

DYNAMIC INVERSE PROBLEMS: ONLINE REGULARISATION THEORY

Jyrki Jauhiainen*

Yassine Nabou†

Tuomo Valkonen‡

Abstract We develop regularisation theory for dynamic inverse problems, solved using online methods with an infinite time horizon. Using concepts of subregularity to treat nonsmooth regularisers, we prove that time-averaged reconstruction errors converge to zero as noise, algorithmic errors, and regularisation vanish as the horizon grows. We illustrate the theory numerically with a dynamic electrical impedance tomography example.

1 INTRODUCTION

In dynamical inverse problems [23, 13, 14, 20, 11, 1], such as flow monitoring using electrical impedance tomography (EIT), one often considers a sequence of nonlinear inverse problems indexed by time. At each time step $k \in \mathbb{N}$, the parameter x_k and the forward operator A_k may change, and only noisy data $b_{k,\delta}$ are available. The underlying *infinite time horizon* operator equations take the form

$$(1.1) \quad (x_1, x_2 \dots) \in \mathcal{X} \quad \text{and} \quad A_k(x_k) = \hat{b}_k \quad \text{for all } k \geq 1,$$

where the set \mathcal{X} encodes a dynamic constraint. It is practically impossible to solve this problem exactly: all of the data will never be available. This naturally leads to the concept of an *online inverse problem*, where reconstructions must be updated sequentially as new data arrive. To place this into context, consider first the classical *static* inverse problem, where the operator and the data do not depend on time:

$$A(x) = \hat{b},$$

where again only noisy measurements b_δ of the exact data \hat{b} are available. A standard Tikhonov-type regularisation scheme reads

$$(1.2) \quad \min_{x \in X} J_\delta(x) + \alpha_\delta R(x), \quad J_\delta(x) = \ell(A(x) - b_\delta),$$

where ℓ is a data fidelity term and R is a regularisation functional. The central asymptotic question is whether minimisers x_δ converge, as $\delta \rightarrow 0$, to an R -minimising solution of the exact equation, that is,

$$\hat{x} \in \arg \min\{R(x) \mid A(x) = \hat{b}\}.$$

This research has been supported by the Academy of Finland grant 345486 and the Special Research Area (SFB) “Mathematics of Reconstruction in Dynamical and Active Models” of the Austrian Science Fund (FWF).

*Department of Mathematics and Scientific Computing, University of Graz, Austria. jyrki.jauhiainen@uni-graz.at, ORCID: 0000-0001-6711-6997

†Department of Mathematics and Statistics, University of Helsinki, Finland. yassine.nabou@helsinki.fi, ORCID: 0009-0004-9805-8039

‡MODEMAT Research Center in Mathematical Modeling and Optimization, Quito, Ecuador and Department of Mathematics and Statistics, University of Helsinki, Finland. tuomo.valkonen@iki.fi, ORCID: 0000-0001-6683-3572

Classical convergence results rely on source conditions, Bregman divergences [7, 6, 3, 15, 9], while we have in [25] exploited concepts of metric subregularity. However, these theories are typically formulated for static problems with fixed operators and data. In contrast, in the dynamic setting both the operators and the data evolve with time. A natural, *formal*, infinite time horizon minimum norm formulation is, therefore

$$(1.3) \quad \min_{(x_1, x_2, \dots) \in \mathcal{X}} \sum_{j=1}^{\infty} R_j(x_j) \quad \text{subject to} \quad A_j(x_j) = \hat{b}_j \quad \text{for all} \quad j = 1, 2, \dots$$

In the presence of noise, the corresponding Tikhonov-type problem becomes

$$(1.4) \quad \min_{(x_1, x_2, \dots) \in \mathcal{X}} \sum_{j=1}^{\infty} \ell_j(A_j(x_j) - b_{j,\delta}) + \sum_{j=1}^{\infty} \alpha_{j,\delta} R_j(x_j).$$

This formulation highlights the structural differences between static and dynamic inverse problems and motivates the development of an online regularisation theory that extends classical convergence concepts to time-evolving operators and data.

Contributions In Section 3, we develop **online regularisation theory** for dynamic inverse problems of the form (1.1), where only noisy measurements $b_{k,\delta}$ of the exact data \hat{b}_k are available. Let $\hat{x}^N = (\hat{x}_0^N, \dots, \hat{x}_N^N)$ solve an initial segment of length $N+1$ of (1.3), and let $x_\delta^N = (x_{\delta,0}^N, \dots, x_{\delta,N}^N)$ *approximately* solve the corresponding initial segment of (1.4). Denote the corresponding solution quality errors by $e_{k,\delta}$. Under mild assumptions on the temporal evolution of the data and of the forward operators A_k , we prove that the average

$$\frac{\|\hat{x}^N - x_\delta^N\|^2}{N+1} \rightarrow 0 \quad \text{as all} \quad \begin{cases} \delta \searrow 0, & \text{noise reduces,} \\ \alpha_\delta \searrow 0, & \text{regularisation reduces,} \\ N \rightarrow \infty, & \text{time advances,} \\ \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \sum_{k=0}^N \frac{e_{k,\delta}}{\alpha_\delta(N+1)} = 0, & \text{solution quality increases.} \end{cases}$$

This extends classical static regularisation theory [12, 7, 27, 24] to the time-dependent setting. Other authors have studied regularisation methods for dynamic inverse problems [22], even with inexact solutions [5], however, no works, before ours, appear to treat infinite time-horizon problems, and online regularisation methods.

As in [27], the solution quality errors $e_{k,\delta}$, allow incorporating information on how practical *online optimisation* methods [10, 26, 11] solve (1.4) inexactly. Using such methods from our companion paper [18], in Section 5, we numerically illustrate the above regularisation theory on an example from dynamic EIT. In these methods, we would have $x_\delta^N = (x_{\delta,0}, \dots, x_{\delta,N})$ for all $N \in \mathbb{N}$ for some $\{x_{\delta,k}\}_{k \in \mathbb{N}}$. That is, the algorithms do not revisit history; they do not refine the solution for time index k after first forming it. Our theory applies to such methods, but equally to x_δ^N formed computationally more naively by exactly solving initial segments of length N of (1.4). Such approximate solutions refine history as $N \rightarrow \infty$. Even in the static case, our results relax some of the assumptions of [27].

Our work is dependent on regularity concepts of set-valued maps, such as metric subregularity, which we recall in the next Section 2. In Section 4, we study the metric subregularity of spatiotemporal infimal convolutions, with the intent of providing ways of verifying the conditions of the theory for regularisation terms arising from the algorithms of [11, 18].

An issue of time indexing We generally use a superscript N to denote a vector x^N that contains information until time index $N \in \mathbb{N}$. To avoid confusion with exponentiation, and to contrast with the

notation A_k for a single time index k , we use the notation $A_{:N}$ to denote quantities indexed up to time N . In particular, $x^N \in X_{:N}$. We will in typical applications have

$$x^N := (x_0^N, \dots, x_N^N) \in X_{:N} \quad \text{with} \quad X_{:N} = \prod_{k=0}^N X_k,$$

but the general theory does not require this. Even when this structure holds, there is not necessarily any correspondence between x_k^N and x_k^M for $0 \leq k \leq \min\{N, M\}$. (See the note about revisiting history above.) When we use the notation $A_{:N}$, and each A_k for $k = 0, \dots, N$ do exist, such a relationship is, however, implied.

Further notation and elementary results We denote the extended reals by $\overline{\mathbb{R}} := [-\infty, \infty]$. We write $H : X \rightrightarrows Y$ when H is a set-valued map from the space X to Y . We write $\mathbb{L}(X; Y)$ for the space of bounded linear operators between the normed spaces X and Y . On a normed space X , for a point $x \in X$ and a set $U \subset X$, we write $\text{dist}(x, U) := \inf_{x' \in U} \|x - x'\|_X$, where $\|\cdot\|_X$ is the norm on X . We also write $\text{dist}^2(x, U) := \text{dist}(x, U)^2$. We write $\langle x, x' \rangle$ for the inner product between two elements x and x' of a Hilbert space X , and $\langle x^* | x \rangle := x^*(x)$ for the dual product or dual pairing in a normed space. We write $\text{Id} : X \rightarrow X$ for the identity operator on X and $\delta_A : X \rightarrow \overline{\mathbb{R}}$ for the $\{0, \infty\}$ -valued indicator function of a set $A \subset X$.

For Fréchet differentiable $F : X \rightarrow \mathbb{R}$, we write $F'(x) \in X^*$ for the Fréchet derivative at $x \in X$. Here X^* is the dual space to X . For a convex function $F : X \rightarrow \overline{\mathbb{R}}$, we write $\partial F : X \rightrightarrows X^*$ for its (set-valued) subdifferential map. The Fenchel conjugate is

$$(1.5) \quad F^*; X^* \rightarrow \overline{\mathbb{R}}, \quad F^*(x^*) := \sup_{x \in X} (\langle x^* | x \rangle_{X^*, X} - F(x)).$$

We write $\text{dom } F := \{x \in X \mid F(x) < \infty\}$ for the effective domain.

For X a Hilbert space, we will frequently use Pythagoras' three-point identity

$$(1.6) \quad \langle x - y, x - z \rangle_X = \frac{1}{2} \|x - y\|_X^2 - \frac{1}{2} \|y - z\|_X^2 + \frac{1}{2} \|x - z\|_X^2 \quad (x, y, z \in X)$$

and (inner product) Young's inequality

$$(1.7) \quad \langle x, y \rangle \leq \|x\|_X \|y\|_X \leq \frac{1}{2\alpha} \|x\|_X^2 + \frac{\alpha}{2} \|y\|_X^2 \quad (x, y \in X, \alpha > 0).$$

2 METRIC SUBREGULARITY AND LOCAL SUBDIFFERENTIABILITY

We briefly recall the main notions of (strong) metric subregularity following [2, 3, 25]. A set-valued mapping $H : X \rightrightarrows Y$ is said to be *metrically subregular* at \hat{x} for $\hat{y} \in H(\hat{x})$ if there exist neighbourhoods $U \ni \hat{x}$, $V \ni \hat{y}$, and a constant $\kappa > 0$ such that

$$\text{dist}(x, H^{-1}(\hat{y})) \leq \kappa \text{dist}(\hat{y}, H(x) \cap V) \quad \text{for all } x \in U.$$

If, in addition, \hat{x} is an isolated point of $H^{-1}(\hat{y})$ and the stronger inequality

$$\|x - \hat{x}\|_X \leq \kappa \text{dist}(\hat{y}, H(x) \cap V) \quad \text{for all } x \in U$$

holds, then H is said to be *strongly metrically subregular* at \hat{x} for \hat{y} .

For convex functionals $F : X \rightarrow \overline{\mathbb{R}}$, the metric and strong metric subregularity of the subdifferential mapping ∂F can be equivalently characterised in terms of local quadratic growth conditions [2]. In particular,

$$(2.1) \quad F(x) \geq F(\hat{x}) + \langle \hat{x}^* | x - \hat{x} \rangle + \gamma \|x - \hat{x}\|_X^2 \quad \text{for all } x \in U \quad \text{and}$$

$$(2.2) \quad F(x) \geq F(\hat{x}) + \langle \hat{x}^* | x - \hat{x} \rangle + \gamma \operatorname{dist}^2(x, [\partial F]^{-1}(\hat{x}^*)) \quad \text{for all } x \in U,$$

correspond, respectively, to *strong metric subregularity* and *metric subregularity* of ∂F at \hat{x} for $\hat{x}^* \in \partial F(\hat{x})$, with the factor $\gamma > 0$ in the neighbourhood U of \hat{x} . The first condition (2.1) expresses a localised version of strong subdifferentiability (equivalent to strong convexity in Hilbert spaces), ensuring isolated minimisers and quadratic stability. By Taylor expansion, the inequality (2.1) holds even for nonconvex F that is twice continuously differentiable. The second condition (2.2) squeezes the set with the same subderivative—in particular, the set of minimisers—into a single point.

For use in our regularisation theory, we extend these concepts to be with respect to an arbitrary seminorm $\|\cdot\|_\circ$ on X , and to incorporate an inexactness $\varepsilon \geq 0$. We also employ the concepts to nonconvex functions. Thus, we call $F : X \rightarrow \overline{\mathbb{R}}$, $\|\cdot\|_\circ$ -*strongly locally ε -subdifferentiable* at $\hat{x} \in X$ for an $\hat{x}^* \in X^*$, if there exists a $\gamma > 0$ and a neighbourhood U of \hat{x} such that

$$(2.3) \quad F(x) \geq F(\hat{x}) + \langle \hat{x}^* | x - \hat{x} \rangle + \gamma \|x - \hat{x}\|_\circ^2 - \varepsilon \quad \text{for all } x \in U.$$

It is $\|\cdot\|_\circ$ -*semi-strongly locally ε -subdifferentiable* at $\hat{x} \in X$ for $\hat{x}^* \in X^*$ if

$$(2.4) \quad F(x) \geq F(\hat{x}) + \langle \hat{x}^* | x - \hat{x} \rangle + \gamma \operatorname{dist}_\circ^2(x, [\partial F]^{-1}(\hat{x}^*)) - \varepsilon \quad \text{for all } x \in U,$$

where

$$\operatorname{dist}_\circ(x, \hat{X}) := \inf_{\tilde{x} \in \hat{X}} \|x - \tilde{x}\|_\circ \quad \text{for any } \hat{X} \subset X.$$

We will omit the mention of ε if it is zero. In practise, as we will see in Section 4, ε will be proportional to the noise level δ .

We also introduce a variant more suitable for use with non-linear Fréchet differentiable forward-operator $A : X \rightarrow Y$. Extending the definition of [25] to semi-norms, we then call $F : X \rightarrow \overline{\mathbb{R}}$ ($A, \|\cdot\|_\circ$)-*semi-strongly locally ε -subdifferentiable* at $\hat{x} \in X$ for an $\hat{x}^* \in X^*$ with respect to a set $\hat{X} \subset X$ if there exist a neighbourhood $U \ni \hat{x}$ and a factor $\gamma > 0$ such that

$$(2.5) \quad F(x) \geq F(\hat{x}) + \langle \hat{x}^* | x - \hat{x} \rangle + \gamma \|A'(\hat{x})(x - \hat{x})\|_Y^2 + \gamma \operatorname{dist}_\circ^2(x, \hat{X}) - \varepsilon \quad (x \in U).$$

This condition interpolates between (2.3) and (2.4) in settings where F may only exhibit growth in certain subspaces (captured by A) and stability with respect to a possibly non-singleton set of ground-truths \hat{X} . This property plays a key role in establishing convergence and stability results for dynamic and online regularisation methods in the subsequent sections.

Remark 2.1. Throughout Section 3, the semi-norm distance $(x, \hat{x}) \rightarrow \|x - \hat{x}\|_\circ$ could be extended to an arbitrary non-negative function; its structure plays no particular role there. In Section 4, we will minimally exploit the specific structure, but that section could also be extended to arbitrary functions if we were content with less explicit results.

3 DYNAMIC REGULARISATION THEORY

We now develop a subregularity-based regularisation theory for dynamic inverse problems and their online solution methods. Our proofs do not require the precise sum-of-data-fidelities-in-product-spaces structure of the introduction, so to reduce notational burden, we slightly relax the structure.

We assume that at each time step $N \in \mathbb{N}$, encompassing the entire history up to that time step, the uncorrupted data $\hat{b}^N \in Y_{:N}$ is generated from a ground-truth $\hat{x}^N \in X_{:N}$ through the ideal model

$$(3.1) \quad A_{:N}(\hat{x}^N) = \hat{b}^N,$$

where $X_{:N}$ is a normed space, and $Y_{:N}$ a Hilbert space. We require that the forward operator $A_{:N} \in C^1(X_{:N}; Y_{:N})$. The true data is not available to us. Instead, we only observe a corrupted measurement

$$b_\delta^N \in Y_{:N}.$$

Our goal, therefore, is not to solve (3.1) at each time step, but rather to track the unknown trajectory \hat{x}^N in an evolving, online fashion. To this end, for the finite but advancing time horizon N , we consider inexact, sequentially constructed, solutions x_δ^N to the regularised cumulative reconstruction problem

$$(3.2a) \quad \min_{x^N \in X_{:N}} J_{:N,\delta}(x^N) + \alpha_\delta R_{:N}(x^N),$$

where $R_{:N} : X_{:N} \rightarrow \overline{\mathbb{R}}$ is a convex spatio-temporal regulariser, $\alpha_\delta > 0$ a regularisation parameter corresponding to the corruption level $\delta > 0$, and

$$(3.2b) \quad J_{:N,\delta}(x^N) = \ell_{:N}(A_{:N}(x^N) - b_\delta^N)$$

for some data fidelity function $\ell_{:N} : Y_{:N} \rightarrow \mathbb{R}$. The regulariser $R_{:N}$ can be used to incorporate the dynamic constraint \mathcal{X} from (1.4) (upto time index N).

The next example explains how, in the typical case, the unknowns and observations encompass the whole history.

Example 3.1 (Simple sequential observations). For a normed space X of unknowns, and a Hilbert space Y of measurements (often \mathbb{R}^n), let us be given $A_k \in C^1(X; Y)$, a data fidelity function $\ell : Y \rightarrow \mathbb{R}$, and corrupted measurements $b_{k,\delta}$ of $A_k \hat{x}_k = \hat{b}_k$. For

$$x^N = (x_0^N, \dots, x_N^N) \in X_{:N} := X^{N+1} \quad \text{and} \quad y^N = (y_0^N, \dots, y_N^N) \in Y_{:N} := Y^{N+1},$$

we then set

$$A_{:N}(x^N) = (A_0(x_0^N), \dots, A_N(x_N^N)) \quad \text{and} \quad \ell_{:N}(y^N) = \sum_{k=0}^N \ell(y_k^N)$$

as well as

$$b_\delta^N = (b_{0,\delta}, \dots, b_{N,\delta}) \quad \text{and} \quad \hat{b}^N = (\hat{b}_0, \dots, \hat{b}_N).$$

Then (3.2b) gives

$$J_{:N,\delta}(x^N) = \sum_{k=0}^N \ell_k(A_k(x_k) - b_{k,\delta}).$$

The regulariser $R_{:N}$ will typically have a more complex structure:

Example 3.2 (Spatiotemporal regularisers arising from online algorithms). For some linear operators $K_k \in \mathbb{L}(X_k; Y_k)$, ($k = 0, \dots, N$), write

$$K_{:N} x^N := (K_0 x_0^N, \dots, K_N x_N^N).$$

Also let $T_k : Y^k \rightarrow \overline{\mathbb{R}}$ be convex, proper, and lower semicontinuous. In the online algorithms of [11, 18], the spatiotemporal regulariser $R_{:N}$ arises as the specific instance¹

$$R_{:N}(x^N) = \mathring{T}_{:N}(K_{:N} x^N)$$

where, for an initial slice $\mathcal{Y}_{:N} \subset Y_{:N}$ of a *dual* dynamic constraint,²

$$\mathring{T}_{:N}(y^N) := \sup_{\tilde{y}^N \in \mathcal{Y}_{:N}} \sum_{k=1}^N [\langle y_k^N, \tilde{y}_k^N \rangle - T_k^*(\tilde{y}_k^N)].$$

If $\mathcal{Y}_{:N}$ were convex, we would have [10] the infimal convolution

$$\mathring{T}_{:N}(y^N) = [T_{:N} \square \delta_{\mathcal{Y}_{:N}}^*](y^N) := \inf_{\tilde{y}^N} \left(T_{:N}(y^N - \tilde{y}^N) + \delta_{\mathcal{Y}_{:N}}^*(\tilde{y}^N) \right)$$

where

$$T_{:N}(y^N) := \sum_{k=1}^N T_k(\eta_k^{-1} y^k) \quad \text{and} \quad \delta_{\mathcal{Y}_{:N}}^*(y^N) = \sup_{\tilde{y}^N \in \mathcal{Y}_{:N}} \langle y^N, \tilde{y}^N \rangle.$$

In typical applications of interest to us, each $T_k = \|\cdot\|_{2,1}$ in a suitable finite element space, and $K_k = \nabla$, so that $T_k \circ K_k$ models an isotropic total variation regulariser for the frame k . We recall that $\|\cdot\|_{2,1}$ models the global L^1 -norm of vector-valued functions with pointwise 2-norms, and $B_{2,\infty}$ is the corresponding dual ball. Thus $R_{:N}$ would be a spatiotemporal infimal convolution of the (dual) dynamic constraint, and the framewise total variation regularisers. When $\mathcal{Y}_{:N}$ is not convex, we instead get a “sub-infimal” convolution: we have $\mathring{T}_{:N} \leq T_{:N} \square \delta_{\mathcal{Y}_{:N}}^*$. We will in [Example 4.4](#) show that equality actually holds here.

We start with our general assumptions and constructions in [Section 3.1](#). We then provide a basic Bregman estimate with minimal assumptions in [Section 3.2](#). To provide norm-convergence to a specific minimum- $R_{:N}$ solution of the noise-free problem, we consider in [Section 3.3](#) a *strong source condition*. We relax this in [Section 3.4](#) to *semi-strong source condition* that gives convergence to the *set* of minimum- $R_{:N}$ solutions.

3.1 GENERAL ASSUMPTIONS AND CONCEPTS

Let $N \in \mathbb{N}$. We denote by $\hat{X}_{:N} \subset X_{:N}$ the set of $R_{:N}$ -minimising solutions

$$(3.3) \quad \hat{X}_{:N} := \arg \min_{x \in X_{:N}} R_{:N}(x) \quad \text{subject to} \quad A_{:N}x = \hat{b}^N$$

We can derive necessary optimality conditions for this problem through the theory of Clarke subdifferentials [8, Section 2.9]. Since $A_{:N} \in C^1(X_{:N}; Y_{:N})$ and $R_{:N}$ is convex and lower semicontinuous, both the functional $x \mapsto R_{:N}(x)$ and the indicator function of the constraint set in (3.3) are regular in the sense of Clarke.³ Consequently, the sum and chain rules for the generalized gradient apply as equalities [8, Thms. 2.9.8–2.9.9]. The Fermat principle yields the desired optimality condition, known in the inverse problems literature as a source condition:

¹To be more precise, in the algorithmic works [11, 18], where the corruption level δ plays no special role, one works instead T_k with $G_k = \alpha_\delta^{-1} T_k$ that already incorporate the regularisation parameter. Then $G_k^* = \alpha_\delta T_k^*(\cdot / \alpha_\delta)$ [9, Lemma 5.8], and one defines $\mathring{G}_{:N}$ analogously to $\mathring{T}_{:N}$. For $T_k = \|\cdot\|_{2,1}$, in particular, $G_k^* = \delta_{\alpha_\delta B_{2,\infty}}$. Hence the dual variables are scaled by α_δ . Correspondingly scaling the dual dynamic constraint $\mathcal{Y}_{:N}$ by α_δ , it follows that $\mathring{G}_{:N} = \alpha_\delta \mathring{T}_{:N}$, so $\alpha_\delta R_{:N}(X_{:N}) = \mathring{G}_{:N}$.

²In [11, Corollary 2.13], the definition of $\mathring{G}_{:N}$ has had the order of the sum and the supremum interchanged. The correct definition is found before [10, Theorem 2.6], on which the former is based.

³That is, for locally Lipschitz functions the Clarke directional derivative f° agrees with the standard directional derivative f' , which is assumed to exist; see [8, Definitions 2.3.4]. For infinite-valued functions, the Clarke tangent cone and the contingent cone of the epigraph have to agree [8, Definitions 2.4.10].

Assumption 3.3 (Basic source condition). For given $N \geq 0$ and $\hat{x}^N \in \hat{X}_{:N}$, there exists $\hat{w}^N \in Y_{:N}^*$ such that

$$(3.4) \quad A_{:N}(\hat{x}^N) = \hat{b}^N \quad \text{and} \quad -A'_{:N}(\hat{x}^N)^* \hat{w}^N \in \partial R_{:N}(\hat{x}^N).$$

For corruption levels $\delta > 0$, we also require inexact regularised solutions that satisfy certain quality guarantees:

Assumption 3.4. For a given $N \geq 0$, let $X_{:N}$ be a normed space and $Y_{:N}$ a Hilbert space, and the regularisation functional $R_{:N} : X_{:N} \rightarrow \overline{\mathbb{R}}$ convex, proper, and lower semicontinuous. Moreover, let $A_{:N} \in C^1(X_{:N}; Y_{:N})$, and let $\ell_{:N} \in C^1(Y_{:N})$. For all corruption levels $\delta \in (0, \delta^*)$ for some $\delta^* > 0$, we then require:

- (i) **Noise levels:** The corrupted measurements $b_\delta^N \in Y_{:N}$ of the ground-truth $\hat{b}^N \in Y_{:N}$ satisfy for some fixed constants $C' > 0$ and $q > 0$

$$(3.5) \quad \frac{1}{N+1} \|\ell'_{:N}(b_\delta^N - \hat{b}^N)\|_{Y_{:N}^*}^2 \leq \delta \quad \text{and} \quad \frac{1}{N+1} \ell_{:N}(b_\delta^N - \hat{b}^N) \leq C' \delta^q.$$

- (ii) **Accuracy of solution of Tikhonov-type problems:** We are given $x_\delta^N \in X_{:N}$ that, for some regularisation parameter $\alpha_\delta > 0$, minimise

$$Q_{:N,\delta}(x^N) := J_{:N,\delta}(x^N) + \alpha_\delta R_{:N}(x^N) \quad \text{where} \quad J_{:N,\delta}(x^N) := \ell_{:N}(A_{:N}(x^N) - b_\delta^N)$$

to an accuracy $e_{N,\delta} \geq 0$ in the sense that, for some (and hence any) minimum- $R_{:N}$ solution $\hat{x}^N \in \hat{X}_{:N}$, we have

$$(3.6) \quad Q_{:N,\delta}(x_\delta^N) \leq Q_{:N,\delta}(\hat{x}^N) + e_{N,\delta}.$$

- (iii) **Quadratic bounds:** For some $C > 0$, we have

$$(3.7) \quad \ell_{:N}(A_{:N}(x_\delta^N) - \hat{b}^N) - \ell_{:N}(A_{:N}(\hat{x}^N) - \hat{b}^N) \leq C \left(\ell_{:N}(A_{:N}(x_\delta^N) - b_\delta^N) + \|\ell'_{:N}(b_\delta^N - \hat{b}^N)\|^2 \right).$$

Remark 3.5 (Quality of algorithmic solutions). In (3.6), we compare the (typically algorithm-constructed) approximate solution x_δ^N for the corruption level $\delta > 0$ against a minimum- $R_{:N}$ solution \hat{x}^N for corruption-free data. We could, instead, compare it against a minimiser \tilde{x}_δ^N of $Q_{:N,\delta}$, i.e., the best regularised solution for the problem with corruption level δ and finite time horizon N . However, the comparison against the ground-truth is, by definition, more permissive, and sufficient for our purposes.

Remark 3.6 (Quadratic bound). The bound (3.7) is variant of the pseudo-Hölder estimate of [27, §5]. If $\ell_{:N} = \frac{1}{2} \|\cdot\|^2$, we have

$$\begin{aligned} \ell_{:N}(A_{:N}(x_\delta^N) - \hat{b}^N) - \ell_{:N}(A_{:N}(x_\delta^N) - b_\delta^N) &= \langle A_{:N}(x_\delta^N), b_\delta^N - \hat{b}^N \rangle_{Y_{:N}} + \frac{1}{2} \|\hat{b}^N\|_{Y_{:N}}^2 - \frac{1}{2} \|b_\delta^N\|_{Y_{:N}}^2 \\ &= \langle A_{:N}(x_\delta^N)_{Y_{:N}} - b_\delta^N, b_\delta^N - \hat{b}^N \rangle_{Y_{:N}} + \frac{1}{2} \|\hat{b}^N - b_\delta^N\|_{Y_{:N}}^2 \\ &\leq \ell_{:N}(A_{:N}(x_\delta^N) - b_\delta^N) + \|\hat{b}^N - b_\delta^N\|_{Y_{:N}}^2, \end{aligned}$$

where in the last step we have used Young's inequality. This verifies (3.7).

Remark 3.7 (Linearisation). Recall that $\ell'_{:N} : Y_{:N} \rightarrow Y_{:N}^*$ is the Fréchet derivative of $\ell_{:N}$, and $B_{\ell_{:N}}(y, y') := \ell_{:N}(y) - \ell_{:N}(y') - \langle \ell'_{:N}(y'), y - y' \rangle$ is the Bregman divergence generated by $\ell_{:N}$. The condition (3.8) is satisfied, for instance, when $\ell_{:N}(z) = \frac{1}{2} \|z\|_{Y_{:N}}^2$ is quadratic and the forward operator $A_{:N}$ is linear (e.g., [27]). A static version of (3.7) has also been considered in [27] when ℓ is quadratic. Thus, our setting allows for more general data fidelity terms.

3.2 A BASIC BREGMAN ESTIMATE

As a starting point for our analysis, we derive preliminary bounds on the distance of x_δ^N to \hat{x}^N in terms of Bregman divergences. These results extend the work of [27, 7] to the online setting. Minding the basic source condition of [Assumption 3.3](#), we define the Bregman divergence generated by $R_{:N}$ at \hat{x}^N for $v \in \partial R_{:N}(\hat{x}^N)$ by

$$B_{R_{:N}}^v(x_\delta^N, \hat{x}^N) := R_{:N}(x_\delta^N) - R_{:N}(\hat{x}^N) + \langle v | x_\delta^N - \hat{x}^N \rangle_{X_{:N}^*, X_{:N}}.$$

We also define the Bregman divergence generated by $\ell_{:N}$,

$$B_{\ell_{:N}}(y, \bar{y}) := \ell_{:N}(y) - \ell_{:N}(\bar{y}) - \langle \ell'_{:N}(\bar{y}) | y - \bar{y} \rangle_{Y_{:N}^*, Y_{:N}}.$$

For the basic estimates of this subsection, in addition to [Assumptions 3.3](#) and [3.4](#), we require the the following approximate linearity assumption.

[Assumption 3.8](#) (Approximate linearised data fidelity condition). There exist $\eta > 0$ such that

$$(3.8) \quad \langle \ell'_{:N}(\hat{b}^N - b_\delta^N) | A_{:N}(x_\delta^N) - A_{:N}(\hat{x}^N) - A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N) \rangle_{Y_{:N}} \\ + B_{\ell_{:N}}(A_{:N}(x_\delta^N) - b_\delta^N, A_{:N}(\hat{x}^N) - b_\delta^N) \geq \eta \|A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N)\|_{Y_{:N}}^2$$

for all $N \in \mathbb{N}$ and $\delta \in (0, \delta^*)$.

[Remark 3.9](#) (Approximate linearity). For $\ell_{:N}(z) = \frac{1}{2} \|z\|_{Y_{:N}}^2$, (3.8) reads

$$(3.9) \quad \langle \hat{b}^N - b_\delta^N | A_{:N}(x_\delta^N) - A_{:N}(\hat{x}^N) - A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N) \rangle_{Y_{:N}} \\ + \frac{1}{2} \|A_{:N}(x_\delta^N) - A_{:N}(\hat{x}^N)\|_{Y_{:N}}^2 \geq \eta \|A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N)\|_{Y_{:N}}^2 \quad \text{for all } N \in \mathbb{N}.$$

This condition was used in [27] for conventional static regularisation. For the quadratic data fidelity, it follows that [Assumption 3.8](#) holds, for instance, when the forward operator $A_{:N}$ is linear.

[Example 3.10](#). Suppose $A_{:N}$ is twice continuously differentiable, and $\gamma M \leq A'_{:N}(x^N)^* A'_{:N}(x^N) \leq CM$ for all $x^N \in X_{:N}$ and some $C, \gamma > 0$ and a positive semi-definite operator M . Then, using the mean value theorem to write $A_{:N}(x_\delta^N) - A_{:N}(\hat{x}^N) = A'_{:N}(\zeta)(x_\delta^N - \hat{x}^N)$ for some $\zeta \in [x_\delta^N, \hat{x}^N]$, it is not difficult to see that (3.9) holds when $\hat{b}^N - b_\delta^N$ is small, i.e., $\delta > 0$ is small.

We now have the following bound.

[Theorem 3.11](#). Suppose that [Assumptions 3.3](#), [3.4](#) and [3.8](#) hold for a $\hat{x}^N \in \hat{X}_{:N}$. If $\delta \in (0, \delta^*)$, then

$$0 \leq \frac{1}{N+1} B_{R_{:N}}^{\hat{v}}(x_\delta^N, \hat{x}^N) \leq \frac{\delta}{2\eta\alpha_\delta} + \frac{e_{N,\delta}}{\alpha_\delta(N+1)} + \frac{\alpha_\delta}{2\eta(N+1)} \|\hat{w}^N\|_{Y_{:N}}^2,$$

where we write $\hat{v} := -A'_{:N}(\hat{x}^N)\hat{w}^N$.

Proof. For brevity, since N is fixed throughout the proof, we drop the indexing by N , writing $\hat{x} := \hat{x}^N$, $\hat{w} := \hat{w}^N$, etc. By [Assumption 3.4](#) and the chosen notation, $A(\hat{x}) = A_{:N}(\hat{x}) = \hat{b}^N = \hat{b}$. The definition of

the Bregman divergence then gives

$$\begin{aligned}
J_{:N,\delta}(x_\delta) - J_{:N,\delta}(\hat{x}) &= \ell(A(x_\delta) - b_\delta) - \ell(A(\hat{x}) - b_\delta) \\
&= \langle \ell'(A(\hat{x}) - b_\delta), A(x_\delta) - A(\hat{x}) \rangle + B_\ell(A(x_\delta) - b_\delta, A(\hat{x}) - b_\delta) \\
&= \langle \ell'(\hat{b} - b_\delta), A(x_\delta) - A(\hat{x}) \rangle + B_\ell(A(x_\delta) - b_\delta, A(\hat{x}) - b_\delta) \\
&\quad - \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle + \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle \\
&= \langle \ell'(\hat{b} - b_\delta), A(x_\delta) - A(\hat{x}) - A'(\hat{x})(x_\delta - \hat{x}) \rangle + B_\ell(A(x_\delta) - b_\delta, A(\hat{x}) - b_\delta) \\
&\quad + \langle \ell'(\hat{b} - b_\delta), A'(\hat{x})(x_\delta - \hat{x}) \rangle - \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle + \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle \\
&\stackrel{(3.8)}{\geq} \eta \|A'(\hat{x})(x_\delta - \hat{x})\|^2 + \langle \ell'(\hat{b} - b_\delta) - \alpha_\delta \hat{w}, A'(\hat{x})(x_\delta - \hat{x}) \rangle \\
&\quad + \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle.
\end{aligned}$$

Continuing with Young's inequality and the triangle inequality, it follows that

$$\begin{aligned}
J_{:N,\delta}(x_\delta) - J_{:N,\delta}(\hat{x}) &\geq -\frac{1}{4\eta} \|\ell'(\hat{b} - b_\delta) - \alpha_\delta \hat{w}\|^2 + \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle \\
&\geq -\frac{1}{2\eta} \|\ell'(\hat{b} - b_\delta)\|^2 - \frac{\alpha_\delta^2}{2\eta} \|\hat{w}\|^2 + \alpha_\delta \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle
\end{aligned}$$

Using the definition of $B_{R,N}^{\hat{\delta}}$ and [Assumption 3.4 \(i\)](#) and [\(ii\)](#), it, therefore, follows that

$$\begin{aligned}
e_{N,\delta} &\geq -\frac{1}{2\eta} \|\ell'(\hat{b} - b_\delta)\|^2 - \frac{\alpha_\delta^2}{2\eta} \|\hat{w}\|^2 + \alpha_\delta (R_{:N}(x_\delta) - R_{:N}(\hat{x}) + \langle A'(\hat{x})^* \hat{w}, x_\delta - \hat{x} \rangle) \\
&\geq -\frac{\delta(N+1)}{2\eta} - \frac{\alpha_\delta^2}{2\eta} \|\hat{w}\|_{Y,N}^2 + \alpha_\delta B_{R,N}^{\hat{\delta}}(x_\delta, \hat{x}),
\end{aligned}$$

Dividing by $\alpha_\delta(N+1)$, the claim follows. The Bregman divergence is $B^{\hat{\delta}}(x_\delta, \hat{x})$ non-negative due its definition and to the source condition [\(3.4\)](#). \square

As a corollary, we get a convergence result for the average performance over long segments, as the corruption level $\delta \rightarrow 0$, and the errors and the regularisation parameter behave appropriately.

Corollary 3.12. *Suppose that [Assumptions 3.3](#), [3.4](#) and [3.8](#) hold for a $\hat{x}^N \in \hat{X}_{:N}$ for all $N \in \mathbb{N}$. If, moreover,*

$$\alpha_\delta \rightarrow 0, \quad \delta/\alpha_\delta \rightarrow 0, \quad \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{k=0}^N \frac{e_{k,\delta}}{\alpha_\delta} = 0, \quad \text{and} \quad \sup_{N \in \mathbb{N}} \frac{1}{N+1} \|\hat{w}^N\|_{Y,N}^2 < \infty,$$

then,

$$\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N+1} B_{R,N}^{\hat{\delta}}(x_\delta^N, \hat{x}^N) = 0.$$

3.3 ESTIMATES BASED ON A STRONG SOURCE CONDITION

Bregman divergences may not be very informative; for example, the Bregman divergence of the absolute value function in \mathbb{R} can only discern signs, but not magnitudes. We therefore next develop (semi-)norm-based estimates. We begin with the following lemma, which will be used to show that the approximate regularised solutions x_δ^N remain close to \hat{x}^N when the noise level, regularisation parameter, and accuracy parameter are sufficiently small.

Lemma 3.13. *Suppose Assumption 3.4 holds at a $\hat{x}^N \in \hat{X}_{:N}$ for an $N \geq 0$. Then*

$$\frac{1}{N+1} \ell_{:N}(A_{:N}(x_\delta^N) - A_k(\hat{x}_k)) \leq \alpha_\delta C r_{N,\delta} \quad \text{and} \quad \frac{1}{N+1} R_{:N}(x_\delta^N) \leq r_{N,\delta}$$

for

$$(3.10) \quad r_{N,\delta} := \frac{C' \delta^q + \delta}{\alpha_\delta} + \frac{e_{N,\delta}}{\alpha_\delta(N+1)} + \frac{1}{N+1} R_{:N}(\hat{x}^N),$$

which is independent of the choice of $\hat{x}^N \in \hat{X}_{:N}$.

Proof. That $r_{N,\delta}$ does not depend on the choice of $\hat{x}^N \in \hat{X}_{:N}$ follows from the definition of $\hat{X}_{:N}$ as the set of $R_{:N}$ -minimising solutions.

By Assumption 3.4 (i) to (iii), and that $A_{:N}(\hat{x}^N) = \hat{b}^N$ for all $k \in \{0, 1, \dots, N\}$, we have

$$\begin{aligned} & C^{-1} \ell_{:N}(A_{:N}(x_\delta^N) - A_{:N}(\hat{x}^N)) + \alpha_\delta R_{:N}(x_\delta^N) \\ & \stackrel{(3.7)}{\leq} \ell_{:N}(A_{:N}(x_\delta^N) - b_\delta^N) + \|\ell'(b_\delta^N - \hat{b}^N)\|_{Y_{:N}}^2 + \alpha_\delta R_{:N}(x_\delta^N) \\ & \stackrel{(3.6)}{\leq} e_{N,\delta} + \|\ell'_{:N}(b_\delta^N - \hat{b}^N)\|_{Y_{:N}}^2 + \ell_{:N}(\hat{b}^N - b_\delta^N) + \alpha_\delta R_{:N}(\hat{x}^N) \\ & \stackrel{(3.5)}{\leq} e_{N,\delta} + (N+1)(C' \delta^q + \delta) + \alpha_\delta R_{:N}(\hat{x}^N). \end{aligned}$$

Hence, after dividing by $N+1$, the assertion follows from the elementary observation that for nonnegative quantities A, B, C , the inequality $A + B \leq C$ implies $A \leq C$ and $B \leq C$. \square

We will need the following “strong source condition” based on strong metric subregularity. We refer to [25] for examples of the satisfaction of the corresponding static condition, in particular, with regard to total variation regularisation. We discuss ways to prove the condition for the spatiotemporal regularisers of Example 3.2 in Section 4.

Assumption 3.14 (Strong source condition). Let $N \in \mathbb{N}$. We say that $\hat{x}^N \in X_{:N}$ satisfies the *strong source condition* if $X_{:N}$ is equipped with a semi-norm $\|\cdot\|_{N,\circ}$, and the following hold:

- (i) The basic source condition (3.4) holds for some for some $\hat{w}^N \in Y_{:N}^*$.
- (ii) There exists a $\delta^* > 0$ such that for all $\delta \in (0, \delta^*)$, for given α_δ , the function $Q_{:N,\delta}$ is $\|\cdot\|_{N,A,\circ}$ -strongly locally $\varepsilon_{N,\delta}$ -subdifferentiable (see (2.3)) at \hat{x}^N for $\hat{x}_{:N}^* := J'_{:N,\delta}(\hat{x}^N) - \alpha_\delta A'_{:N}(\hat{x}^N) \hat{w}^N$ in a neighbourhood $V_{\hat{x}^N,\delta}$ of \hat{x}^N with the factor $\gamma_{N,\delta} > 0$. Here the semi-norm

$$\|x^N\|_{N,A,\circ} := \sqrt{\|A'_{:N}(\hat{x}^N)x^N\|_{Y_{:N}}^2 + \|x^N\|_{N,\circ}^2}.$$

While not included as part of the source condition, to pass to the limit, we also require:

Assumption 3.15. Let Assumption 3.14 hold. We also require:

- (i) For some $\rho > 0$, for all $\delta \in (0, \delta^*)$, we have

$$V_{\hat{x}^N,\delta} \supset U_\rho^N := \left\{ x^N \in X_{:N} \left| \begin{array}{l} \frac{1}{N+1} \ell_{:N}(A_{:N}(x^N) - A_{:N}(\hat{x}^N)) \leq \rho, \\ \frac{1}{N+1} R_{:N}(x^N) \leq \frac{1}{N+1} R_{:N}(\hat{x}^N) + \rho \end{array} \right. \right\}.$$

- (ii) For $r_{N,\delta}$ defined in (3.10), we have $\lim_{\delta \rightarrow 0} \sup_{N \in \mathbb{N}} r_{N,\delta} = \lim_{\delta \rightarrow 0} \sup_{N \in \mathbb{N}} \frac{1}{N+1} R_{:N}(\hat{x}^N) < \infty$, along with $\lim_{\delta \rightarrow 0} \alpha_\delta = 0$

Remark 3.16. [Assumption 3.15 \(i\)](#) is a somewhat strong requirement on the neighbourhood of strong local subdifferentiability of the single point \hat{x}^N : it has to contain the level set neighbourhood U_ρ^N , which is independent of the choice of the point $\hat{x}^N \in \hat{X}_{:,N}$, by the definition of the latter set. We relax this to a covering property when treating the semi-strong source condition of [Section 3.4](#).

The condition [\(ii\)](#) is a uniform-in- N limiting property of the parameters and errors as $\delta \searrow 0$, as well as a bound on the average values of the regularisation term for the minimal solutions. The $\sup_{N \in \mathbb{N}}$ can be relaxed to a $\lim_{N \rightarrow \infty}$ if we only require the lemma that follows the next example to hold for $N \geq N_\delta$ for some $N_\delta \in \mathbb{N}$.

Example 3.17. Suppose $\ell_{:,N}$ is minimised at zero. Let $\hat{x}^N \in \hat{X}_{:,N}$. Then, recalling the basic source condition [\(3.4\)](#), the function Ψ defined by

$$\Psi(x^N) := \ell_{:,N}(A_{:,N}(x^N) - \hat{b}^N) + R_{:,N}(x^N) + \langle A'_{:,N}(\hat{x}^N)^* \hat{w}^N | x^N - \hat{x}^N \rangle$$

is minimised at \hat{x}^N , and satisfies $\Psi(\hat{x}^N) = 0$. (The last two convex terms are minimised at \hat{x}^N and, by assumption, so is the first term.) If Ψ is strongly (globally) $\|\cdot\|_{X_{:,N}}$ -subdifferentiable at \hat{x}^N (for 0), then $U_\rho^N \subset B(\hat{x}^N, r)$ for any $r > 0$ for small enough $\rho > 0$. Consequently, [Assumption 3.15 \(i\)](#) holds.

Note that Ψ differs from $Q_{:,N,\delta}$, which has to be $\|\cdot\|_{N,A,\circ}$ -strongly locally $\varepsilon_{N,\delta}$ -subdifferentiable ([Assumption 3.14 \(ii\)](#)) through the additional linear term, which does not effect the “strength” of the subdifferentiability, and through the true data \hat{b}^N in place of the noisy data b_δ^N . The data can be swapped for small enough δ^* through Lipschitz assumptions on $\ell_{:,N}$, but the required growth here is possibly stronger due to requiring growth with respect to a norm instead of merely a semi-norm. However, $V_{\hat{x}^N,\delta}$ may itself contain a subspace when the growth of $Q_{:,N,\delta}$, is with respect to a proper semi-norm. Therefore, it may also be possible to relax the growth requirement on Ψ .

The next lemma explains why we require [Assumption 3.15](#): it ensures that for small enough $\delta > 0$, the algorithmic solutions x_δ^N belong to the neighbourhood of strong local subdifferentiability.

Lemma 3.18. *Suppose [Assumption 3.15](#) holds. Then there exists $\bar{\delta} \in (0, \delta^*)$ such that $x_\delta^N \in V_{\hat{x}^N,\delta}$ for all $\delta \in (0, \bar{\delta})$ and $N \in \mathbb{N}$.*

Proof. By [Assumption 3.15 \(ii\)](#) and [Lemma 3.13](#), there exists $\bar{\delta} \in (0, \delta^*)$, independent of N , such that $x_\delta^N \in U_\rho^N$ when $\delta \in (0, \bar{\delta})$. By [Assumption 3.15 \(i\)](#), it follows that $x_\delta^N \in V_{\hat{x}^N,\delta}$. \square

We can now obtain our main estimate for the strong source condition. We will afterwards convert it into a convergence result.

Theorem 3.19. *Let $N \in \mathbb{N}$. Suppose [Assumptions 3.4](#), [3.14](#) and [3.15](#) hold at $\hat{x}^N \in X_{:,N}$. Then there exists $\bar{\delta} \in (0, \delta^*)$, independent of N , such that if $\delta \in (0, \bar{\delta})$, we have*

$$(3.11) \quad \frac{1}{N+1} \|x_\delta^N - \hat{x}^N\|_{N,\circ}^2 \leq \frac{1}{\gamma_{N,\delta}} \left(\frac{e_{N,\delta} + \varepsilon_{N,\delta}}{N+1} + \frac{\delta}{2\gamma_{N,\delta}} + \frac{\alpha_\delta^2}{2\gamma_{N,\delta}(N+1)} \|\hat{w}^N\|_{Y_{:,N}}^2 \right).$$

Proof. By [Lemma 3.18](#), there exists $\bar{\delta} > \delta^*$ such that $x_\delta^N \in U_\rho^N$ for all $N \in \mathbb{N}$ when $\delta \in (0, \bar{\delta})$. We also have

$$\begin{aligned} J_{:,N,\delta}(\hat{x}^N) - \alpha_\delta A'_{:,N}(\hat{x}^N)^* \hat{w}^N &= A'_{:,N}(\hat{x}^N)^* (\ell'_{:,N}(A_{:,N}(\hat{x}^N) - b_\delta^N) - \alpha_\delta \hat{w}^N) \\ &= A'_{:,N}(\hat{x}^N)^* (\ell'_{:,N}(\hat{b}^N - b_\delta^N) - \alpha_\delta \hat{w}^N). \end{aligned}$$

Hence, by [Assumption 3.4 \(ii\)](#) and [3.14 \(ii\)](#) followed by Young's inequality [\(1.7\)](#),

$$\begin{aligned}
(3.12) \quad e_{N,\delta} + \varepsilon_{N,\delta} &\stackrel{(3.6)}{\geq} [J_{:N,\delta}(x_\delta^N) + \alpha_\delta R_{:N}(x_\delta^N)] - [J_{:N,\delta}(\hat{x}^N) + \alpha_\delta R_{:N}(\hat{x})] + \varepsilon_{N,\delta} \\
&\stackrel{(2.3)}{\geq} \langle J_{:N,\delta}(\hat{x}^N) - \alpha_\delta A'_{:N}(\hat{x}^N)^* \hat{w}^N | x_\delta^N - \hat{x}^N \rangle + \gamma_{N,\delta} \|x_\delta^N - \hat{x}^N\|_{N,A,\circ}^2 \\
&= \langle \ell'_{:N}(\hat{b}^N - b_\delta^N) - \alpha_\delta \hat{w}^N, A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N) \rangle \\
&\quad + \gamma_{N,\delta} \|A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N)\|_{Y_k}^2 + \gamma_{N,\delta} \|x_\delta^N - \hat{x}^N\|_{N,\circ}^2 \\
&\geq -\frac{1}{4\gamma_{N,\delta}} \|\ell'_{:N}(\hat{b}^N - b_\delta^N) - \alpha_\delta \hat{w}^N\|_{Y_N}^2 + \gamma_{N,\delta} \|x_\delta^N - \hat{x}^N\|_{N,\circ}^2.
\end{aligned}$$

Thus, rearranging and continuing with Young's inequality along with [Assumption 3.4 \(i\)](#), we obtain

$$\begin{aligned}
(3.13) \quad \|x_\delta^N - \hat{x}^N\|_{N,\circ}^2 &\leq \frac{1}{\gamma_{N,\delta}} (e_{N,\delta} + \varepsilon_{N,\delta}) + \frac{1}{4\gamma_{N,\delta}^2} \|\ell'_{:N}(\hat{b}^N - b_\delta^N) - \alpha_\delta \hat{w}^N\|_{Y_N}^2 \\
&\leq \frac{1}{\gamma_{N,\delta}} (e_{N,\delta} + \varepsilon_{N,\delta}) + \frac{\|\ell'_{:N}(\hat{b}^N - b_\delta^N)\|^2}{2\gamma_{N,\delta}^2} + \frac{\alpha_\delta^2}{2\gamma_{N,\delta}^2} \|\hat{w}^N\|_{Y_N}^2 \\
&\leq \frac{1}{\gamma_{N,\delta}} (e_{N,\delta} + \varepsilon_{N,\delta}) + \frac{\delta(N+1)}{2\gamma_{N,\delta}^2} + \frac{\alpha_\delta^2}{2\gamma_{N,\delta}^2} \|\hat{w}^N\|_{Y_N}^2.
\end{aligned}$$

Dividing by $N+1$, we obtain the claim. \square

The next corollary shows norm convergence under similar parameter choices as in [Theorem 3.11](#). Define

$$(3.14a) \quad \mathcal{E}_{1,\delta} := \lim_{N \rightarrow \infty} \frac{e_{N,\delta} + \varepsilon_{N,\delta}}{(N+1)\gamma_{N,\delta}} < \infty,$$

$$(3.14b) \quad \mathcal{E}_{2,\delta} := \lim_{N \rightarrow \infty} \frac{\delta}{\gamma_{N,\delta}^2} < \infty,$$

$$(3.14c) \quad \mathcal{E}_{3,\delta} := \lim_{N \rightarrow \infty} \frac{\alpha_\delta^2}{(N+1)\gamma_{N,\delta}^2} \|\hat{w}^N\|_{Y_N}^2 < \infty.$$

Corollary 3.20. *Suppose [Assumptions 3.4](#), [3.14](#) and [3.15](#) hold at some $\hat{x}^N \in X_{:N}$ for all $N \in \mathbb{N}$. If*

$$\lim_{\delta \rightarrow 0} \max\{\mathcal{E}_{1,\delta}, \mathcal{E}_{2,\delta}, \mathcal{E}_{3,\delta}\} = 0,$$

then

$$\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N+1} \|x_\delta^N - \hat{x}^N\|_{N,\circ}^2 = 0.$$

Proof. Since the threshold $\bar{\delta}$ of [Theorem 3.19](#) is independent of N , it suffices to take $N \rightarrow \infty$ in [\(3.11\)](#) divided by $N+1$ for all $\delta \in (0, \bar{\delta})$, and then let $\delta \rightarrow 0$. \square

3.4 ESTIMATES BASED ON A SEMI-STRONG SOURCE CONDITION

We now reduce the assumption of $\|\cdot\|_{N,A,\circ}$ -strong local $\varepsilon_{N,\delta}$ -subdifferentiability, to $(A_{:N}, \|\cdot\|_{N,\circ})$ -semi-strong local $\varepsilon_{N,\delta}$ -subdifferentiability. This will yield convergence to a *set* of minimal- $R_{:N}$ solutions, instead of a specific minimal- $R_{:N}$ solution. We recall that the former set $\hat{X}_{:N}$ is defined in [\(3.3\)](#). The results here, even in the static case with no N -dependence, relax the assumptions of [\[27\]](#).

Assumption 3.21 (Semi-strong source condition). Let $N \in \mathbb{N}$. We say that $\hat{x}^N \in X_{:N}$ satisfies the *semi-strong source condition*, if $X_{:N}$ is equipped with a semi-norm $\|\cdot\|_{N,\circ}$, and the following hold:

- (i) The basic source condition (3.4) holds for some $\hat{w}^N \in Y_{:N}$.
- (ii) There exist a $\delta^* > 0$ such that for all $\delta \in (0, \delta^*)$, for given $\alpha_\delta > 0$, the function $Q_{:N,\delta}$ is $(A_{:N}, \|\cdot\|_{N,\circ})$ -*semi-strongly locally* $\varepsilon_{N,\delta}$ -*subdifferentiable* (see (2.5)) at \hat{x}^N for $\hat{x}_{:N}^* := J'_{:N,\delta}(\hat{x}^N) - \alpha_\delta A'_{:N}(\hat{x}^N)^* \hat{w}^N$ with respect to $\hat{X}_{:N}$ in a neighbourhood $V_{\hat{x}^N,\delta}$ of \hat{x}^N with the factor $\gamma_{N,\delta} > 0$.

Again, we refer to [25] for examples of the satisfaction of the corresponding static condition. We discuss ways to prove the condition for the spatiotemporal regularisers of Example 3.2 in Section 4.

To pass to the limit, we also require the following relaxation of Assumption 3.15.

Assumption 3.22. We require:

- (i) Let U_ρ^N be defined in Assumption 3.15 (i). Then, there exists a $\delta^* > 0$ such that for all $\delta \in (0, \delta^*)$ and $N \in \mathbb{N}$, there exist collections $\bar{X}_{:N}^\delta \subset \hat{X}_{:N}$ of points satisfying the *semi-strong* source condition of Assumption 3.21 for the same δ^* and $\gamma_{N,\delta}$ with $\bigcup_{\hat{x}^N \in \bar{X}_{:N}^\delta} V_{\hat{x}^N,\delta} \supset U_\rho^N$.
- (ii) For $r_{N,\delta}$ defined in (3.10), we have $\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} r_{N,\delta} = \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N+1} R_{:N}(\hat{x}^N) < \infty$, as well as $\lim_{\delta \rightarrow 0} \alpha_\delta = 0$

Example 3.23. Recall Example 3.17. If, there, we only have semi-strong subdifferentiability of Ψ , then $U_\rho^N \subset \hat{X}_{:N} + B(0, r)$ for any $r > 0$ for small enough $\rho > 0$. Consequently, if there exists a set $\bar{X}_{:N}^\delta \subset \hat{X}_{:N}$ where Assumption 3.21 holds for uniform $\gamma_{N,\delta}$ and δ^* with $V_{\hat{x}^N,\delta} = B(\hat{x}^N, \varepsilon_\delta)$ and $\bar{X}_{:N}^\delta + B(0, \varepsilon_\delta) \supset \hat{X}_{:N}$ for some uniform $\varepsilon_\delta > 0$, then Assumption 3.22 (i) holds. The existence of the uniform $\varepsilon_\delta > 0$ holds, for instance, when $\bar{X}_{:N}^\delta$ is finite, or when $\hat{X}_{:N}$ is compact and all $\hat{x}^N \in \hat{X}_{:N}$ satisfy Assumption 3.21.

We denote the set distance with respect to $\|\cdot\|_{N,\circ}$ by

$$(3.15) \quad \text{dist}_{N,\circ}(x^N, \hat{X}_{:N}) := \inf_{\hat{x}^N \in \hat{X}_{:N}} \|\hat{x}^N - x^N\|_{N,\circ} \quad \text{for any } \hat{X}_{:N} \subset X_{:N}.$$

Then we have the following initial estimate.

Lemma 3.24. *Let $N \in \mathbb{N}$. Suppose Assumptions 3.4 and 3.21 hold at some $\hat{x}^N \in \hat{X}_{:N}$. Let $\delta \in (0, \delta^*)$. If $x_\delta^N \in V_{\hat{x}^N,\delta}$, then*

$$\text{dist}_{N,\circ}^2(x_\delta^N, \hat{X}_{:N}) \leq \frac{\varepsilon_{N,\delta} + \varepsilon_{N,\delta}}{(N+1)\gamma_{N,\delta}} + \frac{\delta}{2\gamma_{N,\delta}^2} + \frac{\alpha_\delta^2}{2(N+1)\gamma_{N,\delta}^2} \|\hat{w}^N\|_{Y_{:N}}^2.$$

Proof. As in (3.12), using the assumed $(A_{:N}, \|\cdot\|_{N,\circ})$ -strong local $\varepsilon_{N,\delta}$ -subdifferentiability of $Q_{:N,\delta}$ (Assumption 3.21), we estimate

$$\begin{aligned} \varepsilon_{N,\delta} + \varepsilon_{N,\delta} &\stackrel{(3.6)}{\geq} \left(J_{:N,\delta}(x_\delta^N) + \alpha_\delta R_{:N}(x_\delta^N) \right) - \left(J_{:N,\delta}(\hat{x}^N) + \alpha_\delta R_{:N}(\hat{x}^N) \right) + \varepsilon_{N,\delta} \\ &\stackrel{(2.5)}{\geq} \left(\langle J'_{:N,\delta}(\hat{x}^N) - \alpha_\delta A'_{:N}(\hat{x}^N)^* \hat{w}^N | x_\delta^N - \hat{x}^N \rangle_{X_{:N}^*, X_{:N}} + \gamma_{N,\delta} \|A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N)\|_{Y_{:N}}^2 \right) \\ &\quad + \gamma_{N,\delta} \text{dist}_{N,\circ}^2(x_\delta^N, \hat{X}_{:N}) \\ &= \left(\langle \ell'_{:N}(\hat{b}^N - b_\delta^N) - \alpha_\delta \hat{w}^N, A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N) \rangle_{Y_{:N}} + \gamma_{N,\delta} \|A'_{:N}(\hat{x}^N)(x_\delta^N - \hat{x}^N)\|_{Y_{:N}}^2 \right) \\ &\quad + \gamma_{N,\delta} \text{dist}_{N,\circ}^2(x_\delta^N, \hat{X}_{:N}) \\ &\geq -\frac{1}{4\gamma_{N,\delta}} \|\ell'_{:N}(\hat{b}^N - b_\delta^N) - \alpha_\delta \hat{w}^N\|_{Y_{:N}}^2 + \gamma_{N,\delta} \text{dist}_{N,\circ}^2(x_\delta^N, \hat{X}_{:N}). \end{aligned}$$

Now estimating as in (3.13) yields the claim. \square

Theorem 3.25. *Suppose Assumptions 3.4 and 3.22 hold. If, for $\mathcal{E}_{i,\delta}$ defined in (3.14),*

$$(3.16) \quad \lim_{\delta \searrow 0} \max\{\mathcal{E}_{1,\delta}, \mathcal{E}_{2,\delta}, \mathcal{E}_{3,\delta}\} = 0,$$

then

$$\lim_{\delta \searrow 0} \lim_{N \rightarrow \infty} \frac{1}{N+1} \text{dist}_{N,\circ}(x_\delta^N, \hat{X}_{:N}) = 0.$$

Proof. Suppose, to reach a contradiction, that there exist $\varepsilon > 0$ and a sequence $\delta_j \searrow 0$ such that for every $j \in \mathbb{N}$ there exists a sequence $\{N_{j,k}\}_{k \in \mathbb{N}}$, $\lim_{k \rightarrow \infty} N_{j,k} = \infty$, satisfying

$$\frac{1}{N_{j,k} + 1} \text{dist}_{N,\circ}(x_{\delta_j}^{N_{j,k}}, \hat{X}^{N_{j,k}}) \geq \varepsilon.$$

We may assume that $\delta_{j_0} \leq \delta^*$ for δ^* given by Assumption 3.21(ii). By Lemma 3.13 and Assumption 3.22(ii), there exists $j_0 \in \mathbb{N}$ and for all $j \geq j_0$ an index $k(j) \in \mathbb{N}$ such that $x_{\delta_j}^{N_{j,k}} \in U_\rho^{N_{j,k}}$ for $k \geq k(j)$. By Assumption 3.22(i), we have $x_{\delta_j}^{N_{j,k}} \in V_{\bar{x}, \delta_j}$ for some $\bar{x} \in \bar{X}_{:N_{j,k}}^{\delta_j}$, for all $k \geq k(j)$ and $j \geq j_0$. By Lemma 3.24 and (3.16), we then find some $n(j) \geq k(j)$ that satisfies the contradiction

$$\lim_{j \rightarrow \infty} \frac{1}{N_{j,n(j)} + 1} \text{dist}_{N,\circ}(\hat{X}^{N_{j,n(j)}}, x_{\delta_j}^{N_{j,n(j)}}) = 0. \quad \square$$

4 STRONG LOCAL SUBDIFFERENTIABILITY AND INFIMAL CONVOLUTIONS

We now provide tools to verify the subregularity assumptions of Assumptions 3.15 and 3.21 when R_N arises from the algorithms of [11, 18] as discussed in Example 3.2. In particular, we show in Section 4.1 that R_N is a proper infimal convolutions, not merely a “sub-infimal convolution”. There we also provide several results of general interest in the analysis of infimal convolutions. Then in Section 4.2 we treat ways to obtain the required subregularity. Through out this section, X is a normed space, unless otherwise indicated.

4.1 BASIC PROPERTIES OF INFIMAL CONVOLUTIONS

We recall that the infimal convolution $G \square H$ of two functions $G, H : X \rightarrow \overline{\mathbb{R}}$ is defined as

$$[G \square H](x) := \inf_{\tilde{x} \in X} (G(\tilde{x}) + H(x - \tilde{x})).$$

In analogy with the Fenchel conjugate (1.5), we also define for $E : X^* \rightarrow \overline{\mathbb{R}}$ the *Fenchel preconjugate* $E_* : X \rightarrow \overline{\mathbb{R}}$ through

$$E_*(x) := \sup_{x^* \in X^*} (\langle x^* | x \rangle_{X^*, X} - E(x^*)).$$

Then the *biconjugate*

$$F^{**} := (F^*)_* : X \rightarrow \overline{\mathbb{R}}.$$

A well-known conjugate formula establishes $[G \square H]^* = G^* + H^*$; in finite dimensions, see, e.g., [21, Theorem 11.23]. It then follows that $[G \square H]^{**} = [G^* + H^*]_*$. Now, if $G \square H$ is convex, proper, and lower semicontinuous, we have $[G \square H]^{**} = G \square H$ [9, Theorem 5.2], so obtain $G \square H = [G^* + H^*]_*$. It is easy to see that $G \square H$ is convex and proper if G and F are, however, the lower semicontinuity requires further assumptions. See [21, Theorem 11.23] in finite dimensions, and [4, Proposition 12.14] in Hilbert spaces. Since the result is not easily found in the literature in normed spaces, we include it here in full together with assumptions that are practical in our applications.

Lemma 4.1. *Let $G, H : X \rightarrow \overline{\mathbb{R}}$ be convex and proper. Then $G \square H$ is convex and proper.*

If, moreover, H or G is Lipschitz continuous, then $G \square H$ is also lower semicontinuous; in consequence, $G \square H = [G^ + H^*]_*$.*

Proof. By assumption, there exist some $\tilde{x} \in \text{dom } G$ and $x \in \text{dom } H$. Now, $x + \tilde{x} \in \text{dom}[G \square H]$, proving properness. Let then $\Psi(\tilde{x}, x) := G(\tilde{x}) + H(x - \tilde{x})$ and $x_1, x_2 \in X$ and $\lambda \in [0, 1]$. The function Ψ is convex, and there exist sequences $\{z_n^1\}_{n \in \mathbb{N}}$ and $\{z_n^2\}_{n \in \mathbb{N}} \subset X$ with $\Psi(x_1, z_n^1) \rightarrow [G \square H](x_1)$ and $\Psi(x_2, z_n^2) \rightarrow [G \square H](x_2)$. Convexity now follows from passing to the limit in

$$[G \square H](\lambda x_1 + (1 - \lambda)x_2) \leq \Psi(\lambda(x_1, z_n^1) + (1 - \lambda)(x_2, z_n^2)) \leq \lambda \Psi(x_1, z_n^1) + (1 - \lambda)\Psi(x_2, z_n^2).$$

Finally, we need to show that $[G \square H](x) \leq \liminf_{n \rightarrow \infty} [G \square H](x_n)$ for all converging sequences $x_n \rightarrow x$. Because $G \square H = H \square G$, we only need to consider the case that H is Lipschitz continuous. Suppose, to reach a contradiction, that there exist $\varepsilon \geq 0$ and $x_n \rightarrow x$ such that $[G \circ H](x) \geq [G \circ H](x_n) + \varepsilon$ with $[G \circ H](x_n) < \infty$. By the definition of the infimal convolution, there then exists $z_n \in X$ such that $[G \circ H](x_n) \geq \Psi(x_n, z_n) - 1/n$. It follows that

$$[G \circ H](x) \geq \Psi(x_n, z_n) + \varepsilon - 1/n = G(z_n) + H(x_n - z_n) + \varepsilon - 1/n.$$

Since H is L -Lipschitz for some L , it follows that

$$[G \circ H](x) \geq G(z_n) + H(x - z_n) - L\|x - x_n\| + \varepsilon - 1/n = \Psi(x, z_n) - L\|x - x_n\| + \varepsilon - 1/n.$$

Thus there exists $N \in \mathbb{N}$ such that for $n \geq N$, we have $[G \circ H](x) \geq \Psi(x, z_n) + \varepsilon/2 \geq [G \circ H](x) + \varepsilon/2$. This is the desired contradiction. \square

Our first contribution in this section shows that the spatiotemporal “sub-infimal convolutions” of [Example 3.2](#) reduce to infimal convolutions with respect to the support function of the convex hull of the dual dynamic constraint set.

Lemma 4.2. *Suppose the assumptions of [Lemma 4.1](#) hold. Let $F = (G^* + \delta_U)_*$ for a convex, proper, and lower semicontinuous $G : X \rightarrow \overline{\mathbb{R}}$, and $\emptyset \neq U \subset X^*$. Then $F = [G \square \delta_{\text{cl conv } U}^*]^{**}$, where $\text{cl conv } U$ is the closed convex hull of U .*

If $G \square \delta_{\text{cl conv } U}^$ is lower semicontinuous (e.g., the full conditions of [Lemma 4.1](#) hold for G and $H = \delta_{\text{cl conv } U}^*$) then $F = G \square \delta_{\text{cl conv } U}^*$.*

Proof. The supremum of a lower semicontinuous concave function over a possibly nonconvex set equals its supremum over the closed convex hull of that set. Therefore,

$$\begin{aligned} (G^* + \delta_U)^*(x) &= \sup_{x^* \in X^*} (\langle x^* | x \rangle - G^*(x^*) - \delta_U(x^*)) \\ &= \sup_{x^* \in X^*} (\langle x^* | x \rangle - G^*(x^*) - \delta_{\text{cl conv } U}(x^*)) \\ &= [G^* + \delta_{\text{cl conv } U}]_*(x) = [G \square \delta_{\text{cl conv } U}^*]^{**}(x). \end{aligned}$$

By [Lemma 4.1](#), $G \square \delta_{\text{cl conv } U}^*$ is convex and proper. When it is also lower semicontinuous, it equals its biconjugate [[9](#), Theorem 5.2]. \square

Corollary 4.3. *Let $F = (\delta_B + \delta_U)_*$ for a non-empty $U \subset X^*$ and a closed, convex, non-empty $B \subset X^*$. Then $F = \delta_{B \cap \text{cl conv } U}^*$.*

Proof. We have $\delta_B + \delta_{\text{cl conv } U} = \delta_{B \cap \text{cl conv } U}$, which is lower semicontinuous by the closedness of $B \cap \text{cl conv } U$. The claim now follows from [Lemma 4.2](#) with $G = (\delta_B)_*$. \square

Example 4.4. In Example 3.2, suppose the $G = T_{:N}$ is Lipschitz continuous, and $H = \delta_{\text{cl conv } \mathcal{Y}_{:N}}^*$. Then $\overset{\circ}{T}_{:N} = T_{:N} \square \delta_{\text{cl conv } \mathcal{Y}_{:N}}^*$, hence

$$R_{:N}(x^N) = [T_{:N} \square \delta_{\text{cl conv } \mathcal{Y}_{:N}}^*](K_{:N}x^N).$$

In particular, with Z a normed space, let

$$T_{:N} : (Z^*)^N \rightarrow \mathbb{R}, \quad T_{:N}(z^N) = \sum_{k=0}^N \|z_k^N\|_{\bullet} = (\delta_{B^{N+1}})_*(z^N)$$

for some norm $\|\cdot\|_{\bullet} = (\delta_B)_*$ in Z^* with the corresponding dual unit ball B . Then

$$R_{:N}(x^N) = \delta_{B^{N+1} \cap \text{cl conv } \mathcal{Y}_{:N}}^*(K_{:N}x^N) = \sup_{y^N \in B^{N+1} \cap \text{cl conv } \mathcal{Y}_{:N}} \langle y^N, K_{:N}x^N \rangle.$$

If $\text{cl conv } \mathcal{Y}_{:N} \supset B^{N+1}$, we obtain $R_{:N}(x^N) = \sum_{k=0}^N \|K_k x_k^N\|_{\bullet}$. When we take $K_k = \nabla$ to model total variation, this says that the spatiotemporal regulariser $R_{:N}$ still penalises the gradients, but the dual temporal dynamics $\text{cl conv } \mathcal{Y}_{:N}$ restrict the penalisation amount.

We also require a characterisation of the subdifferentials of the infimal convolution. In Hilbert spaces, the following result can be found in [4, Proposition 16.48].

Lemma 4.5. Let $F = G \square H$ for a convex, proper, and lower semicontinuous $G, H : X \rightarrow \overline{\mathbb{R}}$. Then

$$\begin{aligned} \partial F(x) &\supset \{x^* \in X^* \mid x^* \in \partial G(\tilde{x}) \cap \partial H(x - \tilde{x}) \text{ for some } \tilde{x} \in X\} \\ &= \{x^* \in X^* \mid x^* \in \partial G(\tilde{x}) \cap \partial H(x - \tilde{x}), F(x) = G(\tilde{x}) + H(x - \tilde{x}) \text{ for some } \tilde{x} \in X\}. \end{aligned}$$

Equality holds when there exists some $x_0^* \in \text{int dom } G^* \cap \text{dom } H^*$.

Proof. By the Fenchel–Young theorem [9, Lemma 5.9], $x^* \in \partial F(x)$ if and only if $x \in \partial F^*(x^*)$. We have $F^* = G^* + H^*$. By the sum rule for convex subdifferentials, $\partial F^*(x^*) \supset \partial G^*(x^*) + \partial H^*(x^*)$. Therefore $x^* \in \partial F(x)$ follows from $x \in \partial G^*(x^*) + \partial H^*(x^*)$, i.e., $\tilde{x} \in \partial G^*(x^*)$ and $x - \tilde{x} \in \partial H^*(x^*)$ for some \tilde{x} , i.e., $x^* \in \partial G(\tilde{x}) \cap \partial H(x - \tilde{x})$. This proves the inclusion.

The sum rule holds with equality when there exists some $x_0^* \in \text{int dom } G^* \cap \text{dom } H^*$. In that case the opposite implication also holds. \square

It is not difficult to see that the infimal convolution of semi-norms is a semi-norm. In fact, we can also square-and-root them:

Lemma 4.6. Let $G, H : X \rightarrow \overline{\mathbb{R}}$ be semi-norms. Write $G^2(x) := G(x)^2$. Then $F(x) := \sqrt{[G^2 \square H^2](x)}$ is a semi-norm.

Proof. The definition of the infimal convolution immediately establishes non-negativity. For any $x, z \in X$ and $\lambda > 0$, since F and G are positively homogeneous and satisfy the triangle inequality,

$$\begin{aligned} [G^2 \square H^2](x + \lambda z) &= \inf_{\tilde{x}, \tilde{z} \in X} (G(\tilde{x} + \lambda \tilde{z})^2 + H(\tilde{x} + \lambda \tilde{z} - x - \lambda z)^2) \\ &\leq \inf_{\tilde{x}, \tilde{z} \in X} ([G(\tilde{x}) + \lambda G(\tilde{z})]^2 + [H(\tilde{x} - x) + \lambda H(\tilde{z} - z)]^2). \end{aligned}$$

Hence, by the triangle inequality in \mathbb{R}^2 ($\|a + b\|_2 \leq \|a\|_2 + \|b\|_2$),

$$\begin{aligned} \sqrt{[G^2 \square H^2](x + z)} &\leq \inf_{\tilde{x}, \tilde{z} \in X} \left(\sqrt{G(\tilde{x})^2 + H(\tilde{x} - x)^2} + \lambda \sqrt{G(\tilde{z})^2 + H(\tilde{z} - z)^2} \right) \\ &= \sqrt{[G^2 \square H^2](x)} + \lambda \sqrt{[G^2 \square H^2](z)}. \end{aligned}$$

This establishes both the triangle inequality and positive homogeneity. \square

4.2 STRONG LOCAL SUBDIFFERENTIABILITY

Recall the concepts (2.3) and (2.4) of strong and semi-strong local subdifferentiability.

Lemma 4.7. *Let $F = G \square H$ for convex, proper, and lower semicontinuous $G, H : X \rightarrow \overline{\mathbb{R}}$ such that the full assumptions of Lemma 4.1 hold. Suppose*

$$S(x) := \{\tilde{x} \in X \mid \partial G(\tilde{x}) \cap \partial H(x - \tilde{x}) \neq \emptyset\}$$

satisfies for a given $\hat{x} \in S(\hat{x})$ and some $\varepsilon, \rho > 0$ the continuity property

$$(4.1) \quad \text{dist}(\tilde{x}, S(x)) \leq \rho \quad \text{for all } x \in B(\hat{x}, \varepsilon).$$

Let $\hat{x}^ \in \partial F(\hat{x})$, and assume further that, for semi-norms $\|\cdot\|_\diamond$ and $\|\cdot\|_\Delta$ on X ,*

(i) *G is $\|\cdot\|_\diamond$ -(semi)strongly locally subdifferentiable with the factor $\gamma > 0$ at \tilde{x} for \hat{x}^* in $B(\hat{x}, \rho)$, and*

(ii) *H is $\|\cdot\|_\Delta$ -(semi)strongly locally subdifferentiable with the factor γ at $\hat{x} - \tilde{x}$ for \hat{x}^* in $B(\hat{x} - \tilde{x}, \rho)$.*

Recalling Lemma 4.6, let $\|\cdot\|_\square := \sqrt{\|\cdot\|_\diamond^2 \square \|\cdot\|_\Delta^2}$. Then $\hat{x}^ \in \partial F(\hat{x})$ and F is $\|\cdot\|_\square$ -(semi)strongly locally subdifferentiable at \hat{x} for \hat{x}^* with the factor γ in $B(\hat{x}, \min\{\varepsilon, 2\rho\})$.*

Proof. We first treat strong local subdifferentiability. Let $x \in B(\hat{x}, \min\{\varepsilon, 2\rho\})$. We have $[G \square H](\hat{x}) = G(\tilde{x}) + H(\hat{x} - \tilde{x})$ for $\tilde{x} \in S(\hat{x})$. Also using the definition of the infimal convolution, it follows that

$$\begin{aligned} F(x) - F(\hat{x}) - \langle \hat{x}^* | x - \hat{x} \rangle_{X^*, X} &= \inf_{x' \in X} (G(x') + H(x - x') - G(\tilde{x}) - H(\hat{x} - \tilde{x})) \\ &\quad - \langle \hat{x}^* | (x' - \tilde{x}) + (x - x' - (\hat{x} - \tilde{x})) \rangle_{X^*, X}. \end{aligned}$$

Due to (4.1) and $x \in B(\hat{x}, \varepsilon)$, there exists a $x' \in B(\tilde{x}, \rho) \cap S(x)$, so we can use the assumed strong local metric subdifferentiabilities to obtain, as required

$$\begin{aligned} F(x) - F(\hat{x}) - \langle \hat{x}^* | x - \hat{x} \rangle_{X^*, X} &\geq \inf_{x' \in B(\tilde{x}, \rho)} (\gamma \|x' - \tilde{x}\|_\diamond^2 + \gamma \|(x - x') - (\hat{x} - \tilde{x})\|_\Delta^2) \\ &\geq \inf_{x' \in X} (\gamma \|x' - \tilde{x}\|_\diamond^2 + \gamma \|(x - \hat{x}) - (x' - \tilde{x})\|_\Delta^2) = \gamma \|x - \hat{x}\|_\square^2. \end{aligned}$$

For semi-strong local subdifferentiability, we similarly have

$$\begin{aligned} F(x) - F(\hat{x}) - \langle \hat{x}^* | x - \hat{x} \rangle &\geq \inf_{x' \in X} (\gamma \text{dist}^2(x', [\partial G]^{-1}(\hat{x}^*)) + \gamma \text{dist}^2(x - x', [\partial H]^{-1}(\hat{x}^*))) \\ &= \inf \left\{ \inf_{x' \in X} (\gamma \|x' - \tilde{x}\|_\diamond^2 + \gamma \|x - x' - \bar{x}\|_\Delta^2) \mid \hat{x}^* \in \partial G(\tilde{x}) \cap \partial H(\bar{x}); \tilde{x}, \bar{x} \in X \right\} \\ &= \inf \left\{ \inf_{x' \in X} (\gamma \|x' - \tilde{x}\|_\diamond^2 + \gamma \|x - \bar{x} - (x' - \tilde{x})\|_\Delta^2) \mid \hat{x}^* \in \partial G(\tilde{x}) \cap \partial H(\bar{x} - \tilde{x}); \tilde{x}, \bar{x} \in X \right\} \\ &= \inf \{ \gamma \|x - \bar{x}\|_\square^2 \mid \hat{x}^* \in \partial G(\tilde{x}) \cap \partial H(\bar{x} - \tilde{x}); \tilde{x}, \bar{x} \in X \}. \end{aligned}$$

On the other hand, Lemma 4.5 establishes that $\partial G(\tilde{x}) \cap \partial H(\bar{x} - \tilde{x}) \subset \partial F(\bar{x})$. It follows that

$$F(x) - F(\hat{x}) - \langle \hat{x}^* | x - \hat{x} \rangle \geq \inf \{ \gamma \|x - \bar{x}\|_\square^2 \mid \hat{x}^* \in \partial F(\bar{x}); \bar{x} \in X \} = \gamma \text{dist}_\square^2(x, [\partial F]^{-1}(\hat{x}^*)),$$

where dist_\square is the set-distance generated by $\|\cdot\|_\square$. □

Remark 4.8. The continuity property (4.1) holds if S has the Aubin property (see [9, Chapter 27]) at \hat{x} for \tilde{x} in $B(\hat{x}, \varepsilon)$ with modulus $\kappa > 0$. Indeed, then

$$\text{dist}(\tilde{x}, S(x)) \leq \kappa \text{dist}(S^{-1}(\tilde{x}), x) \leq \kappa \|\hat{x} - x\| \leq \kappa \varepsilon = \rho \quad \text{for all } x \in B(\hat{x}, \varepsilon).$$

By [9, Theorem 28.3], if $x^* \in \partial G(\tilde{x}) \cap \partial H(\hat{x} - \tilde{x})$, and

$$0 \in D^*[\partial G](\tilde{x}|x^*)(x^{**}) + p \text{ and } p \in D^*[\partial H](\hat{x} - \tilde{x}|x^*)(x^{**}) \implies p = 0, x^{**} = 0$$

then S has the Aubin property at x for x^* .

Example 4.9 (Growth properties of spatiotemporal total variation). Let $R_{:N}$ be formed from a spatial total variation regulariser as in [Examples 3.2](#) and [4.4](#). Suppose

$$G = T_{:N}, \quad T_{:N}(z^N) = \sum_{k=0}^N \|z_k^N\|_{\bullet}, \quad \text{and} \quad H = \delta_{\text{cl conv } \mathcal{Y}_{:N}}^* = (\delta_{B^{N+1}})_*(z^N)$$

satisfy the assumptions of [Lemma 4.7](#). Then $T_{:N} \square \delta_{\text{cl conv } \mathcal{Y}_{:N}}^*$ is $\|\cdot\|_{\square}$ -(semi-)strongly locally subdifferentiable; consequently $R_{:N}(x^N) = [T_{:N} \square \delta_{\text{cl conv } \mathcal{Y}_{:N}}^*](K_{:N}x^N)$ is $\|\cdot\|_{N,\circ}$ -(semi-)strongly locally subdifferentiable for $\|x^N\|_{N,\circ} := \|K_{:N}x^N\|_{\square}$.

The semi-norm $\|\cdot\|_{\diamond}$ may be zero on a very large subspace: it will generally only be non-zero when $z_k^N(\xi) = 0$ for a spatial point ξ , and the corresponding subdifferential is in the interior of the dual ball; compare the analysis of both the Lasso and spatial total variation in [[27](#), Section 4]. Likewise, a small $\mathcal{Y}_{:N}$ may produce $\|\cdot\|_{\square}$ that provides growth only on a very small subspace. We, therefore, have to compensate for this weak growth estimate with the data term.

Suppose now that the basic source condition of [Assumption 3.3](#) holds for \hat{x}^N . To verify the strong or semi-strong source conditions [Assumptions 3.14](#) and [3.21](#), we need to verify the respective (semi-)strong local subdifferentiability of $Q_{:N,\delta} = J_{:N,\delta} + \alpha_{\delta}R_{:N}$ at \hat{x}^N for $\hat{x}_{:N}^* := J'_{:N,\delta}(\hat{x}^N) - \alpha_{\delta}A'_{:N}(\hat{x}^N)\hat{w}^N$. Recall that

$$J_{:N,\delta}(x) = \ell_{:N}(A_{:N}(x) - b_{\delta}^N).$$

Lemma 4.10. *Suppose [Assumption 3.4](#) holds, and that*

(i) $\|[A'_{:N}(\zeta) - A'_{:N}(\hat{x})]h\| \leq \beta\|A'_{:N}(\hat{x})h\|$ for all h and for all $\zeta \in U$ for some $\beta > 0$ and a neighbourhood U of \hat{x}^N .

(ii) $\ell_{:N}$ is γ -strongly convex.

Then, for any $\eta > 0$ and $x^N \in U$,

$$(4.2) \quad J_{:N,\delta}(x^N) - J_{:N,\delta}(\hat{x}^N) \geq \langle J'_{:N,\delta}(\hat{x}^N)|x^N - \hat{x}^N\rangle - \delta(N+1)\eta^{-1} + \frac{\gamma - \eta\beta^2}{2}\|A'_{:N}(\hat{x}^N)(x^N - \hat{x}^N)\|^2.$$

Proof. Using the γ -strong convexity of $\ell_{:N}$ and the mean value theorem, for some $\zeta \in [\hat{x}^N, x^N] \subset U$, we have

$$\begin{aligned} J_{:N,\delta}(x^N) - J_{:N,\delta}(\hat{x}^N) &= \ell_{:N}(A_{:N}(x^N) - b_{\delta}^N) - \ell_{:N}(A_{:N}(\hat{x}^N) - b_{\delta}^N) \\ &= \ell_{:N}(A_{:N}(\hat{x}^N) - b_{\delta}^N + A'_{:N}(\zeta)(x^N - \hat{x}^N)) - \ell_{:N}(A_{:N}(\hat{x}^N) - b_{\delta}^N) \\ &\geq \langle \nabla \ell_{:N}(A_{:N}(\hat{x}^N) - b_{\delta}^N), A'_{:N}(\zeta)(x^N - \hat{x}^N) \rangle + \frac{\gamma}{2}\|A'_{:N}(\zeta)(x^N - \hat{x}^N)\|^2 \\ &= \langle J'_{:N,\delta}(\hat{x}^N)|x^N - \hat{x}^N\rangle + \langle \nabla \ell_{:N}(\hat{b}^N - b_{\delta}^N), [A'_{:N}(\zeta) - A'_{:N}(\hat{x}^N)](x^N - \hat{x}^N) \rangle \\ &\quad + \frac{\gamma}{2}\|A'_{:N}(\hat{x}^N)(x^N - \hat{x}^N)\|^2. \end{aligned}$$

The claim now follows using Young's inequality, (i), and [Assumption 3.4](#) (i). \square

Take $0 < \eta < \gamma/\beta^2$. Suppose $R_{:N}$ is $\|\cdot\|_{N,\circ}$ -(semi-)strongly locally subdifferentiable with the factor $\gamma^R > 0$ (independent of N) at \hat{x} for $-\alpha_{\delta}A'_{:N}(\hat{x}^N)\hat{w}^N$, that is

$$(4.3) \quad R_{:N}(x^N) - R_{:N}(\hat{x}^N) \geq -\langle \alpha_{\delta}A'_{:N}(\hat{x}^N)\hat{w}^N |x^N - \hat{x}^N\rangle + \gamma^R\|x^N - \hat{x}^N\|_{N,\circ}^2.$$

or

$$(4.4) \quad R_{:N}(x^N) - R_{:N}(\hat{x}^N) \geq -\langle \alpha_{\delta}A'_{:N}(\hat{x}^N)\hat{w}^N |x^N - \hat{x}^N\rangle + \gamma^R \text{dist}_{N,\circ}^2(x^N, \hat{X}_{:N})$$

These estimates can in applicable cases be verified with Lemma 4.7; compare Example 4.9. We can then by summing (4.3) (resp. (4.4)) with (4.2) obtain Assumption 3.14 (resp. Assumption 3.21) with

$$\varepsilon_{N,\delta} = \delta(N+1)\eta^{-1} \quad \text{and, independent of } N, \quad \gamma_{N,\delta} = \gamma := \min\{\gamma - \eta\beta^2, \gamma^R\}/2.$$

That is, for $v^N := J'_{:,N,\delta}(\hat{x}^N) - \alpha_\delta A'_{:,N}(\hat{x}^N)\hat{w}^N$, we have in the strong case,

$$(4.5) \quad \begin{aligned} Q_{:,N,\delta}(x^N) - Q_{:,N,\delta}(\hat{x}^N) &\geq \langle v^N | x^N - \hat{x}^N \rangle + \frac{\gamma - \eta\beta^2}{2} \|A'_{:,N}(\hat{x}^N)(x^N - \hat{x}^N)\|^2 + \frac{\gamma^R}{2} \|x^N - \hat{x}^N\|_{N,\circ}^2 - \varepsilon_{N,\delta} \\ &\geq \langle v^N | x^N - \hat{x}^N \rangle + \gamma_{N,\delta} \left(\|A'_{:,N}(\hat{x}^N)(x^N - \hat{x}^N)\|^2 + \|x^N - \hat{x}^N\|_{N,\circ}^2 \right) - \varepsilon_{N,\delta} \end{aligned}$$

and, likewise, in the semi-strong case,

$$(4.6) \quad \begin{aligned} Q_{:,N,\delta}(x^N) - Q_{:,N,\delta}(\hat{x}^N) &\geq \langle v^N | x^N - \hat{x}^N \rangle + \gamma_{N,\delta} \|A'_{:,N}(\hat{x}^N)(x^N - \hat{x}^N)\|^2 + \gamma_{N,\delta} \text{dist}_{N,\circ}^2(x^N, \hat{X}_{:,N}) - \varepsilon_{N,\delta}. \end{aligned}$$

Recalling the discussion in Example 4.9, in the case of the strong source condition of Assumption 3.14, a part of $\|A'_{:,N}(\hat{x}^N)(x^N - \hat{x}^N)\|^2$ in (4.5) can be fused with $\|x^N - \hat{x}^N\|_{N,\circ}^2$ to improve the latter semi-norm. Indeed, compare the treatment of spatial total variation in [27, Section 4.3]. To do the same with (4.6) and the semi-strong source condition of Assumption 3.21 requires further assumptions. In particular, if $A_{:,N}$ is linear then the value of

$$Q_{:,N,\delta}(x^N) - Q_{:,N,\delta}(\hat{x}^N) - \langle v^N | x^N - \hat{x}^N \rangle$$

is independent of the choice of $\hat{x}^N \in \hat{X}_{:,N}$. Therefore, in this case, we can take the supremum over $\hat{x}^N \in \hat{X}_{:,N}$ in (4.6), and use $\sup A + \inf B \geq \inf(A + B)$ to improve the $\text{dist}_{N,\circ}$.

By Examples 3.17 and 3.23, the containment and covering properties of Assumption 3.15 (i) and Assumption 3.22 (i) can be studied using the very same techniques that we just applied to $Q_{:,N,\delta}$.

Recalling (3.14), we now get

$$\mathcal{E}_{1,\delta} = \lim_{N \rightarrow \infty} \frac{e_{N,\delta}}{(N+1)\gamma} + \frac{\delta}{\eta\gamma}.$$

The condition $\lim_{\delta \rightarrow 0} \mathcal{E}_{1,\delta} = 0$ of Theorem 3.25 and Corollary 3.20 can therefore be ensured by controlling the algorithm errors $e_{N,\delta}$. The condition $\lim_{\delta \rightarrow 0} \mathcal{E}_{2,\delta} = 0$ automatically holds from the definition (3.14), while $\lim_{\delta \rightarrow 0} \mathcal{E}_{3,\delta} = 0$ requires $\|\hat{w}^N\|_{Y_{:,N}}^2/(N+1)$ to be bounded and $\alpha_\delta \rightarrow 0$. To ensure the limiting properties of $r_{N,\delta}$ defined in (3.10), and as required by Assumption 3.15 (ii), we further require the relative rate control

$$\lim_{\delta \rightarrow 0} \sup_{N \in \mathbb{N}} \left(\frac{\delta^q + \delta}{\alpha_\delta} + \frac{e_{N,\delta}}{\alpha_\delta(N+1)} \right) = 0.$$

For Assumption 3.22 (ii), the supremum becomes a limit.

5 NUMERICAL ILLUSTRATION

In this section, we numerically verify the regularisation theory of Section 3 in the context of the dynamic EIT problem from the introduction. In the latter, the goal is to reconstruct a time-varying conductivity distribution containing moving (and occasionally disappearing) inclusions. The evolution of the inclusions is governed by a transport equation corresponding to incompressible flow with constant density.

The numerical experiments, including the EIT forward model, data generation, discretisation choices, and noise model, closely follow the setup introduced in [11]. The algorithmic setup follows our companion paper [18]. We therefore only summarise the key ingredients here and refer the reader to those works for the full details. Our Julia implementation of the experiments and algorithms is available on Zenodo [16].

5.1 MATHEMATICAL MODEL OF EIT

On time step k , the forward problem of EIT is governed by the Complete Electrode Model (CEM). The objective is to reconstruct an electric conductivity $\sigma^k \in L^\infty(\Omega)$ inside the measurement domain $\Omega \subset \mathbb{R}^d$ from noisy measurements $\mathcal{J}_\delta^{j,k}$ of electrode currents $I^{j,k} = I(\sigma^k, U^{j,k}) \in \mathbb{R}^{N_{\text{elec}}}$ at $N_{\text{elec}} \in \mathbb{N}$ electrodes over $j = 1, \dots, N_{\text{pat}}$ different potential patterns $U^{j,k} = (U_1^{j,k}, \dots, U_{N_{\text{elec}}}^{j,k}) \in \mathbb{R}^{N_{\text{elec}}}$ at the same electrodes. We enforce the uniform bounds $0 < \sigma_m \leq \sigma^k \leq \sigma_M < \infty$ a.e. in the imaging domain Ω . Then the physics are governed by the PDE

$$\begin{aligned}
 (5.1a) \quad & \nabla \cdot (\sigma^k \nabla u^{j,k}) = 0 \quad \text{in } \Omega, \\
 (5.1b) \quad & u^{j,k} + \zeta_i \sigma^k \nabla u^{j,k} \cdot \nu = U_i^{j,k} \quad \text{on } \partial\Omega_{e_i}, \\
 (5.1c) \quad & \sigma^k \nabla u^{j,k} \cdot \nu = 0 \quad \text{on } \partial\Omega \setminus \cup_i \partial\Omega_{e_i}, \\
 (5.1d) \quad & \int_{\partial\Omega_{e_i}} \sigma^k \nabla u^{j,k} \cdot \nu \, dS = -I_i^{j,k} \quad \text{for } i = 1, \dots, N_{\text{elec}} \text{ and } j = 1, \dots, N_{\text{pat}}.
 \end{aligned}$$

Here $\partial\Omega_{e_i}$ are the electrode surfaces, $\zeta_i > 0$ are contact impedances, and ν is the outward unit normal. This potential-to-current formulation reflects modern EIT measurement devices [17]. In practice, only noisy measurements $\mathcal{J}^{j,k}$ of these currents are available, and the inverse problem is to recover σ^k from the measurements $\mathcal{J}^{j,k}$. For a single time index k , this is commonly formulated as the total variation regularised optimisation problem

$$(5.2) \quad \min_{\sigma \in \text{BV}(\Omega) \cap L^\infty(\Omega)} \frac{1}{2} \sum_{j=1}^{N_{\text{pat}}} \|\Sigma^{-1/2} (I(\sigma, U^{j,k}) - \mathcal{J}_\delta^{j,k})\|_2^2 + \delta_{[\sigma_m, \sigma_M]}(\sigma) + \alpha_\delta \text{TV}(\sigma).$$

where $0 < \sigma_m < \sigma_M$ enforce physical bounds, $\Sigma^{-1/2}$ is a data precision matrix, and TV promotes structured sparsity. Similar formulations have been studied in [28, 29, 19]. In the dynamic framework of Section 3, following Example 3.1, we would in the spaces $X_{:N} = (\text{BV}(\Omega) \cap L^\infty(\Omega))^{N+1}$ and $Y_{:N} = (\mathbb{R}^{N_{\text{elec}} \times N_{\text{pat}}})^{N+1}$ take

$$\ell_{:N}(y^N) = \sum_{k=0}^{N-1} \|\Sigma^{-1/2} y_k^N\|_2^2, \quad b_\delta^N = (b_{0,\delta}, \dots, b_{N,\delta}), \quad b_{k,\delta} = (\mathcal{J}_\delta^{1,k}, \dots, \mathcal{J}_\delta^{N_{\text{elec}},k})$$

and

$$A_{:N}(x^N) = (A_0(x_0^N), \dots, A_N(x_N^N)) \quad \text{for } A_k(x_k^N) = (I(x_1^N, U^{1,k}), \dots, I(x_{N_{\text{elec}}}^N, U^{N_{\text{elec}},k})).$$

The spatiotemporal regulariser $R_{:N}$ arises, as in Examples 3.2 and 4.4, as the temporal infimal convolution of the box constraints and total variation of (5.2), with the algorithm-dependent dual temporal dynamics. The online algorithms of [18, 11] satisfy $x_\delta^N = (x_{0,\delta}, \dots, x_{N,\delta})$, i.e., the iterates for past time indices are not updated afterwards. Each $\sigma = x_{k,\delta}$ is produced exclusively from $b_{k,\delta}$ and an initial prediction of its value formed using the past iterates, as we explain in Section 5.3.

5.2 EXPERIMENTAL SETUP

We perform our experiments on synthetic data in a disk-shaped domain Ω equipped with $N_{\text{elec}} = 16$ equally spaced boundary electrodes. The CEM is solved using a finite element approximation of the electrode potentials and the conductivity, both represented with piecewise linear basis functions. For each time instance, current injection is applied sequentially at a single electrode while all remaining electrodes are grounded, resulting in $N_{\text{pat}} = 16$ excitation patterns. As is standard in EIT, the current measured at the excited electrode is excluded from the data, yielding $(N_{\text{elec}} - 1)N_{\text{pat}} = 240$ measurements per frame.

For the movement of the inclusion, we limit our attention the following test scenario from [11]:

Disappearing Inclusions Two inclusions move at uniform speed along circular paths. One inclusion disappears at time index $1/4$ (near frame 1500) and the other at time index $2/4$ (near frame 2800); both reappear at time index $3/4$ (near frame 4200). This scenario violates the incompressibility assumption.

We make the following important changes:

1. We use the same discretisation (5039 nodes) for both the forward simulations and the inverse problem. Although this would typically be regarded as an “inverse crime”, here it is a deliberate modelling choice: it eliminates discretisation-induced modelling error, allowing us to study the effect of measurement noise in isolation and, in particular, to consider the regime $\delta \rightarrow 0$.
2. We gradually slow down the originally uniform-speed motion of the inclusions for increasing time index $t = \min\{k/7000, 1\}$ by rescaling it to $s(t) := t + t^2 - t^3$. This satisfies $s(0) = 0$, $s(1) = 1$, $s'(0) = 1$, and $s'(1) = 0$. Therefore, initially, the inclusions start at the same speed as in [11], but—to allow evaluating the predictions of the regularisation theory—the movement slows to zero towards frame $t = 7000$, after which, there is no movement. The total number of data frames is $N = 10000$.

We illustrate the ground-truth data in Figure 1. To the ground-truth data, in our experiments, we add various levels of Gaussian noise δ .

The regularisation parameter α_δ is chosen heuristically. Based on a manual fit to empirically optimal values observed across a range of noise levels, we set $\alpha_\delta = 1 + \log_{10}(\delta)/10$. This choice is intended only for the numerically relevant range of noise levels considered here and is not derived from theoretical principles; in particular, it is not suitable for $\delta \leq 10^{-10}$. All remaining experimental parameters, as well as the affine prediction scheme, are adopted from [11].

5.3 ALGORITHM AND RESULTS

For the numerical algorithm, we use the “exact” variant of the online primal-dual method of our companion paper [18] with the step length and other parameters described therein, and the “affine scaling” dual predictor from [10]. This variant of the method, on each data frame k :

1. Uses an optical flow predictor to predict the the primal iterate (conductivity) σ^k to form the prediction $\check{\sigma}^{k+1}$,
2. Uses an “affine scaling” predictor (see [11]) to likewise form a prediction \check{y}^{k+1} from the dual iterate y^k . The dual iterates correspond to the variables that give rise to the total variation as a supremum over the dual ball; compare Example 4.4.
3. Solves both the PDE and its adjoint exactly to form $I(\sigma^k, U^{j,k})$. (This is what the “exact” refers to. In [18], also inexact PDE solves are studied.)

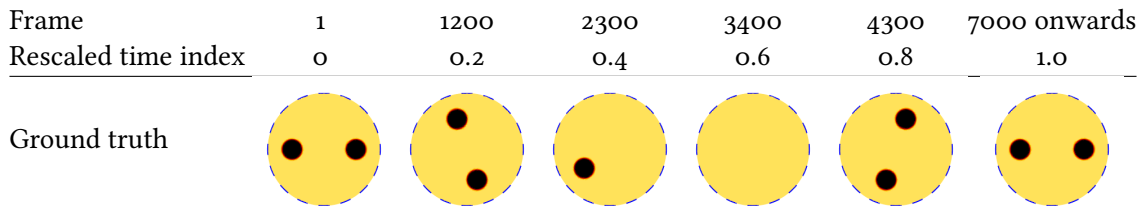
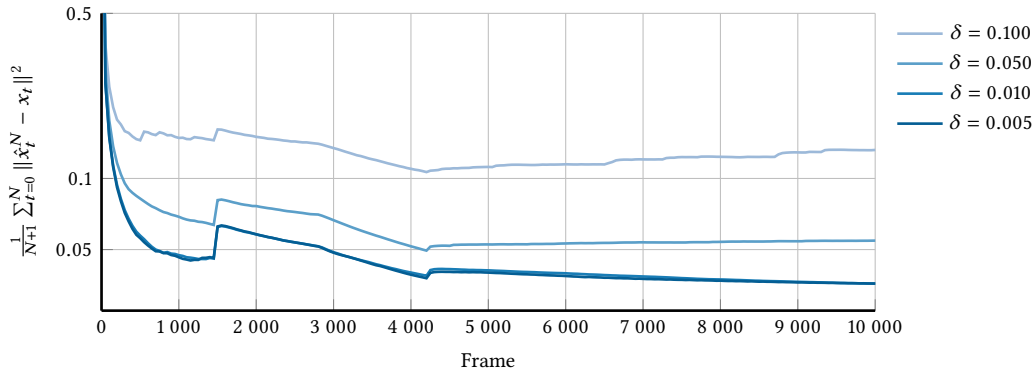


Figure 1: Illustration of the experimental data.

Figure 2: Illustration of asymptotic regularisation behaviour for various noise levels δ and regularisation parameter $\alpha = 1 + \log_{10}(\delta)/10$.

4. Takes *only one* primal-dual proximal splitting step from the prediction $(\check{\sigma}^{k+1}, \check{y}^{k+1})$ to form the next primal and dual iterates (σ^{k+1}, y^{k+1}) .

The predictions, to transfer iterates from one time from to the next one, along with the inexact (single-step) solve of the frame-wise optimisation problem in the final step, introduce the errors $e_{k,\delta}$ in (3.6).

Using the above experimental setup, we can now, in Figure 2, illustrate the claim of Theorem 3.19 on the convergence of the algorithmic solutions to a (minimum norm) ground truth as the noise level $\delta \rightarrow 0$, given that the errors and regularisation parameter have appropriate limiting behaviour.

REFERENCES

- [1] M. Alsaker, J. L. Mueller, and A. Stahel, A multithreaded real-time solution for 2D EIT reconstruction with the D-bar algorithm, *Journal of Computational Science* 67 (2023), 101967, doi:10.1016/j.jocs.2023.101967.
- [2] F. A. Artacho and M. H. Geoffroy, Characterization of metric regularity of subdifferentials, *Journal of Convex Analysis* 15 (2008), 365.
- [3] F. J. A. Artacho and M. H. Geoffroy, Metric subregularity of the convex subdifferential in Banach spaces, *J. Nonlinear Convex Anal.* 15 (2014), 35–47.
- [4] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, 2 edition, 2017, doi:10.1007/978-3-319-48311-5.
- [5] S. E. Blanke, B. N. Hahn, and A. Wald, Inverse problems with inexact forward operator: iterative regularization and application in dynamic imaging, *Inverse Problems* 36 (2020), 124001, doi:10.1088/1361-6420/abb5e1.

- [6] K. Bredies, K. Kunisch, and T. Pock, Total generalized variation, *SIAM Journal on Imaging Sciences* 3 (2010), 492–526.
- [7] M. Burger and S. Osher, Convergence rates of convex variational regularization, *Inverse Problems* 20 (2004), 1411.
- [8] F. Clarke, *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics, 1990, doi:10.1137/1.9781611971309.
- [9] C. Clason and T. Valkonen, *Introduction to Nonsmooth Analysis and Optimization*, MOS-SIAM Series on Optimization, SIAM, 2026, doi:10.1137/1.9781611978995.
- [10] N. Dizon, J. Jauhiainen, and T. Valkonen, Prediction techniques for dynamic imaging with online primal-dual methods, 2024, doi:10.1007/s10851-024-01214-w, arXiv:2405.02497.
- [11] N. Dizon, J. Jauhiainen, and T. Valkonen, Online optimisation for dynamic electrical impedance tomography, 2025, doi:10.1088/1361-6420/adcb66, arXiv:2412.12944.
- [12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, volume 375, Springer Science & Business Media, 1996.
- [13] D. J. Holland, D. M. Malioutov, A. Blake, A. J. Sederman, and L. F. Gladden, Reducing data acquisition times in phase-encoded velocity imaging using compressed sensing, *Journal of Magnetic Resonance* 203 (2010), 236–46.
- [14] A. Hunt, Weighing without touching: applying electrical capacitance tomography to mass flowrate measurement in multiphase flows, *Measurement and Control* 47 (2014), 19–25, doi:10.1177/0020294013517445.
- [15] A. D. Ioffe, *Variational Analysis of Regular Mappings: Theory and Applications*, Springer Monographs in Mathematics, Springer, 2017, doi:10.1007/978-3-319-64277-2.
- [16] J. Jauhiainen, N. Dizon, T. Valkonen, and Y. Nabou, Online optimisation codes for dynamic electrical impedance tomography, 2026, doi:10.5281/zenodo.19154746. Software.
- [17] J. Jauhiainen, P. Kuusela, A. Seppänen, and T. Valkonen, Relaxed Gauss–Newton methods with applications to electrical impedance tomography, *SIAM Journal on Imaging Sciences* 13 (2020), 1415–1445.
- [18] J. Jauhiainen, Y. Nabou, and T. Valkonen, Dynamic inverse problems: Single-loop online algorithms, 2026.
- [19] J. Jauhiainen, M. Pour-Ghaz, T. Valkonen, and A. Seppänen, Nonplanar sensing skins for structural health monitoring based on electrical resistance tomography, *Computer-Aided Civil and Infrastructure Engineering* 36 (2021), 1488–1507.
- [20] A. Lipponen, A. Seppänen, and J. P. Kaipio, Nonstationary approximation error approach to imaging of three-dimensional pipe flow: experimental evaluation, *Measurement Science and Technology* 22 (2011), 104013, doi:10.1088/0957-0233/22/10/104013.
- [21] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, 1998, doi:10.1007/978-3-642-02431-3.

- [22] G. Sarnighausen, T. Hohage, M. Burger, A. Hauptmann, and A. Wald, Regularization for time-dependent inverse problems: geometry of Lebesgue–Bochner spaces and algorithms, *Inverse Problems* 42 (2026), 015008, [doi:10.1088/1361-6420/ae30f9](https://doi.org/10.1088/1361-6420/ae30f9).
- [23] T. Schuster, B. Hahn, and M. Burger, Dynamic inverse problems: modelling—regularization—numerics, *Inverse Problems* 34 (2018), 040301, [doi:10.1088/1361-6420/aabof5](https://doi.org/10.1088/1361-6420/aabof5). Preface to special issue.
- [24] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski, *Regularization methods in Banach spaces*, volume 10, Walter de Gruyter, 2012.
- [25] T. Valkonen, Preconditioned proximal point methods and notions of partial subregularity, *Journal of Convex Analysis* 28 (2021), 251–278, [arXiv:1711.05123](https://arxiv.org/abs/1711.05123).
- [26] T. Valkonen, Predictive online optimisation with applications to optical flow, *Journal of Mathematical Imaging and Vision* 63 (2021), 329–355, [doi:10.1007/s10851-020-01000-4](https://doi.org/10.1007/s10851-020-01000-4), [arXiv:2002.03053](https://arxiv.org/abs/2002.03053).
- [27] T. Valkonen, Regularisation, optimisation, subregularity, *Inverse Problems* 37 (2021), 045010, [doi:10.1088/1361-6420/abe4aa](https://doi.org/10.1088/1361-6420/abe4aa), [arXiv:2011.07575](https://arxiv.org/abs/2011.07575).
- [28] A. Voss, N. Hänninen, M. Pour-Ghaz, M. Vauhkonen, and A. Seppänen, Imaging of two-dimensional unsaturated moisture flows in uncracked and cracked cement-based materials using electrical capacitance tomography, *Materials and Structures* 51 (2018), 68.
- [29] A. Voss, P. Hosseini, M. Pour-Ghaz, M. Vauhkonen, and A. Seppänen, Three-dimensional electrical capacitance tomography—A tool for characterizing moisture transport properties of cement-based materials, *Materials & Design* 181 (2019), 107967.