification (2010): 47H05; 47J06; 45Q05; 65J20; 65R30.

1 Introduction

In this paper we study stable approximations of a class of monotone and possibly nonlinear ill-posed equations of the form

$$\mathcal{A}u = f^\dagger$$

in the presence of noisy data $f^\delta$ satisfying

$$\|f^\dagger - f^\delta\| \leq \delta,$$

where $\delta > 0$ denotes the noise level. Here $\mathcal{A} : \mathcal{X} \to \mathcal{X}^*$ is an operator mapping a reflexive Banach space $\mathcal{X}$ into its dual $\mathcal{X}^*$, and for simplicity we usually write $\mathcal{A}u$ instead of $\mathcal{A}(u)$. 
A common approach to stabilising (1) with noisy data is Tikhonov regularisation (see for instance [15, 34]), which amounts to minimising the functional

\[
\mathcal{T}(u) := \frac{1}{2} \| Au - f_\delta \|^2 + \alpha R(u)
\]

over \( \mathbb{X} \). The first term guarantees that the minimisers \( u_\delta \) approximately solve (1), \( R : \mathbb{X} \to [0, \infty) \) is a suitable lower-semicontinuous and convex regularisation functional that encodes qualitative a priori information about the solution, and \( \alpha > 0 \) is the regularisation parameter balancing the two terms.

One classical example of ill-posed inverse problems is the solution of linear Volterra equations of the first kind [19]. Here we are given some real-valued, time-dependent function \( f^\dagger \in L^2(0, T) \) supported on an interval \( [0, T] \), and a linear Volterra operator

\[
Au(t) := \int_0^t k(t, s) u(s) \, ds,
\]

with \( t \leq T \) and square-integrable kernel \( k \in L^2([0, T]^2) \). If we have the prior information that the true underlying signal \( u^\dagger \) in (1) is piecewise constant, it makes sense to apply total-variation regularisation, in which case the regularisation functional \( R : L^2(0, T) \to [0, \infty) \) is the total variation (TV) of the distributional derivative of \( u \) over \([0, T] \), that is,

\[
R(u) := |u|_{TV} := |Du|(0, T),
\]

as discussed in more detail in Section 6. This type of regularisation has for instance been studied in a related, two-dimensional setting [1, 33].

Although this choice of \( R \) yields a nonsmooth optimisation problem, the additional effort compared to quadratic regularisation is often worthwhile because of the improved reconstructions that are structurally much closer to the true solution. Classical regularisation with a Sobolev regularisation term \( R(u) = \|u'\|^2_{L^2} \) will lead to smooth regularised solutions, whereas the TV term is capable of reconstructing discontinuities. In addition, the recent advances in nonsmooth optimisation have led to the development of efficient numerical algorithms (see for instance [12, 24]), which make nonquadratic regularisers feasible for practical problems.

A particularly efficient algorithm is available for TV regularisation in the case where \( A \) is the identity operator on \( L^2(0, T) \) (which is not of type (3)). Here, Tikhonov's method (2) with TV (4) can be seen as a spline-based nonlinear regression method [26, 37]. Moreover, the (nonlinear) problem of minimising the Tikhonov functional (2) can in this case be solved efficiently by means of the taut-string algorithm [13, 14, 17], which is a dynamic-programming method for the solution of a dual problem.

In the case of a Volterra operator of the form (3), the situation is different. There, standard optimisation theory implies that the minimisers of the Tikhonov functional are precisely the solutions of the variational inclusion

\[
0 \in \partial \mathcal{T}(u_\delta) = A^*(Au_\delta - f_\delta) + \alpha \partial R(u_\delta),
\]

where \( \partial \) denotes the subdifferential. This equation, however, requires the adjoint of \( A \), which reverses the time-causal structure. That is, whereas the operator \( A \) consists of an integration backward in time,
its adjoint $A^*$ consists in the integration forward in time,

$$A^*u(t) = \int_t^T k(s, t) u(s) \, ds,$$

which at each point depends completely on future information. As a consequence, Tikhonov regularisation is not applicable in situations where the input data consists of a continuous data stream and one tries to solve the inverse problem on-the-fly in real time. In addition, it is impossible to implement the same dynamic-programming idea as in the taut-string algorithm in order to obtain a similarly efficient numerical solution method.

An alternative to Tikhonov regularisation for the stable solution of (1) is Lavrentiev regularisation (see for instance [2, 4, 8, 16, 21, 25, 27, 29, 30, 38]) consisting in solving the operator equation

$$Au + \alpha u = f^\delta$$

in the Hilbert-space setting, or more generally

$$Au + \alpha J(u) \ni f^\delta$$

in the Banach-space equivalent, where $J := \partial \left( \frac{1}{2} \| \cdot \|^2 \right)$ is the normalised duality mapping (16). This regularisation method is applicable for monotone operators, that is, for those $A$ that satisfy

$$\langle Au - Av, u - v \rangle \geq 0 \quad \text{for all } u, v \in X.$$

In contrast to Tikhonov regularisation (2), Lavrentiev regularisation (6) simply perturbs the equation itself, has a local nature, and avoids the introduction of noncausal effects in the form of adjoints.

Suggestively, we now propose to regularise (1) by solving the variational inclusion

$$Au + \alpha \partial R(u) \ni f,$$  

which combines the general regularisation functionals in the Tikhonov approach with the simplicity of Lavrentiev regularisation. One may thus view (7) either as the optimality conditions (5) with the adjoint removed or the inclusion (6) for more general subgradients. Observe in passing that we display (7) for a generic $f \in X^*$ (say, $f^\dagger$ or $f^\delta$) in order to accommodate the upcoming analysis.

In this paper, we show that this subgradient-based Lavrentiev regularisation – by solving the inclusion problem (7) – yields a completely new, well-posed regularisation method for a general class of monotone operators $A$ and regularisers $R$ in reflexive Banach spaces. To this end, it is important to emphasise that our setting does not require coercivity of either $A$ or $R$, as will be discussed in more detail in Section 2.

In addition to showing well-posedness, we derive asymptotic error estimates under the assumption of variational source conditions. There, we follow the ideas of [21], where the same approach was used for the analysis of classical Lavrentiev regularisation. All these results as well as the requirements for $A$ and $R$ are collected in Section 2. Moreover, Section 3 provides a review of the functional-analytic background that is required for the well-posedness proof. We discuss a quantitative formulation of the Browder–Minty theorem concerning existence and uniqueness of monotone operator equations that includes an explicit bound on the solution. We then provide the proof of well-posedness in
Section 4 and the derivation of convergence rates in Section 5. Finally, we demonstrate the new approach (7) with TV denoising in both convolutional Volterra integral equations in Section 6 and nonlinear parameter-identification problems for parabolic PDEs in Section 7. This includes new conditions on the convolution kernel (3) that guarantee strict monotonicity of \( A \), and a detailed analysis of coercivity in the PDE problem. We also briefly present numerical results based on a generalisation of the taut-string method, which will be the subject of a forthcoming paper.

### 2 Main results

#### 2.1 Assumptions

We study subgradient-based Lavrentiev regularisation (7) under the following conditions and refer to Section 3 for functional-analytic definitions.

**Assumption 2.1.**

i) **Space:** \((X, \|\cdot\|)\) is a real, reflexive Banach space with topological dual \((X^*, \|\cdot\|_{X^*})\), and \(\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{X^* \times X} \) denotes the duality pairing.

ii) **Operator:** \(A : X \to X^*\) is strictly monotone and hemicontinuous.

iii) **Regulariser:** \(R : X \to (-\infty, +\infty]\) is proper, convex, and lower semicontinuous.

iv) **Solution:** There exists a solution \(u^\dagger \in \text{dom } R\) to the original problem (1), that is, \(Au^\dagger = f^\dagger\).

v) **Coercivity:** The sublevel set \(\{u \in X : \|u\| + R(u) \leq C\}\) is compact for all \(C > 0\).

vi) **Growth:** For all (sufficiently large) \(C > 0\) and some \(t_0 > 0\) we have that

\[
\lim_{r \to \infty} \inf_{u \in U_C} \langle A(ru + (1-r)u^\dagger), u - u^\dagger \rangle = \infty, \tag{8}
\]

where \(U_C := \{u \in X : \|u - u^\dagger\| = t_0 \text{ and } R(u) \leq C\}\).

Note that both \(u^\dagger\) and the regularised solutions are necessarily unique since \(A\) is injective by strict monotonicity. Moreover, if \(A\) is linear, then (hemi)continuity comes for free, since linear monotone operators are bounded. As regards the natural assumption \(u^\dagger \in \text{dom } R\), we point out that for classical Lavrentiev regularisation it is intrinsically satisfied and means simply that \(u^\dagger\) is an element of \(X\).

Switching focus to Assumption 2.1 v), we emphasise that \(\text{dom } R\) is not necessarily a compact subset of \(X\), which otherwise easily would have guaranteed coercivity. In fact, in the typical case when \(R\) is the seminorm of a (not necessarily compact) subspace of \(X\), such as when \(R\) equals total variation or a Sobolev seminorm in Lebesgue spaces (see Section 6), then \(\text{dom } R\) is not compact in \(X\). As such, Assumption 2.1 v) is weaker than assuming the compactness of \(\text{dom } R\).

Furthermore, the growth condition in Assumption 2.1 vi) holds trivially for strictly monotone linear \(A\) and shows up naturally in the proof of well-posedness. On the practical side, we also prove directly that it is satisfied in the nonlinear parameter-identification problem in Section 7. To aid further applications, we include in addition an alternative condition which may be easier to verify.
Proposition 2.2 (Alternative growth conditions). Assume that Assumption 2.1 i–iii) and v) is satisfied. Let \( u_{\text{ref}} \in \mathbb{X} \) be a fixed reference point and assume that there exists a function \( \gamma : (0, \infty) \to (0, \infty) \) with superlinear growth,

\[
\lim_{r \to \infty} \frac{\gamma(r)}{r} = \infty,
\]

such that the “spherical growth condition”

\[
\sup_{\|u\| = 1} \frac{\|A(ru + (1 - r)u_{\text{ref}})\|_{\mathbb{X}}}{\gamma(r)} \xrightarrow{r \to \infty} 0
\]  

(9)

holds and the “directional growth condition”

\[
\langle A(ru + (1 - r)u_{\text{ref}}) - Au_{\text{ref}}, u - u_{\text{ref}} \rangle \geq \frac{\gamma(r)}{r} \langle Au - Au_{\text{ref}}, u - u_{\text{ref}} \rangle
\]  

(10)

holds for all \( u \in \mathbb{X} \) with \( \|u - u_{\text{ref}}\| = 1 \) and sufficiently large \( r > 0 \). Then the growth condition in Assumption 2.1 vi) is satisfied.

In particular, this applies with \( u_{\text{ref}} = 0 \) and \( \gamma(r) = r^2 \) in the case where \( A \) is a strictly monotone linear operator.

The proof of Proposition 2.2 can be found in Appendix A. We also note that condition (9) is superfluous if we can choose \( u_{\text{ref}} = u^\dagger \); for details, we refer to Appendix A.

Observe that (10) is a relatively mild condition, which implies a form of “weak directional coercivity” of \( A \) relative to the point \( u_{\text{ref}} \) in the sense that

\[
\langle A(u_{\text{ref}} + rv) - Au_{\text{ref}}, v \rangle \xrightarrow{r \to \infty} \infty
\]  

(11)

for all \( v \in \mathbb{X} \) with \( \|v\| = 1 \). Since the limit (11) is not necessarily uniform over the unit sphere, that is, \( r \) may be a function of \( v \)—or there might exist a sequence of “bad” \( u \)'s for which the last factor

\[
\langle Au - Au_{\text{ref}}, u - u_{\text{ref}} \rangle
\]

goes to 0 sufficiently fast in (10)—this neither leads to coercivity of \( A \) nor \( A + \alpha \partial \mathcal{R} \) of itself. The compactness of the sublevel set in Assumption 2.1 v) together with strict monotonicity and (9), however, ensure that the possibly bad sequence of \( u \)'s is not an issue—even when the noise level \( \delta \) and the regularisation parameter \( \alpha \) go to 0 appropriately. The interpretation of (8) in light of (11) might thus be labeled as weak directional coercivity relative to \( u^\dagger \) uniformly over \( U_C \).

2.2 Well-posedness. We obtain well-posedness of subgradient-based Lavrentiev regularisation (7) under Assumption 2.1 in the following manner. Remember that \( f^\dagger \) is the exact data for which \( Au^\dagger = f^\dagger \), that \( f^\delta \) is its noisy counterpart, and that \( f \) refers to any element of \( \mathbb{X}^* \).

Theorem 2.3 (Existence). For every \( f \in \mathbb{X}^* \) and \( \alpha > 0 \), there exists a unique solution \( u_\alpha \in \text{dom} \mathcal{R} \) to the regularised problem (7).
Theorem 2.4 (Stability). For every \( f \in \mathbb{X}^* \) and fixed \( \alpha > 0 \), if \( f_k \rightarrow f \) in \( \mathbb{X}^* \), then
\[
u_{a,k} \rightarrow \nu_a \quad \text{in} \quad \mathbb{X}, \quad A\nu_{a,k} \rightarrow A\nu_a \quad \text{weakly in} \quad \mathbb{X}^*, \quad \text{and} \quad \mathcal{R}(\nu_{a,k}) \rightarrow \mathcal{R}(\nu_a),
\]
where \( \nu_{a,k} \) and \( \nu_a \) solve (7) with right-hand sides \( f_k \) and \( f \), respectively.

Theorem 2.5 (Convergence). Let \( \alpha = \alpha(\delta) \) be any a priori parameter choice for which
\[
\alpha \rightarrow 0 \quad \text{and} \quad \frac{\delta}{\alpha} \quad \text{is bounded} \quad \text{as} \quad \delta \rightarrow 0,
\]
and denote by \( \nu^\delta_a := \nu^\delta_{a(\delta)} \) the regularised solutions of (7) with \( f^\delta \in \mathbb{X}^* \) satisfying \( \|f^\delta - f^\dagger\|_{\mathbb{X}^*} \leq \delta \). Then
\[

\nu^\delta_a \rightarrow f^\dagger \quad \text{in} \quad \mathbb{X}, \quad A\nu^\delta_a \rightarrow f^\dagger \quad \text{weakly in} \quad \mathbb{X}^*, \quad \text{and} \quad \mathcal{R}(\nu^\delta_a) \rightarrow \mathcal{R}(f^\dagger)
\]
as \( \delta \rightarrow 0 \).

We wish to draw attention to the unusually weak relationship between \( \delta \) and \( \alpha \) in Theorem 2.5 in that merely boundedness of \( \delta / \alpha \) suffices for convergence. This contrasts results both for Tikhonov-type regularisations
\[
\nu^\delta_a = \arg \min_{u \in \mathbb{X}} \left( \|Au - f^\delta\|_{\mathbb{X}^*}^p + \alpha \mathcal{R}(u) \right)
\]
for \( p > 1 \), where one has to require that \( \delta^p / \alpha \rightarrow 0 \) as \( \alpha(\delta) \rightarrow 0 \); see for instance [22, Proposition 3.1], and for standard Lavrentiev regularisation (6), in which the condition \( \delta / \alpha \rightarrow 0 \) is needed to guarantee convergence [2, Theorem 2.2.4], [21, Proposition 1]. It should still be noted that one has weak convergence in \( \mathbb{X} \) under the parameter choice (12) for Lavrentiev regularisation [2, Theorems 2.1.6 and 2.2.6], which rationalises our need for a compactness condition in Assumption 2.1 v) in order to obtain strong convergence.

Remark 2.6. If \( \mathcal{A} \) is linear, then weak convergence in the codomain \( \mathbb{X}^* \) in Theorems 2.4 and 2.5 always becomes strong convergence, because linear monotone maps are continuous.

We can also upgrade weak convergence \( A\nu^\delta_a \rightarrow A\nu^\dagger \) in Theorem 2.5 to strong convergence if \( \nu^\dagger \) is a continuity point of \( \mathcal{R} \) (which is the case if \( \mathcal{R} = \frac{1}{2}\|\cdot\|^2 \) in classical Lavrentiev regularisation in Hilbert spaces). In this setting, \( \{\nu^\delta_a\}_{\delta} \) lies in the interior of dom \( \mathcal{R} \), ignoring a finite number of \( \nu^\delta_a \)'s if necessary. By the Rockafellar–Veselý theorem [32, Theorem 1], \( \partial \mathcal{R} \) is locally bounded in its interior domain—meaning that for every \( u \in \text{int(dom} \mathcal{R}) \) there exists an \( \epsilon > 0 \) such that the image of an \( \epsilon \)-ball around \( u \) is bounded. This implies that \( \partial \mathcal{R}(\{\nu^\delta_a\}_{\delta}) \) is bounded by compactness of \( \{\nu^\delta_a\}_{\delta} \). In particular, the elements \( \xi^\delta_a \in \partial \mathcal{R}(\nu^\delta_a) \) satisfying \( A\nu^\delta_a + \alpha \xi^\delta_a = f^\delta \) are uniformly bounded in \( \mathbb{X}^* \) as \( \delta \) and \( \alpha \) go to 0, from which it follows that \( A\nu^\delta_a \rightarrow f^\dagger \).

Finally, we mention that it may sometimes be beneficial to consider the modified variant
\[
A\nu + \alpha \delta \mathcal{R}(u - u_{\text{init}}) \ni f^\delta
\]
of the regularisation method (7) in which \( u_{\text{init}} \in \mathbb{X} \) denotes an initial guess. If we assume that \( (u^\dagger - u_{\text{init}}) \in \text{dom} \mathcal{R} \), all the results and proofs carry over to this case, with the only difference being
that
\[ R(u_{a,k} - u_{\text{init}}) \rightarrow R(u_a - u_{\text{init}}) \]
replaces \[ R(u_{a,k}) \rightarrow R(u_a) \]
in Theorem 2.4, and in Theorem 2.5 that
\[ R(u^\delta_{a} - u_{\text{init}}) \rightarrow R(u^\dagger - u_{\text{init}}) \]
replaces \[ R(u^\delta_a) \rightarrow R(u^\dagger) \].

### 2.3 Convergence rates

As regards quantitative error estimates for (7), we use the method of variational source conditions established in [18, 20] and generalised in [21] to classical Lavrentiev regularisation. That is, we consider general variational source conditions of the form
\[ D(u, u^\dagger) \leq R(u) - R(u^\dagger) + \varphi\left(\langle Au - Au^\dagger, u - u^\dagger\rangle\right) \]
for all \( u \in \mathcal{M} \), (13)

where \( D: X \times X \to [0, \infty] \) is any distance-like function and \( \varphi: [0, \infty) \to [0, \infty) \) denotes an index map; see Section 5. The set \( \mathcal{M} \subseteq X \), typically a ball, must contain all regularised solutions \( u^\delta_a \) of interest for sufficiently small \( \delta > 0 \) and \( \alpha > 0 \) chosen appropriately. A priori we do not assume any further properties other than nonnegativity of \( D \). This gives the following main result, where \( \psi \) denotes the convex (Fenchel) conjugate of \( \varphi^{-1} \) (defined in Section 5). Notationally, we write \( A \lesssim B \) when \( A \leq cB \) for an independent constant \( c > 0 \), and \( A \sim B \) means that \( A \lesssim B \lesssim A \).

**Theorem 2.7 (General convergence rates).** Assume that the variational source condition (13) holds and that the ratio \( \delta / \alpha \) is uniformly bounded. Then there exists a constant \( C > 0 \) such that
\[ D(u^\delta_a, u^\dagger) \leq C \frac{\delta}{\alpha} + \frac{\psi(\alpha)}{\alpha} \]
whenever \( \delta > 0 \) and \( \alpha > 0 \) are sufficiently small. In particular, a parameter choice \( \alpha \sim \psi^{-1}(\delta) \) gives the convergence rate
\[ D(u^\delta_a, u^\dagger) \lesssim \frac{\delta}{\psi^{-1}(\delta)} \]
as \( \delta \to 0 \).

In practice, one considers distance measures \( D \) like the Bregman distance with respect to the functional \( R \) or powers of the norm on \( X \). Specialising to the latter case and with \( \varphi \) being of Hölder type in (13), the estimates and convergence rates can be further improved, leading to the following result.

**Theorem 2.8.** Assume that the Hölder-type variational source condition
\[ c_1 \|u - u^\dagger\|^r \leq R(u) - R(u^\dagger) + c_2 \left(\langle Au - Au^\dagger, u - u^\dagger\rangle\right)^{1/p} \]
holds for some \( r > 1, p > 1 \), and \( c_1, c_2 > 0 \). Then the parameter choice
\[ \alpha \sim \delta^\left(\frac{r(p-1)}{rp-1}\right) \]
gives the convergence rate
\[ \|u^\delta_a - u^\dagger\| \lesssim \delta^\left(\frac{1}{rp-1}\right) \]
as \( \delta \to 0 \).
In particular, when \( r = 2 = p \) in (14), we deduce from the variational source condition
\[
\|u - u^\dagger\|^2 \lesssim \mathcal{R}(u) - \mathcal{R}(u^\dagger) + (\langle Au - Au^\dagger, u - u^\dagger \rangle)^{1/2}
\]
a convergence rate
\[
\|u_\alpha^\delta - u^\dagger\| \lesssim \delta^{1/3} \quad \text{if } \alpha \sim \delta^{2/3},
\]
which is in line with expected results for Lavrentiev regularisation [21, Theorem 5]. In fact, Theorem 2.8 generalises the Hölder rates in [21, Section 3.3] from \( r = 2 \) to \( r > 1 \), and it should be noted that all the convergence-rates results in our work applies to the classical case \( \mathcal{R} = \frac{1}{2}\|\cdot\|^2 \) as well. We also point out that if \( \mathcal{A} \) is positively homogeneous—for example, if \( \mathcal{A} \) is linear—then Theorem 2.8 is realistic only for \( p \geq 2 \); see the discussion in Remark 5.2.

3 Functional-analytic background

We refer to [2, 6, 7, 35, 39] for in-depth analysis and extensions of the theory presented freely in this section, noting that [2] thoroughly covers regularisation of ill-posed problems and that [7] focuses on the Hilbert-space setting.

Our main tool in the study of (7) is the fundamental theorem in monotone operator theory.

**Theorem 3.1 (Browder–Minty [11])**. Maximally monotone, coercive operators \( \mathbb{X} \Rightarrow \mathbb{X}^\ast \) are surjective.

This result generalises the fact that continuous functions on \( \mathbb{R} \) which are monotone and coercive, also are surjective. A multivalued map \( \mathcal{A} : \mathbb{X} \Rightarrow \mathbb{X}^\ast \) is monotone on its domain \( \text{dom} \mathcal{A} := \{u \in \mathbb{X} : \mathcal{A} u \neq \emptyset\} \) if
\[
\langle f - g, u - v \rangle \geq 0
\]
for all \( u, v \in \text{dom} \mathcal{A} \) and \( f \in \mathcal{A} u \) and \( g \in \mathcal{A} v \), where we usually write \( \mathcal{A} u \) instead of \( \mathcal{A}(u) \) even though \( \mathcal{A} \) may be nonlinear. In particular, a single-valued operator is monotone if \( \langle Au - Av, u - v \rangle \geq 0 \) for all \( u, v \in \text{dom} \mathcal{A} := \{u \in \mathbb{X} : \mathcal{A} u \in \mathbb{X}^\ast\} \). Moreover, \( \mathcal{A} \) is strictly monotone if equality in (15) only holds when \( u = v \), in which case it follows that \( \mathcal{A} \) is injective on this set. \( \mathcal{A} \) is further said to be maximally monotone if it has no proper monotone extension—that is, if \( \langle f - g, u - v \rangle > 0 \) for all \((v, g)\) \( \in \mathcal{A} \), then \((u, f)\) \( \in \mathcal{A} \), where we have identified \( \mathcal{A} \) with its graph.

Of particular interest is the subdifferential \( \partial \mathcal{R} : \mathbb{X} \Rightarrow \mathbb{X}^\ast \) of a convex function \( \mathcal{R} : \mathbb{X} \rightarrow (-\infty, \infty] \) defined by
\[
\xi \in \partial \mathcal{R}(u) \quad \iff \quad \mathcal{R}(v) - \mathcal{R}(u) \geq \langle \xi, v - u \rangle \quad \text{for all } v \in \mathbb{X},
\]
which is readily seen to be monotone. If \( \mathcal{R} \) additionally is lower semicontinuous and proper, meaning that \( \text{dom} \mathcal{R} := \{u \in \mathbb{X} : \mathcal{R}(u) < \infty\} \) is nonempty, then the Moreau–Rockafellar theorem [31, Theorem 1] shows that \( \partial \mathcal{R} \) is maximally monotone and proper. A special example is the normalised duality mapping
\[
J(u) := \partial \left( \frac{1}{2}\|\cdot\|^2 \right)(u) = \{f \in \mathbb{X}^\ast : \|f\|_{\mathbb{X}^\ast}^2 = \langle f, u \rangle = \|u\|^2\},
\]
which is single-valued if \( \mathbb{X}^\ast \) is strictly convex [2, Lemma 1.5.5]—in particular, it collapses to the identity in Hilbert spaces—and represents the regulariser in classical Lavrentiev regularisation.

Maximal monotonicity also holds for hemicontinuous monotone \( \mathcal{A} : \mathbb{X} \rightarrow \mathbb{X}^\ast \), where hemicontinuity

is a type of weak directional continuity in the sense that
\[
\langle A(u + \epsilon v), w \rangle \xrightarrow{\epsilon \to 0^+} \langle Au, w \rangle
\]
for all \(u, v, w \in X\). Every strong-to-weak continuous (demicontinuous) operator—meaning that \(u_k \to u\) in \(X\) implies \(Au_k \to Au\) weakly in \(X^*\)—is hemicontinuous, and in fact, for monotone \(X \to X^*\) both hemicontinuity, demicontinuity, and maximal monotonicity are equivalent. In the special case of linear \(A\), hemicontinuity is automatic, and paired with monotonicity one even obtains continuity.

Monotonicity is further preserved for sums \(A + B\), and strictly so provided at least one map is strictly monotone. For maximally monotone operators the sum rule is more delicate, but is at least true when one of the terms is a subgradient. In particular, this leads to maximal monotonicity of \(\tilde{A} := A + \alpha \partial R\) for all \(\alpha > 0\) in (7) under Assumption 2.1.

As for coercivity, we shall need a localised, quantitative form of Theorem 3.1. On the one hand, this allows for less restrictive assumptions on the forward operator \(A\), while on the other hand, we gain an a priori estimate of the size of the regularised solutions which is crucial in the proofs of stability and convergence of (7). Specifically, whereas the standard criterion
\[
\frac{\langle y, u \rangle}{\|u\|} \to \infty \quad \text{for all } (u, y) \in \tilde{A} \quad \text{as } \|u\| \to \infty
\]
is uniform over all data \(f \in X^*\) and centred at the origin, we shall instead assume the existence of a point \(u_{\text{coerc.}}\) and an \(r > 0\) (depending on \(f\)) such that
\[
\langle y - f, u - u_{\text{coerc.}} \rangle \geq 0 \quad \text{for all } (u, y) \in \tilde{A} \quad \text{with } \|u - u_{\text{coerc.}}\| \geq r. \tag{17}
\]
A priori we may then conclude that every solution \(u_\ast\) of \(\tilde{A}u \ni f\) satisfies \(\|u_\ast - u_{\text{coerc.}}\| \leq r\) (see [2, Theorem 1.7.9]), which quantitatively relates the solutions and the data.

In fact, Theorem 3.1 adapted to (17) can now be quickly proved as follows [2, along the lines of the proof of Theorem 1.7.5], where we set \(u_{\text{coerc.}} = 0\) for clarity. By Minty–Rockafellar’s theorem (see [31, Theorem 6] for a very short argument), stating that \(\tilde{A} + \lambda J\) is surjective for every \(\lambda > 0\), there exist \((u_\lambda, y_\lambda) \in \tilde{A}\) and \(j_\lambda \in Ju_\lambda\) satisfying \(y_\lambda + \lambda j_\lambda = f\). Then
\[
\langle y_\lambda - f, u_\lambda \rangle = -\lambda \langle j_\lambda, u_\lambda \rangle = -\lambda \|u_\lambda\|^2 \leq 0
\]
for every \(\lambda > 0\). If \(u_\lambda = 0\) for some \(\lambda\), then \(j_\lambda = 0\), and so \(y_\lambda = f\) and \(u_\ast = 0\) is a solution. Otherwise, we must have
\[
\langle y_\lambda - f, u_\lambda \rangle < 0
\]
for all \(\lambda\), from which (17) tells us that \(\|u_\lambda\| < r\) for all \(\lambda\). By reflexivity and up to a subsequence, \(\{u_\lambda\}_\lambda\) therefore converges weakly in \(X\) to some \(u_\ast\) as \(\lambda \to 0\), and since
\[
\langle f - \lambda j_\lambda - y, u_\lambda - u \rangle = \langle y_\lambda - y, u_\lambda - u \rangle \geq 0
\]
for all \((u, y) \in \tilde{A}\) by monotonicity, we may take limits as \(\lambda \to 0\) and deduce that \(\langle f - y, u_\ast - u \rangle \geq 0\), using that \(J\) is bounded. Now maximal monotonicity gives \((u_\ast, f) \in \tilde{A}\), and \(\|u_\ast\| \leq \liminf\|u_\lambda\| \leq r\).

This yields an alternative route to the following result, where we note that (18) is a slightly more restrictive but apparently easier variant of (17) using the definition of \(\partial R(u)\).
Theorem 3.2 (Browder [10]). Let $\mathcal{A} : X \to X^*$ be maximally monotone and $\mathcal{R} : X \to (-\infty, \infty]$ be proper, convex, and lower semicontinuous. Suppose that for a given $f \in X^*$, there exist $u_{\text{coerc.}} \in \text{dom } \mathcal{R}$ and $r > 0$ such that
\[
\langle \mathcal{A} u - f, u - u_{\text{coerc.}} \rangle + \mathcal{R}(u) - \mathcal{R}(u_{\text{coerc.}}) \geq 0
\]
for all $u \in X$ with $\|u - u_{\text{coerc.}}\| \geq r$. Then there exists $u_* \in \text{dom } \mathcal{R}$ satisfying $\|u_* - u_{\text{coerc.}}\| \leq r$ and $\mathcal{A} u_* + \partial \mathcal{R}(u_*) \ni f$.

4 Proof of well-posedness

We now establish existence, stability, and convergence for subgradient-based Lavrentiev regularisation (7) under Assumption 2.1 with help of Theorem 3.2. Without loss of generality we assume in this section that $\mathcal{R}(u^\dagger) = 0$. Recall also that $f^\dagger$ is the exact data for which $\mathcal{A} u^\dagger = f^\dagger$, that $f^\delta$ is its noisy counterpart, and that $f$ refers to any element of $X^*$.

Remark 4.1. In order to obtain convergence, it seems necessary to prove coercivity of $\mathcal{A} + \alpha \partial \mathcal{R}$ relative to the exact solution $u^\dagger$ for the monotonicity arguments to work. This is not an issue for the existence of regularised solutions for fixed $\alpha$, but we choose $u_{\text{coerc.}} = u^\dagger$ also in this case. Observe further that there is no direct relationship between $u_{\text{coerc.}}$ in Theorem 3.2 and $u_{\text{ref}}$ from Proposition 2.2; $u_{\text{coerc.}}$ is a point where the combination of $\mathcal{A}$ and $\partial \mathcal{R}$ is coercive, whereas $u_{\text{ref}}$ relates to $\mathcal{A}$ alone.

4.1 Existence—proof of Theorem 2.3. We apply Theorem 3.2, where it remains to establish (18). We choose in the following $u_{\text{coerc.}} = u^\dagger$. Moreover, given $u \in X$ we denote
\[
u_r := ru + (1 - r)u^\dagger = u^\dagger + r(u - u^\dagger).
\]
Then (18) is equivalent to the existence of an $r_0 > 0$ such that
\[
\langle \mathcal{A} \nu_r - f, r(u - u^\dagger) \rangle + \alpha \mathcal{R}(\nu_r) \geq 0
\]
for all $u \in X$ with $\|u - u^\dagger\| = t_0$ and $r > r_0$. For simplicity, we assume in the following that $t_0 = 1$.

Define now
\[
C := \frac{\|f^\dagger - f\|_{X^*}}{\alpha}
\]
and let $V := \{u \in X : \|u\| + \mathcal{R}(u) \leq C + \|u^\dagger\| + 1\}$. By Assumption 2.1 v), the set $V$ is compact, and thus the lower semicontinuity of $\mathcal{R}$ implies that $\mathcal{R}$ is bounded below on $V$, say, $\mathcal{R}(u) \geq -K$ for all $u \in V$. Next we define
\[
r_0 := \sup \left\{ r > 0 : \inf_{u \in U_C} \langle \mathcal{A} \nu_r, u - u^\dagger \rangle \leq \alpha K + \|f\|_{X^*} \right\},
\]
where $U_C := \{u \in X : \|u - u^\dagger\| = 1 \text{ and } \mathcal{R}(u) \leq C\}$. Such an $r_0$ exists in view of the growth condition in Assumption 2.1 vi).
Assume now that \( r > r_0 \) and that \( u \in \mathcal{X} \) satisfies \( \|u - u^\dagger\| = 1 \). By the decomposition
\[
Au_r - f = (Au_r - Au^\dagger) + (f^\dagger - f),
\]
it follows from monotonicity that
\[
\langle Au_r - f, u - u^\dagger \rangle \geq \langle Au_r - Au^\dagger, u - u^\dagger \rangle - \|f^\dagger - f\|_{\mathcal{X}^*} \|u - u^\dagger\| \\
\geq 0 - \|f^\dagger - f\|_{\mathcal{X}^*} = -C\alpha,
\]
and thus
\[
\langle Au_r - f, r(u - u^\dagger) \rangle + \alpha\mathcal{R}(u_r) \geq -C\alpha r + \alpha\mathcal{R}(u_r).
\]
Therefore, if \( \mathcal{R}(u_r) > Cr \), then (19) holds.

Let us now assume that \( \mathcal{R}(u_r) \leq Cr \). The convexity of \( \mathcal{R} \) and the assumption \( \mathcal{R}(u^\dagger) = 0 \) imply that
\[
r\mathcal{R}(u) = r\mathcal{R}\left(\frac{1}{r}u_r + \frac{r-1}{r}u^\dagger\right) \leq \mathcal{R}(u_r) + (r - 1)\mathcal{R}(u^\dagger) = \mathcal{R}(u_r) \leq Cr,
\]
that is, \( \mathcal{R}(u) \leq C \), which implies that \( u \in U_C \). Since \( r > r_0 \), it follows that
\[
\langle Au_r, u - u^\dagger \rangle > K + \|f\|_{\mathcal{X}^*}.
\]
Moreover, we have that
\[
\|u\| + \mathcal{R}(u) \leq \|u - u^\dagger\| + \|u^\dagger\| + \mathcal{R}(u) \leq 1 + \|u^\dagger\| + C,
\]
and thus \( u \in V \), which in turn implies that \( \mathcal{R}(u) \geq -K \). Therefore
\[
\langle Au_r - f, r(u - u^\dagger) \rangle + \alpha\mathcal{R}(u_r) > (\alpha K + \|f\|_{\mathcal{X}^*} + r\alpha\mathcal{R}(u)) \\
\geq \alpha K r + r\|f\|_{\mathcal{X}^*} - r\|f\|_{\mathcal{X}^*} - r\alpha K = 0.
\]
This shows that (19) holds for all \( u \in \mathcal{X} \) with \( \|u - u^\dagger\| = 1 \) (= \( t_0 \)) and \( \mathcal{R}(u) \leq C \), and all \( r > r_0 \), which concludes the proof.

As a consequence of the above arguments, we note the following preliminary stability result.

**Corollary 4.2.** Let \( \alpha > 0 \) be fixed. If \( \{f_k\} \) is a bounded sequence of data in \( \mathcal{X}^* \), then the corresponding sequence of regularised solutions \( \{u_{a,k}\} \) solving \( Au + \alpha\partial\mathcal{R}(u) \equiv f_k \) is also bounded.

**Proof.** Since \( \{f_k\} \) is bounded, it follows that
\[
\tilde{C} := \sup_k \frac{\|f^\dagger - f_k\|_{\mathcal{X}^*}}{\alpha} < \infty.
\]
Replacing the constant \( C \) in (20) by \( \tilde{C} \) and correspondingly modifying the values of \( K \) and \( r_0 \), we obtain by following the proof of Theorem 2.3 that
\[
\langle Au_r - f_k, r(u - u^\dagger) \rangle + \alpha\mathcal{R}(u_r) > 0
\]
for all \( u \in \mathcal{X} \) with \( \|u - u^\dagger\| = 1 \) (= \( t_0 \)), all \( r > r_0 \), and all \( k \in \mathbb{N} \). As a consequence of Theorem 3.2, we therefore obtain the estimate \( \|u_{a,k} - u^\dagger\| \leq r_0 \) for all \( k \), which proves the assertion. \( \square \)
4.2 Stability—proof of Theorem 2.4. Let $\alpha > 0$ be fixed and $f \in \mathbb{X}^*$ be given, and assume that $\{f_k\} \subseteq \mathbb{X}^*$ is a sequence of data converging to $f$. Furthermore, let $u_\alpha$ and $u_{\alpha,k}$ be the respective regularised solutions of

$$A u_\alpha + \alpha \partial R(u_\alpha) \ni f$$

and

$$A u_{\alpha,k} + \alpha \partial R(u_{\alpha,k}) \ni f_k,$$

so that

$$A u_\alpha + \alpha \xi_\alpha = f$$

and

$$A u_{\alpha,k} + \alpha \xi_{\alpha,k} = f_k$$

(21)

for some $\xi_\alpha \in \partial R(u_\alpha)$ and $\xi_{\alpha,k} \in \partial R(u_{\alpha,k})$. If we now test with $u_{\alpha,k} - u_\alpha$ in both of the equations in (21) and subtract the resulting expressions from each other, we find by monotonicity of $\partial R$ and $A$ the basic stability estimate

$$0 \leq \langle \xi_{\alpha,k} - \xi_\alpha, u_{\alpha,k} - u_\alpha \rangle$$

$$\leq \langle \xi_{\alpha,k} - \xi_\alpha + \frac{1}{\alpha}(A u_{\alpha,k} - A u_\alpha), u_{\alpha,k} - u_\alpha \rangle$$

$$= \frac{1}{\alpha} \langle f_k - f, u_{\alpha,k} - u_\alpha \rangle$$

$$\leq \frac{1}{\alpha} \|f_k - f\|_{\mathbb{X}} \|u_{\alpha,k} - u_\alpha\|.$$  

(22)

Since $\|u_{\alpha,k} - u_\alpha\|$, as a $k$-indexed sequence, is bounded by Corollary 4.2, it then follows that

$$\lim_{k \to +\infty} \langle \xi_{\alpha,k} - \xi_\alpha, u_{\alpha,k} - u_\alpha \rangle = 0.$$  

(23)

Moreover, by reflexivity of $\mathbb{X}$ and boundedness, $\{u_{\alpha,k}\}$ converges weakly—up to a subsequence—to some $u \in \mathbb{X}$, and from (23) we infer that

$$\limsup_k R(u_{\alpha,k}) \leq R(u_\alpha) + \lim \langle \xi_{\alpha,k}, u_{\alpha,k} - u_\alpha \rangle$$

$$= R(u_\alpha) + \langle \xi_\alpha, u - u_\alpha \rangle$$

$$< \infty.$$  

(24)

Assumption 2.1 v) now implies that $u_{\alpha,k} \to u$ strongly, again up to a subsequence, and so demicontinuity results in $A u_{\alpha,k} \rightharpoonup A u$ weakly in $\mathbb{X}^*$. In combination with the convergence $f_k \to f$ and (23), which by virtue of the definitions of $\xi_{\alpha,k}$ and $\xi_\alpha$ can be rewritten as

$$\lim_k \langle A u_\alpha - A u_{\alpha,k} + f - f_k, u_\alpha - u_{\alpha,k} \rangle = 0,$$

we therefore deduce that

$$\langle A u_\alpha - A u_\alpha, u_\alpha - u \rangle = 0.$$  

Strict monotonicity now forces $u = u_\alpha$, and standard subsequence-subsequence reasoning yields that the full sequence $\{u_{\alpha,k}\}$ converges to $u_\alpha$. As a consequence, (24) becomes

$$\limsup_k R(u_{\alpha,k}) \leq R(u_\alpha),$$

so that $R(u_{\alpha,k}) \to R(u_\alpha)$ by lower semicontinuity. Finally, demicontinuity gives $A u_{\alpha,k} \rightharpoonup A u_\alpha$ weakly, and this concludes the proof of Theorem 2.4.
4.3 Convergence—proof of Theorem 2.5. Note first that, similarly to the basic stability estimate (22), we have
\[ \mathcal{R}(u^\delta_a) - \mathcal{R}(u^\dagger) \leq \langle e^\delta_a, u_a - u^\dagger \rangle \]
\[ = \langle e^\delta_a + \frac{1}{\alpha} (A u^\delta_a - Au^\dagger), u^\delta_a - u^\dagger \rangle - \frac{1}{\alpha} \langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle \]
\[ = \frac{1}{\alpha} \langle f^\delta - f^\dagger, u^\delta_a - u^\dagger \rangle - \frac{1}{\alpha} \langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle \]
\[ \leq \frac{\delta}{\alpha} \|u^\delta_a - u^\dagger\| - \frac{1}{\alpha} \langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle. \]

(25)

If \( \{u^\delta_a\}_\delta \) is bounded in \( \mathcal{X} \), where \( \alpha = \alpha(\delta) \) is as in (12), then it converges weakly—up to a subsequence—to some \( u \in \mathcal{X} \) by reflexivity. From (25), monotonicity of \( \mathcal{A} \), and the hypothesis we furthermore have
\[ \limsup_{\delta \to 0} \mathcal{R}(u^\delta_a) \leq \mathcal{R}(u^\dagger) + \limsup_{\delta \to 0} \frac{\delta}{\alpha(\delta)} \|u^\delta_a - u^\dagger\| \leq C < \infty \]

(26)

for some constant \( C \) depending only on \( \sup_\delta (\|u^\delta_a\|) \). By compactness in Assumption 2.1 v), it then follows that \( u^\delta_a \to u \) as \( \delta \to 0 \), and consequently \( Au^\delta_a \to Au \) weakly in \( \mathcal{X}^* \) by demicontinuity. Now we may write (25) as
\[ 0 \leq \langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle \leq \delta \|u^\delta_a - u^\dagger\| - \alpha \left( \mathcal{R}(u^\delta_a) - \mathcal{R}(u^\dagger) \right) \]

and take limits to obtain
\[ \langle Au - Au^\dagger, u - u^\dagger \rangle = 0. \]

Strict monotonicity gives \( u = u^\dagger \), and uniqueness of \( u^\dagger \) also implies that the full sequence \( \{u^\delta_a\}_\delta \) converges to \( u^\dagger \). Thus (26) in fact implies that
\[ \limsup_{\delta \to 0} \mathcal{R}(u^\delta_a) \leq \mathcal{R}(u^\dagger), \]

which, in view of the lower semicontinuity of \( \mathcal{R} \) implies that \( \mathcal{R}(u^\delta_a) \to \mathcal{R}(u^\dagger) \).

Accordingly, it remains to establish boundedness of \( \{u^\delta_a\}_\delta \). By assumption, \( \delta/\alpha(\delta) \) is bounded as \( \delta \to 0 \), and thus there exists \( \tilde{C} > 0 \) such that \( \delta/\alpha(\delta) \leq \tilde{C} \) for all sufficiently small \( \delta \). As such, we can again replace the constant \( C \) in (20) by \( \tilde{C} \) and correspondingly modify the values of \( K \) and \( r_0 \). Then we obtain by following the proof of Theorem 2.3 and applying Theorem 3.2 the estimate \( \|u^\delta_a - u^\dagger\| \leq r_0 \) for all sufficiently small \( \delta \), which establishes the boundedness of \( \{u^\delta_a\}_\delta \) and concludes the proof of Theorem 2.5.

5 Proofs of convergence rates

In the following we will establish Theorems 2.7 and 2.8, where we restate the general variational source condition
\[ D(u, u^\dagger) \leq \mathcal{R}(u) - \mathcal{R}(u^\dagger) + \varphi \left( \langle Au - Au^\dagger, u - u^\dagger \rangle \right) \]

(27)

from (13) for convenience. We write \( Au^\dagger \) instead of \( f^\dagger \) in order to highlight the monotone structure. Here, \( D \) is any nonnegative function on \( \mathcal{X} \times \mathcal{X} \) acting as a distance measure, and \( \varphi : [0, \infty) \to [0, \infty) \)
denotes an index function, meaning that $\varphi$ is concave, strictly increasing, and continuous with $\varphi(0) = 0$. We also assume a sublinear rate of decay of $\varphi$ at the origin—that is, $\lim_{t\to0} \varphi(t)/t = \infty$.

In addition, let $\psi : [0, \infty) \to [0, \infty)$ be the convex (Fenchel) conjugate of $\varphi^{-1}$ defined by

$$\psi(s) = \sup_{t \geq 0} (st - \varphi^{-1}(t)).$$

By construction, this mapping is convex and, due to the nonnegativity of $\varphi$ plus the fact that $\varphi(0) = 0$, also nonnegative with $\psi(0) = 0$. This further implies that $\psi$ is continuous and increasing on $[0, \infty)$. Moreover, since $\varphi$ decays sublinearly at 0, the function $\psi$ is strictly increasing near 0—and therefore also on its whole domain by convexity.

**Proof of Theorem 2.7.** With help of the basic stability estimate

$$\mathcal{R}(u^\delta_a) - \mathcal{R}(u^\dagger) \leq \frac{\delta}{\alpha} \|u^\delta_a - u^\dagger\| - \frac{1}{\alpha} \langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle$$

from (25), we find that the general variational inequality (27) yields

$$D(u^\delta_a, u^\dagger) \leq \mathcal{R}(u^\delta_a) - \mathcal{R}(u^\dagger) + \varphi\left(\langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle\right)$$

$$\leq \frac{\delta}{\alpha} \|u^\delta_a - u^\dagger\| - \frac{1}{\alpha} \langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle + \varphi\left(\langle Au^\delta_a - Au^\dagger, u^\delta_a - u^\dagger \rangle\right)$$

$$\leq \frac{\delta}{\alpha} \|u^\delta_a - u^\dagger\| + \sup_{t \geq 0} (\varphi(t) - \frac{t}{\alpha}).$$

Since by definition of $\psi$ we have

$$\sup_{t \geq 0} \left(\varphi(t) - \frac{t}{\alpha}\right) = \frac{1}{\alpha} \sup_{t \geq 0} (\alpha \varphi(t) - t) = \frac{1}{\alpha} \sup_{s \geq 0} (\alpha s - \varphi^{-1}(s)) = \frac{\psi(\alpha)}{\alpha},$$

it follows that

$$D(u^\delta_a, u^\dagger) \leq \frac{\delta}{\alpha} \|u^\delta_a - u^\dagger\| + \frac{\psi(\alpha)}{\alpha}. \quad (29)$$

Now recall that the regularisation method (7) is convergent by Theorem 2.5—in fact, weak convergence together with reflexivity of $X$ suffice here. Therefore, for sufficiently a small noise level $\delta$ and an appropriate parameter choice for $\alpha$, we know that $u^\delta_a - u^\dagger$ is bounded in $X$ and $u^\delta_a \in \mathcal{M}$. Hence, there exists $C > 0$ such that $\|u^\delta_a - u^\dagger\| \leq C$, and we obtain the estimate claimed in the statement of the theorem. At last, choosing $\alpha \sim \psi^{-1}(\delta)$ in (29) gives the convergence rate

$$D(u^\delta_a, u^\dagger) \lesssim \frac{\delta}{\psi^{-1}(\delta)}. \quad \square$$

**Corollary 5.1.** Assume that a variational source condition of Hölder type

$$D(u, u^\dagger) \leq \mathcal{R}(u) - \mathcal{R}(u^\dagger) + c \left(\langle Au - Au^\dagger, u - u^\dagger \rangle\right)^{1/p}$$

holds for some $p > 1$ and $c > 0$. Then as $\delta \to 0$ we obtain the convergence rate

$$D(u^\delta_a, u^\dagger) \lesssim \delta^{1/p} \quad \text{if } \alpha \sim \delta^{1/q},$$

where $q := p/(p - 1) > 1$ is the conjugate Hölder exponent of $p$. 


\textbf{Proof.} This is immediate from Theorem 2.7, noting that for $\varphi \sim \delta^{1/p}$ we have $\psi \sim \delta^q$, which leads to $\psi^{-1} \sim \delta^{1/q}$ and $\delta/\psi^{-1}(\delta) \sim \delta^{1/p}$.

Finally we specialise Corollary 5.1 to norm-like distance measures of the form $D(u, u^+) \sim \|u - u^+\|_r^r$ for some $r > 1$. This results in Theorem 2.8 by the following arguments.

\textbf{Proof of Theorem 2.8.} We proceed along the proof of Theorem 2.7 until estimate (29), which in this case reads

$$c_1 \|u_a^\delta - u^+\|_r \leq \frac{\delta}{\alpha} \|u_a^\delta - u^+\| + c_3 \alpha^{q-1}$$

for some constant $c_3 > 0$ and with conjugate Hölder exponent $q := p/(p - 1)$. Next we apply Young’s inequality $ab \leq \frac{1}{q} a^q + \frac{1}{p} b^p$ with conjugate exponent $s := r/(r - 1)$ to $a := \|u_a^\delta - u^+\|$ and $b := 2\delta/(rc_1\alpha)$ and obtain that

$$c_1 \|u_a^\delta - u^+\|_r \leq \frac{rc_1}{2} \frac{2\delta \|u_a^\delta - u^+\|}{rc_1\alpha} + c_3 \alpha^{q-1}$$

$$\leq \frac{c_1}{2} \|u_a^\delta - u^+\|_r + \frac{rc_1}{2s} \left( \frac{2\delta}{rc_1\alpha} \right)^s + c_3 \alpha^{q-1}.$$ 

This simplifies to

$$\|u_a^\delta - u^+\|_r \lesssim \frac{\delta}{\alpha} + \alpha^{q-1},$$

and we may now choose

$$\alpha \sim \delta^\left( \frac{s}{s + q - 1} \right) = \delta^\left( \frac{r(p - 1)}{rp - 1} \right),$$

remembering the definitions of $q$ and $s$—in particular, that $(p - 1)(q - 1) = 1$—to arrive at the quantitative error estimate

$$\|u_a^\delta - u^+\| \lesssim \delta^\left( \frac{1}{rp - 1} \right).$$

\textbf{Remark 5.2.} Suppose that $A$ is positively homogeneous with $A(0) = 0$—for instance, that $A$ is linear—and assume that $u^+ \neq 0$. In particular, $\langle Au^+, u^+ \rangle > 0$ by strict monotonicity. If we test with $u = (1 - \varepsilon)u^+$ for $\varepsilon \in (0, 1)$ in the general variational source condition (13), then convexity of $\mathcal{R}$ yields that

$$D((1 - \varepsilon)u^+, u^+) \leq \varepsilon \left( \mathcal{R}(0) - \mathcal{R}(u^+) \right) + \varphi\left( \varepsilon^2 \langle Au^+, u^+ \rangle \right).$$

By requiring $\mathcal{R}(u^+) > \mathcal{R}(0)$ and finiteness of

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} D((1 - \varepsilon)u^+, u^+),$$

which is automatic if $D(u, v) \sim \|u - v\|_r^r$ for some $r \geq 1$, we may now divide by $\sqrt{\varepsilon} := \varepsilon \sqrt{\langle Au^+, u^+ \rangle}$ in (30) and let $\varepsilon \to 0^+$. This gives

$$\varphi(t) \gtrsim \sqrt{t}$$

uniformly near 0, and hence, we infer that the variational source condition (13) is realistic only for $\varphi(t)$ decaying no faster than $\sqrt{t}$ as $t \to 0^+$. In particular, this corresponds to $p \geq 2$ in Theorem 2.8 and Corollary 5.1 and agrees with the discussion in [21, Section 3.2].
We now discuss an application of subgradient-based Lavrentiev regularisation (7) for TV denoising in a class of Volterra integral operators (3) of the first kind, where we focus on the linear, convolutional case

\[ A(u)(t) := \int_0^t k(t - \tau) u(\tau)\,d\tau, \]

with \(A\) acting from the reflexive space \(X := L^p(0, T)\) to \(X^* = L^q(0, T)\) for suitable kernels \(k\) and conjugate exponents \(p, q \in (1, \infty)\). Remember that \(A\) retains causality with respect to time \(t\), that is, \(A\) depends only on \(u(\tau)\) for \(\tau \leq t\). Note nevertheless that the theory may also be applied to more global, noncausal problems, involving, for instance, Fredholm integral operators. One may also extend the results to operators \(u \mapsto Au + a(\cdot, u(\cdot))\) with pointwise terms added, subject to suitable growth conditions.

In the context of TV denoising, we next define the regularisation functional \(R: X \to [0, \infty]\)

\[
R(u) := \begin{cases} 
|u|_{TV} & \text{if } u \in BV(0, T); \\
+\infty & \text{else},
\end{cases}
\]

where \(BV(0, T)\) is the space of integrable functions with bounded variation and seminorm

\[
|u|_{TV} := |Du|(0, T) := \sup \left\{ -\int_0^T \varphi' u\,dt : \varphi \in C^1_c(0, T) \text{ with } \|\varphi\|_{L^\infty(0, T)} \leq 1 \right\}.
\]

Then \(R\) satisfies Assumption 2.1 iii), and since \(BV(0, T)\) is compactly embedded in \(X\) for every \(p \in (1, \infty)\) (see [3, Proposition 3.13]), it follows that the sublevel sets in Assumption 2.1 v) are compact as well. Note also that the growth condition (8) holds automatically because \(A\) is linear.

Accordingly, it remains to establish strict monotonicity of \(A\) on \(X\), which we are able to prove for a class of convolution kernels and exponents \(p\).

### 6.1 Strictly monotone Volterra operators

If \(k\) is identically 1, the inverse problem for \(Au = f^\dagger\) becomes that of numerical differentiation. In this case, the monotonicity of \(A\) on \(L^2(0, T)\), say, follows directly from

\[
\langle Au, u \rangle = \int_0^T A(u)(t) u(t)\,dt = \frac{1}{2} \left( \int_0^T u(t)\,dt \right)^2
\]

by integrating by parts, which implies that \(A\) is strictly monotone on the quotient Hilbert space

\[
L^2(0, T)/\{ u \in L^2(0, T) : \int_0^T u\,d\tau = 0 \}.
\]

More generally, we have the fractional counterpart with singular convolution kernel

\[
k(x) = \frac{x^{s-1}}{\Gamma(s)}
\]

for \(s > 0\), where \(\Gamma\) is the Gamma function, in which (31) is known as an Abel or Riemann–Liouville integral. The latter kernel and also the negative exponential \(x \mapsto \exp(-cx)\) both generate strictly monotone operators satisfying Assumption 2.1 ii), as seen in the following result.
Proposition 6.1. Let $p \in (1, 2]$ and $q$ be its conjugate exponent. Assume that the kernel $k$ is strictly convex and decreasing on $(0, T]$ and lies in $L^{q/2}(0, T)$, with strictly positive mean, that is, $\int_0^T k(t) \, dt > 0$. Then the convolutional Volterra operator (31) defines a strictly monotone, bounded map

$$A : L^p(0, T) \to L^q(0, T).$$

Remark 6.2. Note that we both allow singularities in $k$ at the origin and some negative values as long as the mean of $k$ is strictly positive. The proof below is a simple adaption of [5, Lemma 1.2] and is, in some sense, related to Bochner’s result characterising the Fourier transform of positive-definite functions. We refer to [28] and [36] for further results in this direction, the latter of which includes monotonicity of Volterra operators $\int_0^t u(\tau) \, d\mu(\tau)$ with respect to a class of positive-definite Radon measures $\mu$. It is, however, not entirely clear how these results relate to strict monotonicity.

Proof. Hölder’s inequality yields that

$$|Au(t)| \leq \|k\|_{L^{q/2}(0, T)}^{1/2} \left( \int_0^t |k(t - \tau)|^{p/2} |u(\tau)|^p \, d\tau \right)^{1/p},$$

and so the general Minkowski integral inequality leads to

$$\|Au\|_{L^p(0, T)} \leq \|k\|_{L^{q/2}(0, T)}^{1/2} \left( \int_0^T \left( \int_0^t |k(t - \tau)|^{p/2} |u(\tau)|^p \, d\tau \right)^{q/p} \right)^{1/q} dt$$

$$\leq \|k\|_{L^{q/2}(0, T)}^{1/2} \left( \int_0^T \left( \int_\tau^T |k(t - \tau)|^{q/2} |u(\tau)|^q \, dt \right)^{p/q} \right)^{1/p} dt$$

$$\leq \|k\|_{L^{q/2}(0, T)} \|u\|_{L^p(0, T)},$$

which establishes boundedness of $A : L^p(0, T) \to L^q(0, T)$.

As regards strict monotonicity, we first approximate $k$ pointwise by the shifted, bounded functions

$$k_\varepsilon(t) := \begin{cases} k(t + \varepsilon) & \text{for } t \in [0, T - \varepsilon]; \\ k(T) & \text{for } t \in (T - \varepsilon, T] \end{cases}$$

in order to avoid the potential singularity at 0. Then after extending $k_\varepsilon$ evenly to $[-T, 0]$, Dirichlet’s theorem gives that $k_\varepsilon$ (which is integrable, continuous, and of bounded variation) equals its Fourier series

$$k_\varepsilon(t) = \frac{c_{0,\varepsilon}}{2} + \sum_{n=1}^\infty c_{n,\varepsilon} \cos(\tilde{n} t)$$

pointwise on $(-T, T)$, where $c_{n,\varepsilon} := \frac{2}{T} \int_0^T k_\varepsilon(t) \cos(\tilde{n} t) \, dt$ and $\tilde{n} := 2\pi n / T$. Since $k$ is strictly convex and decreasing on $(0, T]$, it follows from [5, Lemma 1.1] that

$$c_n := \lim_{\varepsilon \to 0^+} c_{n,\varepsilon} = \frac{2}{T} \int_0^T k(t) \cos(\tilde{n} t) \, dt > 0$$

for all $n \geq 1$, and $c_0 := \lim_{\varepsilon \to 0^+} c_{0,\varepsilon} = \frac{2}{T} \int_0^T k(t) \, dt > 0$ by assumption.
We next use dominated convergence repeatedly to find that
\[
\langle Au, u \rangle = \lim_{\epsilon \to 0^+} \int_0^T \int_0^t k_\epsilon(t - \tau) u(\tau) d\tau u(t) dt
\]
\[
= \int_0^T \left( \frac{c_0}{2} \int_0^t u(\tau) d\tau + \sum_{n=1}^{\infty} c_n \int_0^t u(\tau) \cos(n(t - \tau)) d\tau \right) u(t) dt.
\]
Observe now that
\[
\int_0^T \int_0^t u(\tau) \cos(n(t - \tau)) d\tau u(t) dt = \int_0^T \int_0^t v(\tau) d\tau v(t) dt + \int_0^T \int_0^t w(\tau) d\tau w(t) dt
\]
with \( v(t) := u(t) \cos(n t) \) and \( w(t) := u(t) \sin(n t) \). As such, integration by parts gives that
\[
\int_0^T \int_0^t u(\tau) \cos(n(t - \tau)) d\tau u(t) dt = \frac{1}{2} \left( \int_0^T v(t) dt \right)^2 + \frac{1}{2} \left( \int_0^T w(t) dt \right)^2
\]
\[
= \frac{1}{2} \left| \int_0^T u(t) e^{in \tau} dt \right|^2,
\]
and so
\[
\langle Au, u \rangle = \frac{c_0}{2} \left( \int_0^T u(t) dt \right)^2 + \sum_{n=1}^{\infty} \frac{c_n}{2} \left| \int_0^T u(t) e^{in \tau} dt \right|^2 = \sum_{n=0}^{\infty} \frac{c_n}{2} \left| \int_0^T u(t) e^{in \tau} dt \right|^2.
\]
Since \( c_n > 0 \) for all \( n \geq 0 \), we conclude that \( A \) is strictly monotone. \( \square \)

6.2 Description of the subgradient of TV and numerical examples. In order to make the regularisation method (7) practical numerically, let us first characterise the subgradient \( \partial R : X \to X^* \) of (32). Denote by \( (Du)^+ \) and \( (Du)^- \) the positive and negative parts in the Jordan measure decomposition
\[
Du = (Du)^+ - (Du)^-,
\]
and let \( \text{supp} \mu \) symbolise the support of a measure \( \mu \). We then obtain the following description of \( \partial R \), which is a straightforward adaption of [23, Proposition 2.1]. Note that \( W^{1,q}(0, T) \) means the Sobolev space of functions in \( L^q(0, T) \) whose weak derivative also lie in \( L^q(0, T) \).

**Proposition 6.3.** Let \( u \in \text{dom } R = BV(0, T) \subset L^p(0, T) \) for \( R \) in (32) and \( p, q \in (1, \infty) \) be conjugate exponents. Then \( \xi \in \partial R(u) \) if and only if \( \xi = -\varphi' \) for \( \varphi \in W^{1,q}(0, T) \) vanishing at its endpoints (thus \( \xi \) has zero mean) and satisfying
\[
\| \varphi \|_{L^\infty(0, T)} \leq 1, \quad \text{supp}(Du)^+ \subseteq \{ \varphi = 1 \}, \quad \text{and} \quad \text{supp}(Du)^- \subseteq \{ \varphi = -1 \}.
\]
Moreover, \( R(u) = (\xi, u) \).

**Proof.** As in [23, Proposition 2.1], we exploit from the Fenchel–Young inequality that
\[
\xi \in \partial R(u) \iff \langle \xi, u \rangle = R(u) + R^*(\xi),
\]
where
\[ R^*(\xi) := \sup_{v \in L^p(0,T)} \langle \xi, v - R(v) \rangle \]
denotes the convex conjugate of \( R \) similarly as (28). Since \( R(u) < \infty \) by assumption, it remains to characterise \( \text{dom} R^* \).

It suffices to let the supremum run over \( v \in BV(0,T) \) and we need only consider \( \xi \in L^q(0,T) \) with zero mean—if not, we get \( R^*(\xi) = \infty \) by letting \( v \) be constant. It is also clear that \( \xi \) corresponds uniquely to a function \( \varphi \in W^{1,q}(0,T) \) vanishing at its endpoints and given by
\[ \varphi(t) := -\int_0^t \xi(\tau) \, d\tau. \]

By integrating by parts, we obtain that
\[ \int_0^T \varphi \, d(Dv) + \int_0^T \varphi' \, dt = \varphi(T) v(T^+) - \varphi(0) v(0^+) = 0, \]
and so
\[ \langle \xi, v \rangle - |v|_{TV} = -\int_0^T \varphi' \, dt - \int_0^T d\|Dv\| \]
\[ = \int_0^T (\varphi - 1) \, d(Dv)^+ - \int_0^T (\varphi + 1) \, d(Dv)^-. \tag{34} \]

If \( \varphi \geq -1 \) and \( \{ \varphi > 1 \} \) has positive measure (or \( \varphi \leq 1 \) and \( \{ \varphi < -1 \} \) has positive measure), then one finds that \( R^*(\xi) = \infty \) by choosing, say, a sequence of monotonically increasing (decreasing) functions \( v \in BV(0,T) \). As such, we must have \( \|\varphi\|_{L^{\infty}(0,T)} \leq 1 \). But then (34) is nonpositive—and equals 0, corresponding to the supremum over \( v \), precisely when
\[ (Dv)^{+}(\{ \varphi < 1 \}) = 0 \quad \text{and} \quad (Dv)^{-}(\{ \varphi > -1 \}) = 0. \]

In conclusion, \( R^*(\xi) = 0 \) for \( \xi \in \text{dom} R^* \), so that \( R(u) = \langle \xi, u \rangle \) and the supremum in the definition of \( R^* \) is achieved at \( v = u \). \( \square \)

By writing the regularised problem (7) as \( A u - f^\delta \in -\alpha \partial R(u) \), with \( u := u^\delta \) for brevity, and introducing
\[ \mathcal{L}u(t) := \int_0^t (A u(\tau) - f^\delta(\tau)) \, d\tau, \]
the characterisation of \( \partial R \) in Proposition 6.3 now leads after integration to the following relations for \( |Du| \)-almost every \( t \in (0,T) \):
\[
\begin{align*}
|\mathcal{L}u(t)| &\leq \alpha; \quad \text{(35)} \\
\mathcal{L}u(t) &= +\alpha \quad \text{when } t \in \text{supp}(Du)^+; \quad \text{(36)} \\
\mathcal{L}u(t) &= -\alpha \quad \text{when } t \in \text{supp}(Du)^-; \quad \text{(37)} \\
\mathcal{L}u(T) &= 0. \quad \text{(38)}
\end{align*}
\]
We see from (35) that the integrated image \( \int_0^t A u \, d\tau \) of the approximate solution lies within an \( \alpha \)-tube of the integrated data \( \int_0^t f^\delta \, d\tau \). Moreover, touching the upper part of the tube (36) may
lead to an increase in $u$, whereas $u$ has a chance to decrease whenever one touches the lower part of the tube (37). Then it is possible to compute a piecewise-constant solution in $BV(0, T)$ of the system (35)–(38) by means of a new version of the classical taut-string algorithm [13, 14, 17], which we will discuss in detail in a forthcoming paper. The approach is very fast, efficient and employs dynamic programming. In addition to $A$ being strictly monotone in the functional-analytic sense, we also require in this context that $A$ has a positive kernel $k$—so that $A$ preserves the pointwise order. That is, if $u, v \in BV(0, T)$ satisfy $u(t) \geq v(t)$ for almost every $t \in (0, T)$, then also $A u(t) \geq A v(t)$ for almost every $t$.

As illustrations for TV denoising with (7) in convolutional Volterra equations (31), we first consider the Abel operator with kernel (33) for $s = 1/3$. Figure 1 displays the typical performance of the approximate solutions $u_\delta^\alpha$ with moderate noise in the data, and we observe quite faithful reconstructions of the original jump locations and heights.

![Figure 1: TV denoising for the convolutional Abel operator (31)–(33) with $s = 1/3$.](image)

In the second example, we employ the exponential kernel $x \mapsto \exp(-x/10)$ and consider substantially noisy data, as seen in Figure 2. Again we find the performance to be pretty good, although there are some deviations from the true jump heights for jumps occurring close to another one.

![Figure 2: TV denoising for the Volterra operator (31) with exponential kernel $x \mapsto \exp(-t/10)$.](image)
A parameter-identification problem for a semilinear parabolic PDE

7.1 Setup. We now provide an application of subgradient-based Lavrentiev regularisation (7) to a class of parameter-identification problems for semilinear parabolic PDEs. To that end, let \( \Omega \subset \mathbb{R}^d \) be a Lipschitz domain, \( I := (0, 1) \) denote the unit interval, \( g \in L^\infty(\Omega) \) be some given initial data, and \( \varphi(y) := \text{sgn}(y)|y|^\lambda \) for some \( \lambda > 0 \)

be a strictly monotone nonlinearity. With \( \mathbb{X} \cong \mathbb{X}^* \) denoting the Hilbert space \( L^2(I \times \Omega) \), we then define \( \mathcal{A} : \mathbb{X} \to \mathbb{X} \) to be the operator that maps \( u \in \mathbb{X} \) to the weak solution of the PDE

\[
\begin{aligned}
\frac{\partial}{\partial t} y - \Delta y + \varphi(y) &= u \quad \text{in } I \times \Omega; \\
\partial_n y &= 0 \quad \text{on } I \times \partial \Omega; \\
y(0, \cdot) &= g \quad \text{on } \Omega.
\end{aligned}
\]

(40)

Note that this equation can alternatively be written as the gradient flow

\[
y_t + \partial G(y) \ni u, \quad y(0) = g,
\]

(41)

where \( G : L^2(\Omega) \to (-\infty, \infty] \) is the convex and lower-semicontinuous functional

\[
G(y) := \begin{cases} 
\int_{\Omega} \left( \frac{1}{2} |\nabla y|^2 + \frac{1}{\lambda+1} |y|^{\lambda+1} \right) \, dx & \text{if } y \in H^1(\Omega) \cap L^{\lambda+1}(\Omega); \\
+\infty & \text{else}.
\end{cases}
\]

Consequently, the standard theory on gradient flows in Hilbert spaces (see [9] or [6, Theorems 4.2, 4.5 and 4.11]) implies that the PDE (40) has a unique (mild/strong) solution in \( \mathbb{X} \) that depends continuously on \( u \) with respect to strong convergence in \( \mathbb{X} \). Put differently, the operator \( \mathcal{A} \) is well-defined and continuous. In addition, analogous to the standard proof of uniqueness for (41), one finds the comparison principle

\[
y \leq z \quad \text{a.e. in } I \times \Omega \quad \text{whenever} \quad u \leq v \text{ and } g \leq h \quad \text{a.e.},
\]

(42)

where \( y = \mathcal{A}u \) and \( z = \mathcal{A}v \) solve (40) with initial data \( g \) and \( h \), respectively. Moreover, in the case \( g = h \), one easily obtains the equality

\[
\langle y - z, u - z \rangle = \frac{1}{2} \int_{\Omega} (y(1, \cdot) - z(1, \cdot))^2 \, dx + \int_I \int_{\Omega} \left( |\nabla(y - z)|^2 + (\varphi(y) - \varphi(z))(y - z) \right) \, dx \, dt,
\]

which in view of the strict monotonicity of \( \varphi \) shows that \( \mathcal{A} \) is strictly monotone.

We now consider the parameter-identification problem of solving \( \mathcal{A}(\bar{u}) = y^\delta \) given possibly noisy data \( y^\delta \in \mathbb{X} \). For the stable solution of this problem, we assume that the true solution associated with the noise-free data \( y^\dagger \) has the form

\[
\bar{u}^\dagger = u^\dagger + f,
\]

where \( f \in L^\infty(I \times \Omega) \) is a known function and \( u^\dagger \in \text{BV}(I \times \Omega) \) is an unknown function that is constant in the spatial variable—that is, there exists \( v^\dagger \in \text{BV}(I) \) such that

\[
u^\dagger(t, x) = v^\dagger(t) \quad \text{for a.e. } (t, x) \in I \times \Omega.
\]
In other words, \( u^\dagger = \mathcal{I}(v^\dagger) \), where 
\[
\mathcal{I} : L^2(I) \to X
\]
is the natural embedding. With \( \mathcal{R} : X \to [0, \infty] \) defined by 
\[
\mathcal{R}(u) := \begin{cases} 
|u|_{TV(I \times \Omega)} & \text{if } u \in BV(I \times \Omega) \cap \text{ran } \mathcal{I}; \\
+\infty & \text{else,}
\end{cases}
\]
we see that subgradient-based Lavrentiev regularisation (7) adapted to this context requires finding the solution \( u^\dagger_\alpha \) of the inclusion 
\[
\mathcal{A}(u^\delta_\alpha + f) + \alpha \partial \mathcal{R}(u^\delta_\alpha) \ni y^\delta.
\]
(43)

In order to show that this is a well-posed regularisation method, we need to verify that Assumption 2.1 is satisfied. The requirements on the space \( X \) and the regulariser \( \mathcal{R} \) are directly satisfied, and, as discussed above, the operator \( \mathcal{A} \) is both continuous and strictly monotone on \( X \). Accordingly, it remains to establish the coercivity and the growth conditions in Assumption 2.1 v) and vi).

### 7.2 Verification of the conditions.

We assume for simplicity that \(|\Omega| = 1\). Then \( \|iv\|_X = \|v\|_{L^2(I)} \) and \( \mathcal{R}(iv) = |v|_{TV(I)} \) for all \( v \in L^2(I) \). By compactness of \( BV(I) \) in \( L^2(I) \), it is then evident that the sublevel set 
\[
\widehat{W} := \{ v \in L^2(I) : \|v\|_{L^2(I)} + |v|_{TV(I)} \leq C \}
\]
is compact for every \( C > 0 \), and thus the same is true for 
\[
W := \mathcal{I}\widehat{W} = \{ u \in X : \|u\|_X + \mathcal{R}(u) \leq C \},
\]
which shows that Assumption 2.1 v) holds.

Now let \( C > 0 \) and define 
\[
U_C := \{ u \in X : \|u - u^\dagger\|_X = 1 = t_0 \} \text{ and } \mathcal{R}(u) \leq C \}.
\]
Moreover, for \( u \in U_C \) and \( r > 0 \), let \( u_r := ru + (1 - r)u^\dagger \). We then need to prove that 
\[
\inf_{u \in U} \langle \mathcal{A}(u_r), u - u^\dagger \rangle \to \infty \quad \text{as } r \to \infty.
\]
Introducing 
\[
V := \{ u \in L^2(I) : \|v - v^\dagger\|_{L^2(I)} = 1 \text{ and } |v|_{TV} \leq C \}
\]
and 
\[
\nu_r := rv + (1 - r)v^\dagger
\]
for \( v \in V \) and \( r > 0 \), we see that this is equivalent to showing that 
\[
\inf_{v \in V} \langle \mathcal{A}(\nu_r), \mathcal{I}(v - v^\dagger) \rangle \to \infty \quad \text{as } r \to \infty.
\]
The proof of this result will be based on comparison principles and relating solutions of the PDE (40) with the flow of an ODE.
Remark 7.1. In the remaining text we identify \( v \in BV(I) \) with any of its good representatives \( \widetilde{v} \), for which the measure-theoretic TV (32) coincides with the classical pointwise TV. Here \( \widetilde{v}: I \to \mathbb{R} \) is pointwise defined and assumes a value between (typically the average of) the one-sided essential limits

\[
\nu_{\text{left}}(t) := \text{ess lim}_{s \downarrow t} \nu(s) \quad \text{and} \quad \nu_{\text{right}}(t) := \text{ess lim}_{s \uparrow t} \nu(s)
\]

for every \( t \in I \), so that \( \widetilde{v} \) agrees with the essential pointwise limit of \( \nu \) wherever \( \nu \) is essentially continuous.

To this end, for every \( w \in L^1(I) \) let \( z_w: I \to \mathbb{R} \) denote the unique mild solution of the ODE

\[
z'_w + \varphi(z_w) = w, \quad z_w(0) = 0
\]

provided by the theory of gradient flows [6, Theorem 4.11], where we remember that \( \varphi \) is the subgradient of \( \frac{1}{p+1} \cdot | \cdot |^{p+1} \). Observe also that if \( w \) is continuous, the Peano existence theorem would have guaranteed a global (since \( \varphi \) is weakly coercive) solution in \( I \) — in spite of \( (t, z) \mapsto -\varphi(z) + w(t) \) failing to be locally Lipschitz continuous in \( z \). Moreover, uniqueness follows from the fact that \( -\varphi \) is decreasing.

We then show that condition (8) for the PDE (40) may be reduced to that of the ODE (44).

Lemma 7.2. There exists a constant \( K > 0 \) independent of \( r > 0 \) and \( v \in V \) such that

\[
|z_{r(v-v^\dagger)}(t) - A(iv_r)(t, x)| \leq K
\]

for almost every \( (t, x) \in I \times \Omega \).

Proof. Define

\[
f^+ := \max\{0, \text{ess sup}(f + v^\dagger)\}, \quad g^+ := \max\{0, \text{ess sup} g\},
\]

\[
f^- := \min\{0, \text{ess inf}(f + v^\dagger)\}, \quad \text{and} \quad g^- := \min\{0, \text{ess inf} g\},
\]

and for fixed \( r > 0 \) and \( v \in V \) denote by \( y_{r,v}^\pm \) the solutions of the PDEs

\[
\begin{cases}
\partial_t y_{r,v}^\pm - \Delta y_{r,v}^\pm + \varphi(y_{r,v}^\pm) = f^\pm + r(v(t) - v^\dagger(t)) & \text{in } I \times \Omega; \\
\partial_n y_{r,v}^\pm = 0 & \text{on } I \times \partial \Omega; \\
y_{r,v}^\pm(0, \cdot) = g^\pm & \text{on } \Omega.
\end{cases}
\]

Recall here that \( v^\dagger \in BV(I) \subset L^\infty(I) \), \( f \in L^\infty(I \times \Omega) \), and \( g \in L^\infty(\Omega) \), which implies that \( f^\pm \) and \( g^\pm \) are finite numbers. Then the comparison principle (42) yields that

\[
y_{r,v}^\pm(t, x) \leq A(iv_r)(t, x) \leq y_{r,v}^+(t, x)
\]

for almost every \( (x, t) \in I \times \Omega \). In addition, it is straightforward to see that \( y_{r,v}^\pm \) actually are constant with respect to the \( x \)-variable and can be written in the form \( y_{r,v}^\pm(t, x) = z_{r,v}^\pm(t) \), where the functions \( z_{r,v}^\pm: I \to \mathbb{R} \) solve the ODEs

\[
\frac{dz_{r,v}^\pm}{dt} + \varphi(z_{r,v}^\pm) = f^\pm + r(v - v^\dagger), \quad z_{r,v}^\pm(0) = g^\pm.
\]
Thus
\[
\frac{d}{dt}(z_{r,v}^+ - z_{r,v}^-) = f^+ - f^- - (\varphi(z_{r,v}^+) - \varphi(z_{r,v}^-)) \leq f^+ - f^-,
\]  
which implies that
\[
|z_{r,v}^+(t) - z_{r,v}^-(t)| \leq g^+ - g^- + t(f^+ - f^-)
\]
for all \( t \in I \). Since we also have the ODE-variant
\[
z_{r,v}^-(t) \leq z_{r,v}(t) \leq z_{r,v}^+(t)
\]
of the comparison principle (42), we obtain the assertion with \( K = g^+ - g^- + f^+ - f^- \).
\(\square\)

As a consequence of Lemma 7.2, it suffices to show that
\[
\lim_{r \to \infty} \inf_{v \in V} \{z_{r,v}, v - v^\dagger\}_\{L^2(I)\} = +\infty.
\]

On this path, we estimate
\[
\inf_{v \in V} \{z_{r,v}, r v - v^\dagger\}_\{L^2(I)\} \geq \inf_{v \in V} \{z_{r,v}, r v\}_\{L^2(I)\} - \sup_{v \in V} \|z_{r,v}\|_{L^1(I)} \|v^\dagger\|_{L^\infty(I)},
\]  
and note that for all \( v \in V \) we have
\[
\|v\|_{L^\infty(I)} \leq \|v\|_{L^1(I)} + |v|_{TV(I)} \leq \|v\|_{L^2(I)} + |v|_{TV(I)} \leq 1 + C.
\]

Thus we may estimate
\[
-(1 + C)rt \leq z_{r,v}^-(t) \leq z_{r,v}(t) \leq z_{r,v}^+(t) \leq (1 + C)rt
\]
for all \( r > 0 \) and \( v \in V \), similarly as (45), which leads to
\[
\sup_{v \in V} \|z_{r,v}\|_{L^1(I)} \leq \frac{1}{2}r(1 + C),
\]
expressing that the last term in (46) has at most linear rate in \( r \). Since moreover
\[
\{z_{r,v}, r v\}_\{L^2(I)\} = \frac{1}{2}|z_{r,v}(1)|^2 + \langle \varphi(z_{r,v}), z_{r,v}\rangle_{L^2(I)} \geq \langle \varphi(z_{r,v}), z_{r,v}\rangle_{L^2(I)}
\]
it is sufficient to show that
\[
r \mapsto \inf_{v \in V} \langle \varphi(z_{r,v}), z_{r,v}\rangle_{L^2(I)}
\]
grows superlinearly.

**Lemma 7.3.** There exist numbers \( c > 0 \) and \( \tau > 0 \), such that for each \( v \in V \) there exists an interval \( J_v := (t_v - 2\tau, t_v + 2\tau) \subseteq I \) with either \( v(t) - v^\dagger(t) \geq c \) for all \( t \in J_v \) or \( v(t) - v^\dagger(t) \leq -c \) for all \( t \in J_v \).

**Proof.** Due to the weak compactness of \( V \) in \( L^2(I) \), the convex and lower-semicontinuous function \( v \mapsto \|v - v^\dagger\|_{L^1(I)} \) attains its minimum on \( V \), which must be nonzero in light of the definition of \( V \). Thus there exists an \( a > 0 \) such that \( \|v - v^\dagger\|_{L^1(I)} \geq 2a \) for all \( v \in V \). Since \( |I| = 1 \), we find from (47) that
\[
\left| \{ t \in I : |v(t) - v^\dagger(t)| \geq a \} \right| \geq \frac{a}{\|v - v^\dagger\|_{L^\infty(I)}} \geq \frac{a}{1 + C + |v^\dagger|_{BV}} =: 2m.
\]
Consequently, at least one of the sets \( \{ t : v(t) - v^r(t) > a \} \) and \( \{ t : v(t) - v^r(t) <= -a \} \) has measure greater than or equal to \( m \), and without loss of generality we may assume the former.

Denote by \( A_k \), for \( k \) in some countable index set \( K \), the connected components of the (possibly larger) set \( \{ t : v(t) >= a/2 \} \), and let \( K' \subseteq K \) be the set of indices for which \( \sup_{t \in A_k} v(t) >= a \). Note that for any distinct \( k_1, k_2 \in K' \), the total variation of \( v \) on the convex hull of \( A_{k_1} \cup A_{k_2} \) is at least \( a \), because \( v \) must pass from \( a \) to \( a/2 \) and back to \( a \) again. Since \( |v|_{TV(t)} <= c \), this implies that \( |K'| <= C + 1 =: n \), which also covers the trivial case with \( v \) being constant \( (C = 0) \). On the other hand, we have that \( \sum_{k \in K'} |A_k| >= m \) by construction. Thus there exists at least one connected component \( A_k \) for which \( |A_k| >= m/n \).

Hence, the assertion holds with \( c = a/2 \) and \( \tau = m/n \), both being independent of \( v \in V \). \hfill \( \Box \)

**Lemma 7.4.** Let \( c \) and \( \tau \) be as in Lemma 7.3 and denote by \( w_r : \mathbb{R} \to \mathbb{R} \) the solution of the ODEs

\[
w'_r = r - \varphi(w_r), \quad w_r(0) = 0.
\]

Then

\[
\inf_{v \in V} \left\{ \varphi(z_{r,v}), z_{r,v} \right\}_{L^2(I)} \geq \tau w_{rc}(\tau) \varphi(w_{rc}(\tau)) \tag{48}
\]

for all \( r > 0 \).

**Proof.** Let \( v \in V \) and \( c, \tau, \) and \( J_v = (t_v - 2\tau, t_v + 2\tau) \) be as in Lemma 7.3. Then either \( v(t) >= c \) for all \( t \in J_v \) or \( v(t) <= -c \) for all \( t \in J_v \). We assume and prove the result for the former case; the argument for the other situation is similar. In particular, we find from (44) that \( z'_{r,v} >= cr - \varphi(z_{r,v}) \) on \( J_v \) also.

If \( z_{r,v}(t_v) >= 0 \), then \( z_{r,v}(t) >= \tilde{w}(t) \) for all \( t \in [t_v, t_v + 2\tau] \), where \( \tilde{w} \) solves the initial-value problem

\[
\tilde{w}' = cr - \varphi(\tilde{w}), \quad \tilde{w}(t_v) = 0. \tag{49}
\]

Moreover, \( \tilde{w} \) is nonnegative and increasing on \([t_v, t_v + 2\tau] \). Since \( \varphi \) is increasing and \( \varphi(0) = 0 \), this leads to

\[
\left\{ z_{r,v}, \varphi(z_{r,v}) \right\}_{L^2(I)} \geq \int_{t_v + \tau}^{t_v + 2\tau} \tilde{w}(t) \varphi(\tilde{w}(t)) \, dt \geq \tau \tilde{w}(t_v + \tau) \varphi(\tilde{w}(t_v + \tau)).
\]

Now we note that \( \tilde{w}(t_v + \tau) = w_{rc}(\tau) \), which proves (48).

If, however, \( z_{r,v}(t_v) <= 0 \), then \( z_{r,v}(t) <= \tilde{w}(t) \) for all \( t \in [t_v - 2\tau, t_v] \), where \( \tilde{w} \) again solves (49), but now considered as a final-value problem. Similarly as before, we find that

\[
\left\{ z_{r,v}, \varphi(z_{r,v}) \right\}_{L^2(I)} \geq \int_{t_v - 2\tau}^{t_v - \tau} \tilde{w}(t) \varphi(\tilde{w}(t)) \, dt \geq \tau \tilde{w}(t_v - \tau) \varphi(\tilde{w}(t_v - \tau)) = \tau w_{rc}(-\tau) \varphi(w_{rc}(-\tau)),
\]

and this establishes (48) by noting that \( w_{rc}(-\tau) = -w_{rc}(\tau) \) and \( \varphi(w_{rc}(-\tau)) = -\varphi(w_{rc}(\tau)) \) due to antisymmetry of \( \varphi \). \hfill \( \Box \)

As a consequence of Lemma 7.4, it is sufficient to show that the function \( r \mapsto w_r(\tau) \varphi(w_r(\tau)) \) grows superlinearly as \( r \to \infty \). To that end, we distinguish between the two cases \( 0 < \lambda < 1 \) and \( \lambda > 1 \) concerning the growth (39) of \( \varphi \).
If $0 < \lambda \leq 1$, then $\varphi(z) \leq 1 + z$ for all $z \geq 0$. Thus $w_r$ is bounded below by the solution of the ODE

$$y'_r = r - 1 - y_r, \quad y_r(0) = 0,$$

and a brief computation reveals that $y_r(t) = (r - 1)(1 - e^{-t})$. Accordingly,

$$w_r(\tau) \varphi(w_r(\tau)) = w_r(\tau)^{1 + \lambda} \geq y_r(\tau)^{1 + \lambda} \geq r^{1 + \lambda}$$

as $r \to +\infty$, proving the superlinear growth.

Next we consider the case $\lambda > 1$, where we first note that $w_r$ is bounded above by the stationary point $\varphi^{-1}(r) = r^{1/\lambda}$. Since $r \mapsto \varphi(r) = r^{\lambda}$ is convex on $[0, +\infty)$, we may estimate

$$r - z^{\lambda} = \frac{\varphi(r^{1/\lambda}) - \varphi(z)}{r^{1/\lambda} - z} \geq \frac{\varphi(r^{1/\lambda}) - \varphi(0)}{r^{1/\lambda} - 0} = r^{\frac{\lambda - 1}{\lambda}}$$

for $0 \leq z \leq r^{1/\lambda}$, which becomes

$$r - z^{\lambda} \geq r^{\frac{\lambda - 1}{\lambda}} (r^{1/\lambda} - z).$$

As such, $w_r$ is bounded below by the solution of the ODE

$$y'_r = r^{\frac{\lambda - 1}{\lambda}} (r^{1/\lambda} - y_r), \quad y_r(0) = 0.$$

One sees that $y_r(t) = r^{1/\lambda}(1 - \exp(-r^{\frac{1}{\lambda}} t))$. Since the exponential term in $y_r$ tends, for all fixed $t > 0$, to 0 as $r \to +\infty$, we thus obtain that

$$w_r(\tau) \varphi(w_r(\tau)) = w_r(\tau)^{1 + \lambda} \geq y_r(\tau)^{1 + \lambda} \geq r^{1 + \frac{1}{\lambda}}$$

for sufficiently large $r$, again resulting in superlinear growth.

Collecting all the results above, it follows that subgradient-based Lavrentiev regularisation of the form (43) is a well-posed regularisation method.

### 7.3 Numerical formulation and example

We now reformulate the regularisation problem (43) in order to make it numerically tractable. To that end, let $\pi := i^* : X \to L^2(I)$ be the projection

$$\pi u(t) = \frac{1}{|\Omega|} \int_\Omega u(t, x) \, dx,$$

and define $S : L^2(I) \to [0, \infty]$ by

$$S(v) := \begin{cases} |v|_{TV(I)} & \text{if } v \in BV(I); \\ +\infty & \text{else.} \end{cases}$$

Then we have that $\mathcal{R} = S \circ \pi + \chi_{\text{ran } i}$, where

$$\chi_{\text{ran } i}(u) := \begin{cases} 0 & \text{if } u \in \text{ran } i; \\ \infty & \text{else,} \end{cases}$$
is the indicator function on \( \text{ran} \). Since \( \partial \chi_{\text{ran}}(u) = \ker \pi \), subdifferential calculus \([7, \text{Thm. 16.47}]\) gives that
\[
\partial R(u) = \begin{cases} 
\partial S(\pi u) + \ker \pi & \text{if } u \in \text{ran} \\
\emptyset & \text{else}
\end{cases}
\]

Instead of solving for \( u_a^\delta \in X \), we may therefore rewrite this as a problem for \( v_a^\delta \in L^2(I) \) with \( iv_a^\delta = u_a^\delta \). Thus we arrive at the equivalent reformulation of (43) as \( u_a^\delta = iv_a^\delta \), where \( v_a^\delta \in L^2(I) \) solves the inclusion
\[
A(iv_a^\delta + f) + \alpha \partial S(v_a^\delta) + \ker \pi \ni y_a^\delta,
\]
which can again be rewritten as
\[
-\pi(A(iv_a^\delta + f) - y_a^\delta) \in \alpha \partial S(v_a^\delta),
\]
using that \( \pi \circ i = \text{id}_{L^2(I)} \). Or, introducing the auxiliary variable \( \xi_a^\delta \in \partial S(v_a^\delta) \), we obtain the system
\[
\begin{cases} 
\pi A(iv_a^\delta + f) + \alpha \xi_a^\delta = \pi y_a^\delta \\
\xi_a^\delta \in \partial S(v_a^\delta).
\end{cases}
\]

\( (a) \) Exact data: \( y^\dagger \).
\( (b) \) Noisy data: \( y^\delta \).

\( (c) \) Projected data:
- \( \pi y^\dagger \) (dashed black);
- \( \pi y_a^\delta \) (solid gray);
- \( \pi y_a^\delta \) (solid red).

\( (d) \) Solutions:
- \( v^\dagger \) (dashed black);
- \( v_a^\delta \) (solid red).

Figure 3: TV denoising in the parameter-identification problem (43) with the choice (51) in (40). By \( y_a^\delta \) we mean the solution of (40) corresponding to the reconstruction \( u_a^\delta = iv_a^\delta \).
In order to solve (50), we employ the classical taut-string algorithm [13, 14, 17] for \( v_\alpha^\delta \) on \( I \), noting that at each step in the algorithm, one has to solve the PDE (40) and perform projection onto \( L^2(I) \). As an example, we perform TV denoising in (40) with

\[
\varphi(y) = y^3,
\]

\[
f(x) = 20(\sin(2\pi x))^3 + 2\sqrt{x},
\]

and \( g(x) = \frac{1}{2}\sin(\pi x) \).

The typical performance under moderate gaussian noise in \( I \times \Omega \) can be seen in Figure 3. The PDE-solution \( y^\dagger \) has been computed (from \( v^\dagger \)) using a much finer resolution in order to avoid inverse crimes. Observe that we essentially only work with the projected noisy data in \( I \) rather than the spatially-dependent data on \( I \times \Omega \), which explains the adequate reconstructions.

With regards to convergence rates, it is not yet clear how one can verify the conditions in Theorem 2.8. Nevertheless, we demonstrate numerical convergence in Figure 4, with estimated rates in the \( L^2 \) (that is, \( X \)) and \( L^1 \) norms. Note that the \( L^1 \) rate is approximately twice that of the \( L^2 \) rate. We suspect that this behaviour is due to how errors in the estimated positions of the jumps of the solution are treated by the \( L^2 \) and \( L^1 \) norm, respectively: If the true solution has a jump at a position \( x_0 \), but the reconstruction places it at \( x_0 \pm \epsilon \), the resulting \( L^1 \)-error will be of order \( \epsilon^1 \), whereas the \( L^2 \)-error is only of order \( \epsilon^{1/2} \). We remark also that the deviations for small noise levels \( \delta \) are due to the fact that the PDE-reconstructions \( y_\alpha^\delta \) (the solution of (40) corresponding to the reconstruction \( u_\delta^\alpha = v_\alpha^\delta \)) catch up with the finite resolution of the reference PDE-solution \( y^\dagger \).

Figure 4: Observed convergence rates in the \( L^2 \) and \( L^1 \) norms for the parameter-identification problem (43) with (51) in (40).

8 Conclusions and further work

In this paper we have introduced subgradient-based Lavrentiev regularisation as an alternative to Tikhonov regularisation for linear and nonlinear inverse problems with monotone forward operators. We have established a well-posedness theory in Banach spaces subject to weak assumptions on both the forward operator and the regulariser, and proved general convergence-rates results under variational source conditions. In addition, we have studied examples with TV denoising in convolutional Volterra-integral equations and nonlinear parameter-identification problems in parabolic PDEs.

These results open up for new investigations. For instance, one may consider generalisations that include noisy forward operators. It would also be interesting to study the case of nonreflexive Banach...
Proof of Proposition 2.2

We show that conditions (9) and (10) imply the growth condition Assumption 2.1 vi) with $t_0 = 1$.

For convenience, note first that (10) can equivalently be stated as

$$\langle A(u_{\text{ref}} + w) - A_{\text{ref}}, w \rangle \geq \gamma(\|w\|)\left(\|A(u_{\text{ref}} + \frac{w}{\|w\|}) - A_{\text{ref}}\|_{\|w\|}\right)$$

(52)

for all $w \in \mathcal{X}$ with $\|w\|$ sufficiently large. Now let $C > 0$ and define

$$w_{r,u} := u^\dagger - u_{\text{ref}} + r(u - u^\dagger),$$

(53)

for $u \in U_C$ and $r > 0$. By Assumption 2.1 iii) and v) the set $U_C$ is compact, which implies that

$$K := \left\{ \frac{w_{r,u}}{\|w_{r,u}\|} : u \in U_C \text{ and } r \geq \|u^\dagger - u_{\text{ref}}\| + 1 \right\},$$

is a precompact subset of the unit sphere in $\mathcal{X}$. Since $A$ is demicontinuous (being monotone and hemicontinuous), the mapping

$$\nu \mapsto \langle A(u_{\text{ref}} + \nu) - A_{\text{ref}}, \nu \rangle$$

is continuous, and because of the strict monotonicity of $A$, it follows that

$$c_0 := \inf_{\nu \in K} \langle A(u_{\text{ref}} + \nu) - A_{\text{ref}}, \nu \rangle > 0.$$

In order to establish (8), we see from (53) that $ru + (1 - r)u^\dagger = u_{\text{ref}} + w_{r,u}$, meaning that

$$\langle A(ru + (1 - r)u^\dagger), u - u^\dagger \rangle = \frac{1}{r} \langle A(u_{\text{ref}} + w_{r,u}), w_{r,u} - (u^\dagger - u_{\text{ref}}) \rangle.$$

We then decompose the latter expression as follows (ignoring $1/r$ for the moment):

$$\langle A(u_{\text{ref}} + w_{r,u}), w_{r,u} - (u^\dagger - u_{\text{ref}}) \rangle = \langle A(u_{\text{ref}} + w_{r,u}) - A_{\text{ref}}, w_{r,u} \rangle + \langle A_{\text{ref}}, w_{r,u} \rangle - \langle A(u_{\text{ref}} + w_{r,u}), u^\dagger - u_{\text{ref}} \rangle.$$

The growth condition (52) and the definition of $c_0$ imply that

$$I \geq \gamma(\|w_{r,u}\|)\left(\|A(u_{\text{ref}} + \frac{w_{r,u}}{\|w_{r,u}\|}) - A_{\text{ref}}\|_{\|w_{r,u}\|} \right) \geq c_0 \gamma(\|w_{r,u}\|)$$

whenever $r \geq \|u^\dagger - u_{\text{ref}}\| + 1$ (so that $w_{r,u}/\|w_{r,u}\| \in K$). Moreover,

$$II \geq -\|A_{\text{ref}}\|_{\mathcal{X}} \|w_{r,u}\| \quad \text{and} \quad III \leq \|A(u_{\text{ref}} + w_{r,u})\|_{\mathcal{X}} \|u^\dagger - u_{\text{ref}}\|,$$
so that, in total,
\[
\langle A(ru + (1 - r)u^\dagger), u - u^\dagger \rangle = \frac{1}{r} (I + II - III) \\
\geq \frac{\gamma (\|w_{r,\text{ref}}\|)}{r} \left( c_0 - \frac{\|A(u_{\text{ref}} + w_{r,\text{ref}})\|_{X^*}}{\gamma (\|w_{r,\text{ref}}\|)} \|u^\dagger - u_{\text{ref}}\| \right) \\
- \|Au_{\text{ref}}\|_{X^*} \frac{\|w_{r,\text{ref}}\|}{r}.
\]
(54)

The latter term is uniformly bounded for all \( r > 2\|u^\dagger - u_{\text{ref}}\| \) (and \( u \in U_C \)), as seen from
\[
\frac{\|w_{r,\text{ref}}\|}{r} = \left\| \frac{u^\dagger - u_{\text{ref}}}{r} + u - u^\dagger \right\|,
\]
which together with \( \|u - u^\dagger\| = 1 \) implies that \( \frac{1}{2} \leq \frac{\|w_{r,\text{ref}}\|}{r} \leq \frac{3}{2} \). Since
\[
\sup_{u \in U_C} \frac{\|A(u_{\text{ref}} + w_{r,\text{ref}})\|_{X^*}}{\gamma (\|w_{r,\text{ref}}\|)} \xrightarrow{r \to \infty} 0
\]
due to (9), it follows that the big parenthesis in (54) eventually becomes positive for sufficiently large \( r > 0 \) independent of \( u \). Thus the entire right-hand side converges to \( \infty \) as \( r \to \infty \), uniformly over \( u \in U_C \), in light of the superlinear growth of \( \gamma \). This proves the growth condition (8) in Assumption 2.1 vi).

**Remark A.1.** As seen from (54), the spherical growth condition (9) is unnecessary if \( u_{\text{ref}} = u^\dagger \).

### References

[1] R. Acar and C. R. Vogel, Analysis of bounded variation penalty methods for ill-posed problems, *Inverse Problems*, 10 (6), 1217–1229, 1994.

[2] Y. Alber and I. Ryazantseva, *Nonlinear Ill-posed Problems of Monotone Type*, 1st ed. Springer Netherlands, 2006.

[3] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems* (Oxford Mathematical Monographs). Oxford: Clarendon Press, 2000.

[4] I. K. Argyros, Y. J. Cho, and S. George, Expanding the applicability of Lavrentiev regularization methods for ill-posed problems, *Bound. Value Probl.*, 2013 (114), 2013.

[5] S. N. Askhabov, Equations of convolution type with a monotone nonlinearity on an interval, *Differ. Equ.*, 51 (9), 1173–1179, 2015.

[6] V. Barbu, *Nonlinear Differential Equations of Monotone Type in Banach Spaces* (Springer Monographs in Mathematics). Springer, New York, 2010.

[7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (CMS Books in Mathematics), 2nd ed. Springer International Publishing, 2017.

[8] R. I. Bot and B. Hofmann, Conditional stability versus ill-posedness for operator equations with monotone operators in Hilbert space, *Inverse Problems*, 32 (12), 125003, 2016.

[9] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert* (North-Holland Mathematics Studies). North-Holland Pub. Co., 1973, vol. 5.

[10] F. E. Browder, On the unification of the calculus of variations and the theory of monotone nonlinear operators in Banach spaces, *Proc. Natl. Acad. Sci. USA*, 56 (2), 419–425, 1966.

[11] ——, Nonlinear maximal monotone operators in Banach space, *Math. Ann.*, 175 (2), 89–113, 1968.

[12] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-point algorithms for inverse problems in science and engineering*, ser. Springer Optim. Appl. Vol. 49, Springer, New York, 2011, pp. 185–212.

[13] L. Condat, A Direct Algorithm for 1-D Total Variation De-noising, *IEEE Signal Processing Letters*, 20 (11), 1054–1057, 2013.

[14] P. L. Davies and A. Kovacevic, Local Extremes, Runs, Strings and Multiresolution, *Ann. Statist.*, 29 (1), 1–65, Feb. 2001.

[15] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Mathematics and its Applications). Kluwer Academic Publishers Group, Dordrecht, 1996, vol. 375.
[16] S. George and C. Sreedeep, Lavrentiev's regularization method for nonlinear ill-posed equations in Banach spaces, *Acta Math. Sci.*, 38 (1), 303–314, 2018.

[17] M. Grasmair, The Equivalence of the Taut String Algorithm and BV-Regularization, *J. Math. Imaging Vision*, 27 (1), 59–66, 2007.

[18] ——, Generalized Bregman distances and convergence rates for non-convex regularization methods, *Inverse Problems*, 26 (11), 115014, 2010.

[19] C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (Research Notes in Mathematics). Pitman (Advanced Publishing Program), Boston, MA, 1984, vol. 105.

[20] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators, *Inverse Problems*, 23 (3), 987–1010, 2007.

[21] B. Hofmann, B. Kaltenbacher, and E. Resmerita, Lavrentiev’s regularization method in Hilbert spaces revisited, *Inverse Probl. Imaging*, 10 (3), 741–764, 2016.

[22] B. Hofmann and M. Yamamoto, On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems, *Appl. Anal.*, 89 (11), 1705–1727, 2010.

[23] K. Kielak, P. B. Mucha, and P. Rybka, Almost classical solutions to the total variation flow, *J. Evol. Equ.*, 13 (1), 21–49, 2013.

[24] N. Komodakis and J.-C. Pesquet, Playing with Duality: An overview of recent primal–dual approaches for solving large-scale optimization problems, *IEEE Signal Processing Magazine*, 32 (6), 31–54, 2015.

[25] P. Mahale and M. T. Nair, Lavrentiev regularization of nonlinear ill-posed equations under general source condition, *J. Nonlinear Anal. Optim.*, 4 (2), 193–204, 2013.

[26] E. Mammen and S. van de Geer, Locally adaptive regression splines, *Ann. Statist.*, 25 (1), 387–413, 1997.

[27] M. T. Nair and U. Tautenhahn, Lavrentiev Regularization for Linear Ill-Posed Problems under General Source Conditions, *Z. Anal. Anwend.*, 23 (1), 167–185, 2004.

[28] J. A. Nohel and D. F. Shea, Frequency domain methods for Volterra equations, *Adv. Math.*, 22 (3), 278–304, 1976.

[29] R. Plato and B. Hofmann, A Regularized Variational Inequality Approach for Nonlinear Monotone Ill-Posed Equations, *J. Optim. Theory Appl.*, 182 (2), 525–539, 2019.

[30] R. Plato, P. Mathé, and B. Hofmann, Optimal rates for Lavrentiev regularization with adjoint source conditions, *Math. Comp.*, 87 (310), 785–801, 2018.

[31] A. Răşcanu, On the maximal monotonicity of subdifferential operators, *Rev. Roumaine Math. Pures Appl.*, 66 (1), 237–242, 2021.

[32] R. T. Rockafellar, Local boundedness of nonlinear monotone operators, *Michigan Math. J.*, 16 (4), 397–407, Dec. 1969.

[33] L. I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D*, 60 (1), 259–268, 1992.

[34] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational methods in imaging* (Applied Mathematical Sciences). Springer, New York, 2009, vol. 167.

[35] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations* (Mathematical Surveys and Monographs). American Mathematical Society, 1997, vol. 49.

[36] O. J. Staffans, Positive definite measures with applications to a Volterra equation, *Trans. Amer. Math. Soc.*, 218, 219–237, 1976.

[37] G. Steidl, S. Didas, and J. Neumann, Relations Between Higher Order TV Regularization and Support Vector Regression, in *Scale Space and PDE Methods in Computer Vision*, Springer Berlin Heidelberg, 2005, pp. 515–527.

[38] U. Tautenhahn, On the method of Lavrentiev regularization for nonlinear ill-posed problems, *Inverse Problems*, 18 (1), 191–207, 2002.

[39] E. Zeidler, *Nonlinear Functional Analysis and its Applications, II/B: Nonlinear Monotone Operators*. Springer-Verlag New York, 1990.