DYNAMIC INVERSE PROBLEMS: SINGLE-LOOP ONLINE ALGORITHMS

Jyrki Jauhiainen* Yassine Nabou† Tuomo Valkonen‡

Abstract We study efficient online methods for dynamic inverse problems with infinite time horizon. We concentrate, in particular, on problems whose forward model arises from a PDE. Our motivating application is flow monitoring with Electrical Impedance Tomography (EIT). The idea of such online methods is to take single steps of of standard optimisation algorithms, on each time index; each data frame. A predictor, based on problem dynamics, is used to transfer iterates one from time index to the next one. If we monitor a fast flow with a correspondingly fast measurement modality, such as EIT, basic methods are unable to solve the PDE before new data arrives. Our idea, then, is to not solve it, and instead, on each iteration, each time index, take single or few steps of standard iterative solvers towards the solution of both the PDE and an adjoint PDE. This is what "single loop" refers to. To the overall problem, we apply standard online optimisation methods, at the outside developed for exact gradients $\nabla E_k(x^k)$ of the iteration-dependent data fidelity E_k that incorporates the PDE. We replace the gradient by a single-loop estimate $\widetilde{\nabla E_k}(x^k)$ that satisfies standard smoothness properties with summable errors. This allows standard regret proofs to go through. Our numerical experiments on dynamic EIT validate the theoretical predictions and highlight the potential of the proposed approach for the real-time solution of PDE-constrained dynamic inverse problems.

1 INTRODUCTION AND MOTIVATION

Electrical Impedance Tomography (EIT) is a non-invasive imaging modality in which the internal electrical conductivity of a body is reconstructed from boundary voltage–current measurements. Due to its low cost, portability, and absence of ionising radiation, EIT has found applications in medical imaging (e.g. lung ventilation monitoring), geophysics, and industrial process monitoring. In many of these applications, the conductivity distribution evolves over time, for instance due to physiological dynamics, transport processes, or changing environmental conditions. As a consequence, measurements are acquired sequentially and reconstructions must be updated in real time, giving rise to *online* or *dynamic* inverse problems [22, 12, 13, 21, 5, 1, 11].

If we wish to solve the dynamic EIT problem, or other challenging dynamic inverse problem, in real-time, as the data progresses, we need very fast algorithms. If we want a high resolution solution, we may not have time to wait for partial differential equations (PDE) governing the problem to be solved; new data will arrive faster. In fact, we will usually also need to solve an adjoint PDE, adding to

This research has been supported by the Academy of Finland grant 345486. This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/F100800.

*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

the computational cost. In this work, based on the idea of *single-loop online differential estimation*, we will develop algorithms that are able to provide real-time solutions—by never solving the PDEs.

Mathematical model of EIT At time step k, the forward problem of dynamic EIT is governed by the Complete Electrode Model (CEM). Given a conductivity distribution $\sigma^k : \Omega \to \mathbb{R}$, the electric potential $u^{j,k} : \Omega \to R$ corresponding to the jth current injection ($j = 1, \dots, N_{\text{inj}}$) satisfies

$$\nabla \cdot (\sigma^k \nabla u^{j,k}) = 0 \qquad \text{in } \Omega, \tag{1.1a}$$

$$u^{j,k} + \zeta_i \sigma^k \nabla u^{j,k} \cdot \nu = U_i^{j,k} \qquad \text{on } \partial\Omega_{e_i}, \tag{1.1b}$$

$$\sigma^k \nabla u^{j,k} \cdot \nu = 0 \qquad \text{on } \partial\Omega \setminus \cup_i \partial\Omega_{e_i}, \tag{1.1c}$$

$$\int_{\partial\Omega_{e_i}} \sigma^k \nabla u^{j,k} \cdot \nu \, dS = -I_i^{j,k} \qquad \text{for } i = 1, \dots, N_{\text{elec}}. \tag{1.1d}$$

Here $\Omega \subset \mathbb{R}^d$ denotes the imaging domain, $\partial\Omega_{e_i}$ are the electrode surfaces, $\zeta_i > 0$ are contact impedances, and ν is the outward unit normal. The electrode potentials $U_i^{j,k}$ for electrodes $i = 1, \dots, N_{\text{elec}}$ are prescribed, while the resulting electrode currents $I_i^{j,k}$ are determined by the model. We adopt a potential-to-current formulation of the CEM, in which the electrode potentials $U^{j,k}$ are prescribed and the corresponding electrode currents $I^{j,k} = I(\sigma^k, U^{j,k})$ are computed. This reflects modern EIT measurement devices [15]. Here $\mathcal{I}^{j,k}$ denotes noisy measurements of the electrode currents corresponding to the applied electrode potentials $U^{j,k}$. The inverse problem consists in recovering σ^k from noisy boundary measurements $\mathcal{I}^{j,k}$. Applying total variation regularisation, this is commonly formulated the optimisation problem

$$\min_{\sigma} E_k(\sigma) + F_k(\sigma) + G_k(K_k\sigma) \tag{1.2}$$

for K_k a (discretised) gradient operator,

$$E_k(\sigma) := \frac{1}{2} \sum_{j=1}^{N_{\text{inj}}} \|\Sigma^{-1/2}(I(\sigma, U^{j,k}) - \mathcal{I}^{j,k})\|_2^2, \quad F_k(\sigma) := \delta_{[\sigma_m, \sigma_M]}(\sigma), \quad \text{and} \quad G_k(y) := \alpha \|y\|_{2,1}, \tag{1.3}$$

where $0 < \sigma_m < \sigma_M$ enforce physical bounds, $\Sigma^{-1/2}$ is a data precision matrix, and $\alpha > 0$ a regularisation parameter. Similar formulations have been studied in [24, 25, 18].

General online optimisation framework Motivated by online EIT and related applications, we consider the general infinite time horizon optimisation problem

$$\min_{(x^0, x^1, \dots) \in \bar{\mathcal{X}}} \sum_{k=0}^{\infty} V_k(x^k), \quad \text{for} \quad V_k(x) := E_k(x) + G_k(K_k x) + F_k(x) \quad \text{and} \quad \bar{\mathcal{X}} \subset \prod_{k=0}^{\infty} X_k, \tag{1.4}$$

where $E_k : X_k \to \mathbb{R}$ is smooth but possibly nonconvex, $K_k \in \mathbb{L}(X_k; Y_k)$, $G_k : Y_k \to \mathbb{R}$ and $F_k : X_k \to \mathbb{R}$ are convex and possibly nonsmooth. X_k, Y_k are normed spaces. The set $\bar{\mathcal{X}}$ encodes the dynamics of the problem. The data fidelity term is of the composite form

$$E_k(x) = J_k(S_{k,u}(x)), \qquad 0 = T_k(S_{k,u}(x), x), \tag{1.5}$$

and $S_{k,u} : X_k \to U_K$ denotes the solution operator for the PDE constraint $T_k(u, x) = 0$. Here, the subscript u is symbolic, indicating that the operator maps into the space U_k (to distinguish it from the adjoint solution map $S_{k,w}$ defined later). That is, for each x, the state $u = S_{k,u}(x) \in U_k$ satisfies the underlying (possibly nonlinear) PDE. This formulation encompasses EIT as well as a broad class of dynamic inverse and control problems.

Contributions With the goal of developing efficient online methods for the problem (1.4), in Sections 3 and 4, we develop **single-loop estimates $\widetilde{E'_k}(\breve{x}^k)$ of differentials** $E'_k(\breve{x}^k)$ of composite functions $E_k = J_k \circ S_{k,u}$, where $S_{k,u}$ denotes the solution mapping of the inner PDE constraint. The idea is that on each data frame or outer iteration k, we only make few very simple operations to update our estimate $\widetilde{E'_k}(\breve{x}^k)$ into $\widetilde{E'_{k+1}}(\breve{x}^{k+1})$. Here $\breve{x}^k$ is a prediction of x^k, based on the dynamics of the problem. The latter is formed by a correction step performed by the optimisation method. This prediction is essential, as, in a dynamic problem, x^{k-1} may no longer be valid or good starting point for the optimisation problem, the data of the problem having changed. This fact is also modelled by the time-indexed spaces X_k.

The differential estimates lift standard iterative algorithms, such as Gauss–Seidel splitting, for estimating $S_{k,u}(\breve{x}^k)$, as well as an adjoint equation that defines $E'_k(\breve{x}^k) = J'_k(S_{k,u}(\breve{x}^k))S'_{k,u}(\breve{x}^k)$. However, on a single outer iteration or data frame k, we never iterate these *inner and adjoint algorithms* to completion: our goal is to—justifiably—only take a single step or a small number of step of these algorithms. This is what *single-loop* refers to.

The core idea, rigorously developed in Section 3, is that all the complexity of the analysis single-loop methods reduces to the estimates $\widetilde{E'_k}(\breve{x}^k)$ of $E'_k(\breve{x}^k)$ satisfying standard smoothness properties with controlled error. Rather than recomputing inner and adjoint solutions from scratch on each time step, the key to this working, is to recursively exploit already computed information from previous data frames, and to track and control the propagation of errors. In Section 4 we show that the steps of standard splitting algorithms are capable of this, when combined with appropriate *predictors* to account for the dynamic nature of the problem.

These new online differential estimates can directly be used in the (outer) online primal-dual method of [5], which we recall in Section 2. We also previously applied to method to dynamic EIT, however using intermittent PDE solutions from a background thread to obtain real-time performance. We demonstrate numerically in Section 5 that the new single-loop differential estimates can provide much more stable solutions without thread synchronisation difficulties and—as follows from Sections 2 and 3—with less analytical complexity.

Our approach to generating the estimates extends the static theories of [19, 6, 23]. Alternative real-time algorithms for dynamic EIT include the multi-threaded approach of [1], based on the D-bar method. Our companion paper [17] studies relevant dynamic regularisation theory: how do the algorithmic solutions behave as the noise level $\delta \searrow 0$, the corresponding regularisation parameter $\alpha_\delta \searrow 0$, and the frame $k \nearrow \infty$.

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. For Fréchet differentiable $F : X \to R$, 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 $F : X \to \overline{\mathbb{R}}$, we write $\partial F : X \rightrightarrows X^*$ for its subdifferential map. On a normed space X, for a point $x \in X$ and a set $U \subset X$, we write $\operatorname{dist}(x, U) := \inf_{x' \in U} \|x - x'\|_X$, where $\|\,\cdot\,\|_X$ is the norm on X. We also write $\operatorname{dist}^2(x, U) := \operatorname{dist}(x, U)^2$. 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 Banach space. We only work with real Hilbert spaces. We write $\mathrm{Id} : X \to X$ for the identity operator on X and $\delta_A : X \to \overline{\mathbb{R}}$ for the $\{0, \infty\}$-valued indicator function of a set $A \subset X$.

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

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

and (inner product) Young's inequality

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

Finally, given a set $\mathcal{X} \subset \prod_{k=0}^{\infty} X_k$, we define for $n \leq m$ the slices $\mathcal{X}_{n:m} := \{x_{n:m} \mid, (x_0, x_1, \ldots) \in \mathcal{X}\}$, where $x_{n:m} := (x_n, \ldots x_m)$. We may use superscripts instead of subscripts in the case of iterates of algorithms.

2 AN INEXACT ONLINE PRIMAL-DUAL METHOD

Recall the infinite-horizon optimisation problem (1.4),

$$\min_{(x^0,x^1,\ldots)\in\breve{\mathcal{X}}} \sum_{k=0}^{\infty} V_k(x^k), \quad \text{for} \quad V_k(x) := E_k(x) + G_k(K_k x) + F_k(x), \tag{2.1}$$

where E_k is a possibly nonconvex but finite-valued function, F_k and G_k are convex, proper, lower semicontinuous functions, and K_k is a linear operator. The set $\breve{\mathcal{X}} \subset \prod_{k=0}^{\infty} X_k$ encodes the primal dynamics. Note that, due to the Hilbert-space nature of Algorithm 2.1, in this section we talk about gradients $\nabla E_k(\breve{x}^k) \in X_k$ and their estimates, while in most of the rest of the manuscript we talk about differentials $E'_k(\breve{x}^k) \in X_k^*$ and their estimates in general normed spaces. The former is the Riesz representation of the latter. Often, even in Hilbert spaces, it is more convenient to not work with the presentation. For the online solution of (2.1), in this section, we recall the primal-dual method of [5], presented in Algorithm 2.1, and its assumptions and regret theory. *Regret* is replacement in online optimisation of the concept of *convergence* in static optimisation. The rough idea is to bound the regret of past decisions (steps taken by the optimisation algorithm) subject to new information. In practise, mathematically, regret results are analogous to convergence results, but often weaker.

Algorithm 2.1 consists of three main steps:

1. *Predict* the primal and dual variables $\breve{u}^{k+1} = (\breve{x}^{k+1}, \breve{y}^{k+1})$ for frame $k+1$, from the iterate $u^k = (x^k, y^k)$ of frame k (and earlier frames). This should use appropriate dynamical models; in the EIT experiments of [5] and Section 5, the transport equation (optical flow) is used.

2. *Form* the gradient estimate $\widetilde{\nabla E}_{k+1}(\breve{x}^{k+1})$. Typically E_{k+1} is the data term, and in our EIT application of interest, computing $\nabla E_{k+1}(\breve{x}^{k+1})$ exactly would involve solving a PDE and an adjoint PDE. We want to avoid this. In the present paper, to do this, we will use single or few *inner* and *adjoint* steps that arise from standard algorithms. This is the topic of Section 4.

3. *Correct* the primal and dual variables to $u^{k+1} = (x^{k+1}, y^{k+1})$ using a standard primal-dual proximal splitting step based on the predicted iterate, and employing the differential estimate.

This algorithm was, thus, already designed with differential estimates in mind. As we will see in this section, these estimates only need to satisfy simple smoothness-type properties with controlled error. We will in the next Section 3 represent a rigorous theory of single-loop online differential estimation for forming estimates that satisfy those properties.

2.1 ASSUMPTIONS AND BASIC SETUP

We need to make several assumptions. Compared to [5], for simplicity, we choose step length parameters and growth and smoothness factors to be independent of the frame number k. To allow evolving data to be located anywhere in the problem, all the functions and operators in (2.1), and spaces, however, remain dependent on the frame index k. We refer to [5] for the full theory.

The first assumption formally introduces the relevant functions, predictors, and comparison sets.

Assumption 2.1 (Basic structural assumptions [5, Assumption 2.1]). Let $N \in \mathbb{N}$. On real Hilbert spaces X_k and Y_k for $0 \leq k \leq N$, we are given:

Algorithm 2.1 Nonconvex predictive online primal-dual proximal splitting (POPD-N)

Require: For all $k \in \mathbb{N}$, on real Hilbert spaces X_k and Y_k: convex, proper, lower semicontinuous $E_{k+1}: X_{k+1} \to \mathbb{R}$, $F_{k+1}: X_{k+1} \to \overline{\mathbb{R}}$, and $G^*_{k+1}: Y_{k+1} \to \overline{\mathbb{R}}$; a primal-dual predictor $P_k: X_k \times Y_k \to X_{k+1} \times Y_{k+1}$; and an operator $K_{k+1} \in \mathbb{L}(X_{k+1}; Y_{k+1})$. Estimates $\widetilde{E'_{k+1}}(\breve{x}^{k+1})$ of the gradients $E'_{k+1}(\breve{x}^{k+1})$. Step length parameters $\tau, \sigma > 0$.

Pick initial iterates $x^0 \in X_0$ and $y^0 \in Y_0$.
for $k \in \mathbb{N}$ **do**
  $(\breve{x}^{k+1}, \breve{y}^{k+1}) := P_k(x^k, y^k)$ ▷ prediction step
  $x^{k+1} := \operatorname{prox}_{\tau F_{k+1}}(\breve{x}^{k+1} - \tau\widetilde{\nabla E}_{k+1}(\breve{x}^{k+1}) - \tau K^*_{k+1}\breve{y}^{k+1})$ ▷ primal step
  $y^{k+1} := \operatorname{prox}_{\sigma G^*_{k+1}}(\breve{y}^{k+1} + \sigma K_{k+1}(2x^{k+1} - \breve{x}^{k+1}))$ ▷ dual step
end for

(i) Convex, proper, lower semicontinuous functions $F_k: X_k \to \overline{\mathbb{R}}$ and $G^*_k: Y_k \to \overline{\mathbb{R}}$, with uniform strong convexity factors $\gamma \ge 0$ and $\rho \ge 0$, respectively; operators $K_k \in \mathbb{L}(X_k; Y_k)$; and a possibly nonconvex but finite-valued function $E_k: X_k \to \mathbb{R}$.

(ii) A bounded set $\mathscr{U} \subset U$ of primal-dual comparison sequences. Here

$$U := (X_0 \times Y_0) \times (X_1 \times Y_1) \times (X_2 \times Y_2) \cdots .$$

(iii) Primal-dual predictors $P_k: X_k \times Y_k \to X_{k+1} \times Y_{k+1}$ that produce $\breve{u}^{k+1} = P_k(u^k)$.

The predictor P_k is merely a notational convenience. Unlike the notation suggests, we allow $(\breve{x}^{k+1}, \breve{y}^{k+1}) = P_k(x^k, y^k)$ to depend on the entire history, not merely (x^k, y^k).

From the point of view of the satisfaction of our remaining assumptions, we would ideally be able to take $\mathscr{U}$ as the whole space. However, our main regret result does not allow this. Another practical option would be the set of solutions to (2.1), without observing the dynamic constraint $\bar{\mathscr{X}}$. However, intuitively, both

$$\mathscr{X} := \{\bar{x}^{0:\infty} \in X_{0:\infty} \mid (\bar{x}^{0:\infty}, \bar{y}^{0:\infty}) \in \mathscr{U}\} \tag{2.2}$$

and the predictors P_k should approximate the original primal dynamics $\bar{\mathscr{X}}$.

Our second assumption formalises the requirement of a "smooth estimate" $\widetilde{\nabla E}_k(\breve{x}^k)$ of each $\nabla E_k(\breve{x}^k)$. The monotonicity-like conditions in (i) and (ii) are superficially similar, but have slight differences: the first one is supposed to hold globally for any x, the second one, only locally with better factors. The regret bound proofs in [5] use the global version to derive an *a priori* estimate that ensures that the iterates stay in the neighbourhood where the second *a posteriori* estimate holds. Verifying this assumption for single-loop estimates generated by standard inner and adjoint algorithms, is the main content of this paper in the following Section 4. In [5] a background solver was used to generate these estimates. Moreover, [5] also formalises corresponding fully global assumption. For simplicity, we concentrate here on the local version, only.

Assumption 2.2 (Local smoothness and growth, and step length bounds [5, Assumption 2.3]). [1] Let $(\breve{x}^k, \breve{y}^k) = P_{k-1}(u^{k-1})$. Given $N \in \mathbb{N}$, Assumption 2.1 holds, and

(i) **Global a priori smoothness and growth:** For all $1 \le k \le N$, E_k satisfies for some growth and smoothness factors $\bar{\gamma}_E \in \mathbb{R}$, $\bar{\lambda}_E \ge 0$, and errors $\bar{e}_k \ge 0$, for all $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$, and for all $x^k \in X_k$ the "erroneous" three-point monotonicity-like property

$$\langle \widetilde{\nabla E}_k(\breve{x}^k) - \nabla E_k(\hat{x}^k), x^k - \hat{x}^k \rangle \ge \bar{\gamma}_E \|x^k - \hat{x}^k\|^2 - \bar{\lambda}_E \|x^k - \breve{x}^k\|^2 - \bar{e}_k.$$

(ii) **Local a posteriori smoothness and growth:** For every comparison sequence $\bar{x}^{1:N} \in \mathcal{X}_{1:N} \cap \prod_{i=1}^{N} B(\check{x}^i, \delta)$, for all $1 \le k \le N$, E_k satisfies for some smoothness and growth factors $\lambda_E \ge 0$ and $\hat{\gamma}_E \ge \gamma_E \ge 0$, errors $e_k, \hat{e}_k \ge 0$, and a radius $\delta > 0$, for all $x^k \in B(\bar{x}^k, \delta)$, the inequality

$$\langle \widetilde{\nabla} E_k(\check{x}^k), x^k - \bar{x}^k \rangle \ge E_k(x^k) - E_k(\bar{x}^k) + \frac{\gamma_E}{2}\|x^k - \bar{x}^k\|^2 - \frac{\lambda_E}{2}\|x^k - \check{x}^k\|^2 - e_k,$$

as well as, for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0) \cap B(\check{x}^k, \delta) \times Y_k$ and $x^k \in B(\hat{x}^k, \delta)$ that

$$\langle \widetilde{\nabla} E_k(\check{x}^k) - \nabla E_k(\hat{x}^k), x^k - \hat{x}^k \rangle \ge \hat{\gamma}_E \|x^k - \hat{x}^k\|^2 - \hat{\lambda}_E \|x^k - \check{x}^k\|^2 - \hat{e}_k.$$

Here the predicted primal variable $\check{x}^k$ is given by Assumption 2.1.

(iii) **Step length factors:** The step length parameters $\tau, \sigma > 0$ along with a $\kappa \in (0, 1)$, satisfy

$$1 \ge \max\Big\{\lambda_E,\ -2(\gamma + \bar{\gamma}_E),\ \tfrac{2}{1-\kappa}\bar{\lambda}_E,\ \tfrac{2}{1-\kappa}\hat{\lambda}_E\Big\}\tau + \tau\sigma\|K_k\|^2. \tag{2.3}$$

Our final assumption relates the comparison set to the set of critical points of H, and sets bounds on step lengths. To state the assumption, we need to introduce additional notation. First of all, the first-order optimality conditions for the static problem $\min F_k + E_k + G_k \circ K_k$ for a single frame k in (2.1), can be expressed with the general notation $u = (x, y)$ as [2]

$$0 \in H_k(\hat{u}^k) \quad \text{for} \quad H_k(u) := \begin{pmatrix} \partial F_k(x) + \nabla E_k(x) + K_k^* y \\ \partial G_k^*(y) - K_k x \end{pmatrix}. \tag{2.4}$$

We also write

$$H_{n:m}(u^{n:m}) := H_n(u^n) \times \cdots \times H_m(u^m) \subset U_{n:m}.$$

Algorithm 2.1 may then be written in implicit form as the iterative solution of u^k from

$$0 \in \widetilde{H}_k(u^k) + M_k(u^k - \breve{u}^k), \quad \breve{u}^k := P_k(u^k),$$

where

$$\widetilde{H}_k(u) := \begin{pmatrix} \partial F_k(x) + \widetilde{\nabla} E_k(\check{x}^k) + K_k^* y \\ \partial G_k^*(y) - K_k x \end{pmatrix} \quad \text{and} \quad M_k = \begin{pmatrix} \tau^{-1}\,\mathrm{Id} & -K_k^* \\ -K_k & \sigma^{-1}\,\mathrm{Id} \end{pmatrix}.$$

[1]Compared to [5, Assumption 2.3], this fixes the *testing parameters* $\varphi_k \equiv 1$, $\eta_k \equiv \tau$, and $\psi_k \equiv \tau/\sigma$.

Moreover, we use a slightly weaker version of the smoothness and growth assumptions. In [5], the corresponding estimates are formulated for arbitrary test points in the indicated neighbourhoods and arbitrary framewise comparison points. In the present paper, we only require them at the specific primal iterate x^k generated by the algorithm, and only for the kth component $\bar{x}^k$ of the comparison sequence under consideration.

This restriction is sufficient for the arguments used for [5, Theorem 2.12]. Indeed, in the theorem, the relevant estimates are only applied at the current iterate. Likewise, the comparison point is not used as an arbitrary isolated element of the framewise slice $\mathcal{X}_k$, but as the kth component of a fixed comparison sequence $\bar{x}^{1:N} \in \mathcal{X}_{1:N}$. More precisely, in the proof of [5, Theorem 2.12], $\bar{x}^k$ is the primal component of the kth component of the fixed comparison sequence $\bar{u}^{1:N} \in \mathcal{U}_{1:N}$. This is the reason for formulating the present assumption for comparison sequences $\bar{x}^{1:N} \in \mathcal{X}_{1:N}$ rather than for arbitrary framewise comparison points. Once such a sequence is fixed, the proof only uses its kth x component at time k. Further, [5, Assumption 2.5(i)] is applied in [5, Lemma 2.10, before equation (17)] with $x = x^k$. The second estimate in [5, Assumption 2.5(ii)] is applied in [5, Lemma 2.11, before equation (24)], again with $x = x^k$. The first estimate in [5, Assumption 2.5(ii)] is applied in [5, Theorem 2.12, equation (29)], also with $x = x^k$.

The remaining use of these constants is indirect. In [5, Lemma 2.9], they are used to prove positive definiteness of the testing operators, in particular of $\hat{\Gamma}_k$. For this step, it is enough to retain the coefficient ordering $\hat{\gamma}_{E,k} \ge \gamma_{E,k}$; the estimates themselves are not applied to arbitrary additional test points. This positive definiteness estimate is later used in [5, Lemma 2.11] for the term $u^k - \bar{u}^k$.

Hence, for the present verification, it is enough to prove the smoothness and growth inequalities at the realized algorithmic point x^k. The neighbourhood assumptions are still needed to ensure that the predictor, comparison point, and reference point lie in the region where the local estimates are valid.

This formulation facilitates convergence and regret analysis. We also define the growth operator

$$\Gamma := \operatorname{diag}\big((\gamma + \gamma_E)\,\mathrm{Id},\ \rho\,\mathrm{Id}\big).$$

Recalling that $\breve{u}^{k+1} := P_k(u^k)$, we then define for[2] $[\![x]\!]_A^2 := \langle Ax, x\rangle$ the *prediction errors*

$$\varepsilon^\dagger_{k+1}(u^k, \bar{u}^{k:k+1}) := \frac{1}{2}[\![\breve{u}^{k+1} - \bar{u}^{k+1}]\!]^2_{\tau M_{k+1}} - \frac{1}{2}[\![u^k - \bar{u}^k]\!]^2_{\tau(M_k+\Gamma)} \quad \text{for all} \quad \bar{u}^{k:k+1} \in \mathscr{U}_{k:k+1}. \tag{2.5}$$

They measure the difference of deviation from a chosen comparison sequence between the current iterate and its prediction. When u^k and $\bar{u}^{k:k+1}$ are clear from the context, we abbreviate $\varepsilon^\dagger_{k+1} := \varepsilon^\dagger_{k+1}(u^k, \bar{u}^{k:k+1})$.

Assumption 2.3 (Critical point proximity [5, Assumption 2.6]). Let $N \in \mathbb{N}$ and suppose Assumption 2.2 holds. Then:

(i) For constants $\hat{r}_k > 0$ ($1 \le k \le N$), the primal-dual comparison set $\mathscr{U}_{1:N}$ stays close to the critical points of $H_{1:N}$ in the sense that

$$\mathscr{U}_{1:N} \subset \left\{ \bar{u}^{1:N} \in U_{1:N} \;\middle|\; \inf_{\hat{u}^{1:N} \in H^{-1}_{1:N}(0)} \|\bar{u}^k - \hat{u}^k\|_{M_k + \hat{\Gamma}_k} \le \hat{r}_k \text{ for } 1 \le k \le N \right\}.$$

(ii) Let $\theta_k := 1 + 2\tau \min\{\gamma + \bar{\gamma}_E, \bar{\lambda}_E\} - \tau\sigma\|K_k\|^2 > 0$, where positivity follows from (2.3). Then for some $\xi_k, \Delta > 0$, $\tilde{\delta} \in (0, \delta)$, and for all $\bar{u}^{0:N} \in \mathscr{U}_{0:N}$, we have[3]

$$\begin{aligned} 0 < d_N(\bar{u}^{0:N}) := \inf_{0 \le n \le N-1} \Bigg(& \frac{\theta_{n+1}(\delta - \tilde{\delta})^2}{\xi_{n+1}} - \frac{1+\Delta}{\xi_{n+1}} \tau \hat{r}^2_{n+1} - 2\varepsilon^\dagger_{n+1}(u^n, \bar{u}^{n:n+1}) \\ & - \sum_{k=1}^{n} \left(\frac{1+\Delta}{2\kappa} \tau \hat{r}^2_k + \varepsilon^\dagger_k(u^{k-1}, \bar{u}^{k-1:k}) + \tau \hat{e}_k \right) \Bigg) \end{aligned}$$

and

$$(2-\kappa)(\xi_k^{-1} + 1)(\delta - \tilde{\delta})^2 + 2\theta_k^{-1}\tau \bar{e}_k \le \tilde{\delta}^2.$$

Here κ arises from Assumption 2.2 (iii).

The first part of the assumption ensures that the comparison set, affected by problem dynamics, does not stray too far from the sequences of static solutions. It holds for some $\hat{r}_{n+1} > 0$ subject to boundedness assumptions on $H^{-1}_{1:N}(0)$. If we can *choose* $\mathscr{U}_{1:N}$ to include $H^{-1}_{1:N}(0)$, then also $\hat{r}_{n+1} = 0$. This, however, affects the prediction errors (2.5). If the prediction errors, as well as the smoothness errors $\tau\hat{e}_k$ and $\bar{e}_k$ from Assumption 2.2 are small (compared to τ) and with a bounded sum, the second part can also be made to hold for some choices of τ, σ, $\tilde{\delta}$, ξ_k, and Δ. (Note that definition of the prediction error $\varepsilon^\dagger_{k+1}(u^k, \bar{u}^{k:k+1})$ also includes multiplication by τ.)

2.2 REGRET

Our main regret results talks about "temporal sub-infimal convolutions" of the original objective, and the dual dynamics of the problem. To state the result, besides the primal projection (2.2) of the set of comparison sequences, also define the dual projection

$$\mathscr{Y} := \left\{ \bar{y}^{0:\infty} \in Y_{0:\infty} \;\middle|\; (\bar{x}^{0:\infty}, \bar{y}^{0:\infty}) \in \mathscr{U} \right\}.$$

[2] We use this notation when A might not be positive semi-definite of self-adjoint, and therefore not define a (semi-)norm

[3] The infimum here is over $0 \le n \le N-1$, while in [5] it is over $0 \le n \le N$. The former is sufficient, as [5, (23) or (2.19) in arXiv version, in the proof of Lemma 2.11] only has to hold for $1 \le n \le N$, and avoids an inconvenient indexing of $\bar{u}^{0:N}$ in this definition when $n = N$. In either case, passing to the limit $N \to \infty$, the same final bounds will be assumed.

Then define

$$\mathring{G}_{1:N}(z^{1:N}) := \sup_{\tilde{y}^{1:N}\in\mathscr{Y}_{1:N}} \sum_{k=1}^{N}\left[\langle z^k, \tilde{y}^k\rangle - \tau G_k^*(\tilde{y}^k)\right], \quad \text{and} \quad Q_{1:N}(x^{1:N}) := \sum_{k=1}^{N} \tau[F_k + E_k](x^k).$$

Also let[4]

$$K_{1:N}x^{1:N} := (\tau_1 K_1 x^1, \ldots, \tau_N K_N x^N) \quad \text{and} \quad G_{1:N}(z^{1:N}) := \sum_{k=1}^{N}[\tau G_k^*]^*(z^k) = \sum_{k=1}^{N} \tau G_k(z^k/\tau).$$

We have $\mathring{G}_{1:N} \le G_{1:N} \,\square\, \delta^*_{\mathscr{Y}_{1:N}}$. Thus $\mathring{G}_{1:N}$ behaves as a “sub-infimal convolution” of $G_{1:N}$ with the dual temporal dynamics. In our companion paper [17], we show that, actually,

$$\mathring{G}_{1:N} = G_{1:N} \,\square\, \delta^*_{\operatorname{cl}\operatorname{conv}\mathscr{Y}_{1:N}},$$

where $\square$ denotes the infimal convolution, and $\delta_{\operatorname{cl}\operatorname{conv}\mathscr{Y}_{1:N}}$ the indicator function of the closed convex hull of $\mathscr{Y}_{1:N}$. Its conjugate is the so-called support function of this set.

Example 2.4. Take $G_k = \alpha\|\cdot\|_{2,1}$, and $K_k = \nabla$ in a suitable finite element space. Thus $G_k \circ K_k$ models a frame-wise total variation regulariser for the regularisation parameter α. Then $\mathring{G}_{1:N}(K_{1:N}\,\cdot\,)$ becomes a spatiotemporal regulariser that bridges between the frame-wise total variation regularisers, and the dual dynamics encoded in $\operatorname{cl}\operatorname{conv}\mathscr{Y}_{1:N}$.

We may now state the main regret result for Algorithm 2.1. Until each index N, it bounds the values of a modified primal objective function that, compared to an initial segment of (1.4), replaces $G_{1:N} + \delta_{\tilde{\mathscr{X}}_{1:N}}$ by $\mathring{G}_{1:N}$.

Corollary 2.5 ([5, Corollary 2.13]). *Let $N \ge 1$, and suppose Assumptions 2.1 to 2.3 holds for $u^N = (x^N, y^N)$ generated by Algorithm 2.1, with initialisation $u^0 \in X_0 \times Y_0$. If*

$$\frac{1}{2}\|u^0 - \hat{u}^0\|^2_{\tau(M_0+\Gamma_0)} \le d_N(\hat{u}^{0:N}), \tag{2.6}$$

then

$$\begin{aligned}\left[Q_{1:N}(x^{1:N}) + \mathring{G}_{1:N}(K_{1:N}x^{1:N})\right] \le \sup_{\hat{x}^{1:N}\in\mathscr{X}_{1:N}} &\left[Q_{1:N}(\hat{x}^{1:N}) + \mathring{G}_{1:N}(K_{1:N}\hat{x}^{1:N})\right] \\ &+ \sup_{\hat{u}^{0:N}\in\mathscr{U}_{0:N}} \left(\tfrac{1}{2}\|u^0 - \hat{u}^0\|^2_{\tau M_0+\Gamma_0} + c_N(\hat{x}^{1:N}, y^{1:N}) + e_N^{\Sigma}(u^{0:N-1}, \hat{u}^{0:N})\right),\end{aligned}$$

where the cumulative prediction and gradient error

$$e_N^{\Sigma}(u^{0:N-1}, \hat{u}^{0:N}) := \sum_{k=0}^{N-1}\left(\varepsilon^{\dagger}_{k+1}(u^k, \hat{u}^{k:k+1}) + \tau e_{k+1}\right).$$

and the comparison set solution discrepancy

$$c_N(\hat{x}^{1:N}, y^{1:N}) := \inf_{\tilde{y}^{1:N}\in\mathscr{Y}_{1:N}} \langle K_{1:N}\hat{x}^{1:N}, y^{1:N} - \tilde{y}^{1:N}\rangle + G^*_{1:N}(\tilde{y}^{1:N}) - G^*_{1:N}(y^{1:N}).$$

Remark 2.6 (Comparison set solution discrepancy [3, Remark 2.7 and Section 3]). We have $c_N \le 0$ when $G_k \circ K_k = \alpha\|\nabla\,\cdot\,\|_{2,1}$ is the total variation regulariser, the dual initialisation achieves the total variation in the sense that $\langle y^0, x^0\rangle = \alpha\|\nabla x^0\|_{2,1}$ with $\|y^0\|_{2,\infty} \le \alpha$, and the dual predictor is total variation preserving. Examples of such predictors are found in [3].

[4] In [5], incorrectly, $G_{1:N}(z^{1:N}) := \sum_{k=1}^{N} \tau G_k(z^k)$, however, same as here, $G_{1:N}$ is only ever used in discussions. With the definitions here, $G_{1:N}(Kx^{1:N}) = \sum_{k=1}^{N}[\tau G_k^*]^*(x^k) = \sum_{k=1}^{N} \tau G_k(K_k x^k)$, consistently with $Q_{1:N}$.

3 Online differential estimation

Our idea for a low-cost estimate of $E_k'(\check{x}^k) = J'(S_{k,u}(\check{x}^k))S_{k,u}'(\check{x}^k)$ is grounded in the work of [19, 6] for static problems: on each iteration of the outer Algorithm 2.1 for the solution of (1.4)—for each new data frame as it arrives—we take only a *single step* or few steps of standard solvers, such as Gauss–Seidel splitting, towards the solution of (1.1) and its corresponding adjoint equation. As we do this on each iteration, *if* the data does not change too abrubtly between frames, we expect to still get convergence, or, in case of online optimisation, bounded regret.

We start by deriving the general scheme of approximation from adjoint equations in Section 3.1. Then in Section 3.2 we formulate an *online tracking assumption* that specific algorithms for the construction of the estimates have to satisfy. In Section 3.3 we use the tracking assumption to prove smoothness properties of the estimates $\widetilde{E_k'}(\check{x}^k)$ of $E_k'(\check{x}^k)$, in particular, to prove Assumption 2.2. Finally, we study quadratic E_k in detail in Section 3.4. In the next Section 4 we will prove the tracking assumption for specific inner and adjoint algorithms. To ensure the applicability of our results to a broad class of problems—including PDE-constrained and dynamic inverse problems—we work in abstract normed spaces. Our theory is also aplicable to online bilevel optimisation.

3.1 Adjoint equations and the general approach

Let $J_k : U_k \to \mathbb{R}$ and $S_{k,u} : X_k \to U_k$, for $k \in \{1, \ldots, n\}$, be Fréchet differentiable mappings on normed spaces X_k and U_k. We estimate

$$E_k'(\check{x}^k) = J_k'(S_{k,u}(\check{x}^k))\, S_{k,u}'(\check{x}^k) \tag{3.1}$$

by approximating $S_{k,u}(\check{x}^k)$ with a computable state $u^k \in U_k$ and its derivative $S_{k,u}'(\check{x}^k)$ by a $p^k \in \mathbb{L}(X_k; U_k)$. However, as $p^k \approx S_{k,u}'(\check{x}^k) \in \mathbb{L}(X_k; U_k)$ can be very high-dimensional, we want to avoid constructing it directly. Instead, we seek to only construct the necessary projections w^k. This is illustrated by the next lemma, where T_k can model, e.g., a PDE or the first-order optimality conditions of the inner optimisation problem (1.5) (in which case $T_k = E_k'$), both parametrised by x. The proof follows standard techniques for deriving reduced adjoint equations of PDEs; see, e.g., [10, §1.6.2].

Lemma 3.1 (Reduced adjoint). *Let $k \in \mathbb{N}$ and $T_k : U_k \times X_k \to W_k^*$ be Fréchet differentiable, and suppose that $S_{k,u}(x)$ arises from the satisfaction of (1.5). If $S_{k,u}$ is Fréchet differentiable in a neighbourhood of x, then*

$$E_k'(x) = wT_k^{(x)}(S_{k,u}(x), x),$$

where $w \in W_k \hookrightarrow W_k^{**}$ (assuming sufficient regularity that a solution in W_k exists) is a solution of the reduced adjoint

$$wT_k^{(u)}(S_{k,u}(x), x) + J_k'(S_{k,u}(x)) = 0. \tag{3.2}$$

Proof. By implicit differentiation and (1.5) holding for x in neighbourhood of x, we get the basic adjoint

$$0 = T_k^{(u)}(S_{k,u}(x), x)S_{k,u}'(x) + T_k^{(x)}(S_{k,u}(x), x), \tag{3.3}$$

where $S_{k,u}'(x) \in \mathbb{L}(X_k; U_k)$, $T_k^{(u)}(S_{k,u}(x), x) \in \mathbb{L}(U_k; W_k^*)$, and $T_k^{(x)}(S_{k,u}(x), x) \in \mathbb{L}(X_k; W_k^*)$. Applying w from the left to (3.3), and $S_{k,u}'(x)$ from the right to (3.2), we deduce that $E_k'(x) = J_k'(S_{k,u}(x))S_{k,u}'(x) = wT_k^{(x)}(S_{k,u}(x), x)$. ◻

In practice, we will take $w^k \in W_k$ as an approximate operator splitting solution of (3.2), instantiated at $u^k \approx S_{k,u}(\check{x}^k)$ in place of $S_{k,u}(\check{x}^k)$, and then set

$$\widetilde{E_k'}(\check{x}^k) := w^k T_k^{(x)}(u^k, \check{x}^k) \approx J_k'(S_{k,u}(\check{x}^k))S_{k,u}'(\check{x}^k) = E_k'(\check{x}^k).$$

When X_k is a real Hilbert space (such as $\mathbb{R}^n$), we write $\widetilde{\nabla E}_k(\breve{x}^k)$ for the Riesz representation of $\widetilde{E_k'}(\breve{x}^k)$.

This is our recursively constructed differential estimate, which we wish to use in Algorithm 2.1. We just need to show that it satisfies Assumption 2.2. We do this in Section 3.3, given *tracking estimates* for u^k and w^k, derived from the contractivity of splitting methods as in [19, 23]. We state those tracking estimates in Section 3.2, and provide examples of their satisfaction in Section 4.

Example 3.2. We apply Lemma 3.1 to the EIT problem discussed in Section 1. Let $\sigma \in L^\infty(\Omega)$ and $U^k \in \mathbb{R}^{N_{\text{elec}}}$ be fixed, and for simplicity assume $N_{\text{inj}} = 1$. Define the mixed state space $Y := H^1(\Omega) \times \mathbb{R}^{N_{\text{elec}}}$, $y = (u, I) \in Y$. Introduce the constraint operator

$$T_k : Y \times L^\infty(\Omega) \to Y^* \quad \langle T_k(y,\sigma), w\rangle := B_\sigma(y, w) - L_{U^k}(w),$$

where $B_\sigma : (H^1(\Omega) \times \mathbb{R}^{N_{\text{elec}}}) \times (H^1(\Omega) \times \mathbb{R}^{N_{\text{elec}}}) \to \mathbb{R}$ for $w = (v, J) \in H^1(\Omega) \times \mathbb{R}^{N_{\text{elec}}}$ is given by

$$B_\sigma((u,I),(v,J)) := \int_\Omega \sigma \nabla u \cdot \nabla v dx + \sum_{i=1}^{N_{\text{elec}}} \frac{1}{\zeta_i} \int_{e_i} uv ds - \sum_{i=1}^{N_{\text{elec}}} \frac{J_i}{\zeta_i} \int_{e_i} u ds + \sum_{i=1}^{N_{\text{elec}}} J_i I_i$$

and $L_{U^k} : H^1(\Omega) \times \mathbb{R}^{N_{\text{elec}}} \to \mathbb{R}$ by

$$L_{U^k}(v, J) := \sum_{i=1}^{N_{\text{elec}}} \frac{1}{\zeta_i} \int_{e_i} U_i (v - J_i) ds. \tag{3.4}$$

The weak solution of (1.1), i.e. the forward state $S_{k,u}(\sigma) = y(\sigma, U^k)$, satisfies $T_k(S_{k,u}(\sigma), \sigma) = 0$. Moreover $\langle \partial_y T_k(y,\sigma)[h_y], w\rangle = B_\sigma(h_y, w)$ and $\langle \partial_\sigma T_k k(y,\sigma)[h_\sigma], w\rangle = \int_\Omega h_\sigma \nabla u \cdot \nabla v dx$ to all directions $h_y \in Y$ and $h_\sigma \in L^\infty(\Omega)$. Taking $E_k(\sigma) = J_k(S_{k,u}(\sigma))$ with $J_k(y) = \frac{1}{2}\|\Sigma^{-1/2}(I - \mathcal{I}^k)\|_2^2$ gives the data misfit term of (1.3). For this term, Lemma 3.1 yields an adjoint state $w(\sigma, U^k) = (v, V) \in Y$ solving

$$B_\sigma(\tilde{y}, w) = -J_k'(S_{k,u}(\sigma))[\tilde{y}] = \tilde{I}^\top \Sigma^{-1}(\mathcal{I}^k - I(\sigma, U^k)), \quad \text{for all} \quad \tilde{y} = (\tilde{u}, \tilde{I}) \in Y.$$

The differential is then $E_k'(\sigma) = \nabla v \cdot \nabla u$.

3.2 BASIC CONSTRUCTIONS AND ASSUMPTIONS

Recall that our goal is, at each time and iteration index k, to construct $(u^k, w^k) \approx \big(S_{k,u}(\breve{x}^k), S_{k,w}(\breve{x}^k)\big)$, since the latter may be computationally infeasible or expensive to compute. Indeed, $S_{k,u}$ is defined as the solution of an inner problem, such as a PDE or an optimization problem, and solving it exactly at every iteration can be expensive. Therefore, in practice, we work with approximate solutions (u^k, w^k). This naturally introduces approximation errors, whose behavior must be controlled in order to ensure stability and convergence of the overall method. The following principal assumption formalises this requirement in the online setting, and can be seen as a counterpart of the corresponding offline assumptions in [23, 6].

Assumption 3.3. For all $k \in \mathbb{N}$, let the normed space X_k be equipped with a semi-norm $\|\cdot\|_{k,\circ}$, and X_k^* with the corresponding support function (see [6, Section 2])

$$\|x^*\|_{k,*} := \sup\{\langle x^* | x\rangle_{X^*,X} \mid x \in X,\ \|x\|_{k,\circ} \le 1\}.$$

Also let the abstract spaces U_k, and W_k be equipped with the distance functions $d_{U_k} : U_k \times U_k \to [0, \infty]$, and $d_{W_k} : W_k \times W_k \to [0, \infty]$. Let $\Omega_k \subset X_k$, and let $S_{k,u} : X_k \to U_k$ and $S_{k,w} : X_k \to W_k$ denote the inner and adjoint solution maps, respectively. Then, on each iteration $k \in \mathbb{N}$, given $((x^n, \breve{x}^n, u^n, w^n))_{n=0}^k \in \prod_{n=0}^k \Omega_n \times \Omega_n \times U_n \times W_n$ and the prediction $\breve{x}^{k+1} \in \Omega_{k+1}$:

(i) An *inner algorithm* produces $u^{k+1} \in U_{k+1}$ satisfying, for some $\pi_u > 0$, $\kappa_u > 1$, $\varepsilon_{u,k} \geq 0$, *when* $k \geq 1$, the inner tracking inequality

$$\kappa_u d_{U_{k+1}}(u^{k+1}, S_{k+1,u}(\breve{x}^{k+1})) \leq d_{U_k}(u^k, S_{k,u}(\breve{x}^k)) + \pi_u \|x^k - \breve{x}^k\|_{k,\circ} + \varepsilon_{u,k}.$$

(ii) An *adjoint algorithm* produces $w^{k+1} \in W_{k+1}$ satisfying for some $\mu_u, \pi_w > 0$, $\kappa_w > 1$, $\varepsilon_{w,k} \geq 0$, *when* $k \geq 1$, the adjoint tracking inequality

$$\begin{aligned}\kappa_w d_{W_{k+1}}(w^{k+1}, S_{k+1,w}(\breve{x}^{k+1})) \leq d_{W_k}(w^k, S_{k,w}(\breve{x}^k)) + \mu_u d_{U_{k+1}}(u^{k+1}, S_{k+1,u}(\breve{x}^{k+1})) \\ + \pi_w \|x^k - \breve{x}^k\|_{k,\circ} + \varepsilon_{w,k}.\end{aligned}$$

(iii) A *differential transformation*, $(u^{k+1}, w^{k+1}) \in U_{k+1} \times W_{k+1}$, given produces $\widetilde{E'_{k+1}}(\breve{x}^{k+1}) \in X_{k+1}^*$ that satisfies for some $\alpha_u, \alpha_w \geq 0$, $\tilde{\varepsilon}_k \geq 0$ the bound

$$\|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\|_{k+1,*} \leq \alpha_u d_{U_{k+1}}(u^{k+1}, S_{k+1,u}(\breve{x}^{k+1})) + \alpha_w d_{W_{k+1}}(w^{k+1}, S_{k+1,w}(\breve{x}^{k+1})) + \tilde{\varepsilon}_k.$$

Remark 3.4. We observe regarding Assumption 3.3:

1. If $\|\cdot\|_{k,\circ}$ is a norm, then $\|\cdot\|_{k,*}$ is the dual norm.
2. The inner and adjoint tracking conditions (i) and (ii) are parameter change aware contractivity conditions for the inner and adjoint algorithms: if $\breve{x}^k = x^k$ and $S_{k+1,u} = S_{k,u}$ (i.e., there is no data updates between frames), the former reduces to a standard contractivity condition. The differential transformation condition (iii) bounds the construction error of $\widetilde{E'_{k+1}}(\breve{x}^{k+1})$ in terms of the tracking errors of the iterates relative to the corresponding solution maps.
3. The distance functions d_{U_k}, d_{W_k} will in the simplest case be given by norms or semi-norms. More generally, they may be problem-adapted measures of discrepancy between iterates and solution maps. Their exact form plays no role in the theory of this section.
4. The spaces X_k and U_k that depend on the time and iteration index k allow our framework to accommodate dynamic online optimisation [9, 4, 5] with growing data sets.
5. For the inner and adjoint variables, the tracking inequalities are formulated starting from $k \geq 1$, and thus do not explicitly involve the initial iterates. At each iteration, the approximate inner and adjoint variables are computed using the prediction $\breve{x}^k$ and compared against the corresponding exact solution maps $S_{k,u}(\breve{x}^k)$ and $S_{k,w}(\breve{x}^k)$. The role of the prediction is to provide a consistent reference point across iterations, allowing the tracking errors to be propagated even when the underlying spaces vary with k. The initial discrepancies at the first iteration enter the final estimates, but can be controlled by computing the first step with sufficiently high accuracy. For contractive algorithms, these initial errors are naturally reduced relative to the initialisation.

3.3 SMOOTHNESS OF DIFFERENTIAL ESTIMATES

In this section, we derive descent and Lipschitz type inequalities for the differential estimates $\widetilde{E'_k}$, given the tracking estimates. These results are straightforward consequences of the scalar tracking estimates of Appendix A. We recall that if E' is L-Lipschitz, then it satisfies the three point inequality [2, Corollary 7.2]

$$\langle E'_k(\breve{x}^k) | x - \bar{x} \rangle_{X_k^*, X_k} \geq E_k(x) - E_k(\bar{x}) - \frac{L}{2}\|x - \breve{x}^k\|_{k,\circ}^2, \quad \text{for all} \quad x, \breve{x}^k, \bar{x} \in X_k, k \geq 0, \tag{3.5}$$

Our goal is to derive similar estimates for $\widetilde{E'_k}(\breve{x}^k)$. Indeed, such an estimate arises by summing the claim of the next lemma with (3.5)

Lemma 3.5. *Suppose that Assumption 3.3 holds, and that $((x^n, \breve{x}^n, u^n, w^n))_{n=0}^k \in \prod_{n=0}^k \Omega_n \times \Omega_n \times U_n \times W_n$ as well as $\breve{x}^{k+1} \in \Omega_{k+1}$ for some $k \in \mathbb{N}$. Then, for any $\tilde{\gamma} > 0$, and $x, \bar{x} \in X_{k+1}$, we have*

$$\langle \widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1}) | x - \bar{x} \rangle_{X^*_{k+1}, X_{k+1}} \geq -\frac{\tilde{\gamma}}{2}\|x - \bar{x}\|^2_{k+1,\circ} - \frac{5}{4\tilde{\gamma}}\varsigma_1^2 \|\breve{x}^{k+1} - x^{k+1}\|^2_{k+1,\circ} - \frac{1}{2\tilde{\gamma}}\mathring{e}_k, \tag{3.6}$$

where

$$\mathring{e}_k := \frac{5}{2}\varsigma_1^2 \left(\max\left(\frac{1}{\pi_u}\varepsilon_{u,k}, \frac{1}{\pi_w}\varepsilon_{w,k}\right)\right)^2 + e_{1,k}$$

for some $e_{1,k} \in \mathbb{R}$ and $\varsigma_1 \geq 0$ that, with $\kappa := \min(\kappa_u, \kappa_w)$ and $\overline{\kappa} := \max(\kappa_u, \kappa_w)$, satisfy

$$\sum_{n=0}^k e_{1,n} \leq \Psi_1 := \frac{5}{4}\left(\frac{d^1_{U_1}(u^1, S_{1,u}(\breve{x}^1))}{\pi_u}\left(\frac{\varsigma_1\alpha_u\kappa}{\kappa - 1} + \frac{\varsigma_1\alpha_w\mu_u}{(\kappa-1)^2}\right) + \frac{d^2_{W_1}(w^1, S_{1,w}(\breve{x}^1))}{\pi_w}\left(\frac{\varsigma_1\alpha_w\kappa}{\kappa - 1}\right) + \sum_{n=0}^k \tilde{\varepsilon}_n^2\right) \tag{3.7}$$

and

$$\varsigma_1 \leq \frac{(\alpha_u\pi_u + \alpha_w\pi_w)\kappa\overline{\kappa}}{\kappa - 1} + \frac{\alpha_w\mu_u\pi_u\overline{\kappa}}{(\kappa-1)^2}. \tag{3.8}$$

Proof. By [6, Lemma 2.1 (iii)], we have $\langle x^* | x \rangle_{X^*_k, X_k} \leq \frac{1}{2}\|x^*\|^2_{k,*} + \frac{1}{2}\|x\|^2_{k,\circ}$ for all $x \in X_k$ and $x^* \in X^*_k$. Hence,

$$\langle \widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1}) | x - \bar{x} \rangle_{X^*_{k+1}, X_{k+1}} \geq -\frac{1}{2\tilde{\gamma}}\|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\|^2_{k+1,*} - \frac{\tilde{\gamma}}{2}\|x - \bar{x}\|^2_{k+1,\circ}. \tag{3.9}$$

Let

$$\begin{aligned} d^u_k &:= d_{U_k}(u^k, S_{k,u}(\breve{x}^k)), & \varrho_k &:= \|\breve{x}^k - x^k\|_{k,\circ}, & \varepsilon_k &:= \max\left(\varepsilon_{u,k}/\pi_u, \varepsilon_{w,k}/\pi_w\right), \\ d^w_k &:= d_{W_k}(w^k, S_{k,w}(\breve{x}^k)), & &\text{and} & \widetilde{d}_k &:= \|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\|_{k+1,*}. \end{aligned}$$

Then Assumption 3.3 and our assumptions $((x^n, \breve{x}^n, u^n, w^n))_{n=0}^k \in \prod_{n=0}^k \Omega_n \times \Omega_n \times U_n \times W_n$ as well as $\breve{x}^{k+1} \in \Omega_{k+1}$ imply Assumption A.1 for the same $\tilde{\varepsilon}_k$. Thus, Lemma A.4 gives

$$\begin{aligned} \|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\|^2_{k+1,*} &\leq \frac{5}{4}\varsigma_1^2(\varrho_{k+1} + \varepsilon_{k+1})^2 + e_{1,k} \\ &\leq \frac{5}{2}\varsigma_1^2\varrho_{k+1}^2 + \frac{5}{2}\varsigma_1^2\left(\max\left(\frac{1}{\pi_u}\varepsilon_{u,k}, \frac{1}{\pi_w}\varepsilon_{w,k}\right)\right)^2 + e_{1,k} \\ &= \frac{5}{2}\varsigma_1^2\|\breve{x}^{k+1} - x^{k+1}\|^2_{k+1,\circ} + \mathring{e}_k. \end{aligned}$$

Combining this inequality with (3.9) proves (3.6). The bounds (3.7) and (3.8) follow from Lemmas A.3 and A.4. ◻

The next result, which verifies Assumption 2.2, relates the error $\mathring{e}_k$ to the error expression in the regret Corollary 2.5 for the outer method. We will further develop the complete error expression in Remark 4.11. The result depends on the exact differential each E_k essentially satisfying Assumption 2.2. We will in the next Section 3.4 verify this for quadratic energies.

Theorem 3.6 (Verification of Assumption 2.2). *Let $k \geq 1$. Suppose Assumption 3.3 holds and that $((x^n, \breve{x}^n, u^n, w^n))_{n=0}^{k-1} \in \prod_{n=0}^{k-1} \Omega_n \times \Omega_n \times U_n \times W_n$ as well as $\breve{x}^k \in \Omega_k$ for some $k \in \mathbb{N}$. Further, suppose E'_k is L'_k-Lipschitz with respect to $\|\cdot\|_{k,\circ}$ in the sense that*

$$\|E'_k(u) - E'_k(v)\|_{k,*} \leq L'_k\|u - v\|_{k,\circ} \quad \textit{for all } u, v \in X_k,$$

and that for some constants $\gamma_1, \gamma_2, L_1, L_2 > 0$ and that for some $\delta > 0$, every comparison sequence $\bar{x}^{1:k} \in \mathscr{X}_{1:k} \cap \prod_{n=1}^{k} B(\check{x}^n, \delta)$ and all $x \in B(\bar{x}^k, \delta)$, we have

$$\langle E_k'(\check{x}^k)|x - \bar{x}^k\rangle_{X_k^*,X_k} \ge E_k(x) - E_k(\bar{x}^k) + \gamma_1\|x - \bar{x}^k\|_{k,\circ}^2 - L_1\|x - \check{x}^k\|_{k,\circ}^2, \tag{3.10}$$

and for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$ and all $x \in B(\hat{x}^k, \delta)$ we have

$$\langle E_k'(\check{x}^k) - E_k'(\hat{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} \ge \gamma_2\|x - \hat{x}^k\|_{k,\circ}^2 - L_2\|x - \check{x}^k\|_{k,\circ}^2. \tag{3.11}$$

Let $\mathring{e}_{k-1}$ and ς_1 be as in Lemma 3.5. Then the following is true

(i) For any $\tilde{\gamma} > 0$, Assumption 2.2 (i) holds with

$$\bar{\gamma}_E = -L_k' - \frac{\tilde{\gamma}}{2}(L_k' + 1), \quad \bar{\lambda}_E = \frac{L_k'}{2\tilde{\gamma}} + \frac{5}{4\tilde{\gamma}}\varsigma_1^2, \quad \text{and} \quad \bar{e}_k = \frac{1}{2\tilde{\gamma}}\mathring{e}_{k-1},$$

That is, for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$ and for all $x^k \in X_k$, we have

$$\langle \widetilde{E_k'}(\check{x}^k) - E_k'(\hat{x}^k)|x^k - \hat{x}^k\rangle_{X_k^*,X_k} \ge \bar{\gamma}_E\|x^k - \hat{x}^k\|_{k,\circ}^2 - \bar{\lambda}_E\|x^k - \check{x}^k\|_{k,\circ}^2 - \bar{e}_k. \tag{3.12}$$

(ii) For $0 < \tilde{\gamma} < 2\min\{\gamma_1, \gamma_2\}$, Assumption 2.2 (ii) holds with

$$\hat{\gamma}_E := \gamma_2 - \frac{\tilde{\gamma}}{2}, \quad \gamma_E := \min\{\hat{\gamma}_E, 2\gamma_1 - \tilde{\gamma}\}, \quad \lambda_E := 2L_1 + \frac{5\varsigma_1^2}{2\tilde{\gamma}}, \quad \hat{\lambda}_E := L_2 + \frac{5\varsigma_1^2}{4\tilde{\gamma}}, \quad e_k := \hat{e}_k := \frac{1}{2\tilde{\gamma}}\mathring{e}_{k-1}.$$

That is, for every $\bar{x}^{1:k} \in \mathscr{X}_{1:k} \cap \prod_{n=1}^{N} B(\check{x}^n, \delta)$, and for all $x^k \in B(\bar{x}^k, \delta)$, we have

$$\langle \widetilde{E_k'}(\check{x}^k)|x^k - \bar{x}^k\rangle_{X_k^*,X_k} \ge E_k(x^k) - E_k(\bar{x}^k) + \frac{\gamma_E}{2}\|x^k - \bar{x}^k\|_{k,\circ}^2 - \frac{\lambda_E}{2}\|x^k - \check{x}^k\|_{k,\circ}^2 - e_k. \tag{3.13}$$

and, for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0) \cap B(\check{x}^k, \delta) \times Y_k$ and $x^k \in B(\hat{x}^k, \delta)$,

$$\langle \widetilde{E_k'}(\check{x}^k) - E_k'(\hat{x}^k)|x^k - \hat{x}^k\rangle_{X_k^*,X_k} \ge \hat{\gamma}_E\|x^k - \hat{x}^k\|_{k,\circ}^2 - \hat{\lambda}_E\|x^k - \check{x}^k\|_{k,\circ}^2 - \hat{e}_k. \tag{3.14}$$

Proof. (i): Since Assumption 3.3 holds and $((x^n, \check{x}^n, u^n, w^n))_{n=0}^{k-1} \in \prod_{n=0}^{k-1} \Omega_n \times \Omega_n \times U_n \times W_n$ as well as $\check{x}^k \in \Omega_k$, Lemma 3.5 applied with index $k-1$ together with [6, Lemma 2.1 (iii)] yields that for any $\tilde{\gamma} > 0$, any $x, z \in X_k$,

$$\begin{aligned}\langle \widetilde{E_k'}(\check{x}^k) - E_k'(\check{x}^k)|x - z\rangle_{X_k^*,X_k} &= -\left\langle -\tilde{\gamma}^{-1/2}\left(\widetilde{E_k'}(\check{x}^k) - E_k'(\check{x}^k)\right)\middle|\tilde{\gamma}^{1/2}(x - z)\right\rangle_{X_k^*,X_k} \\ &\ge -\frac{1}{2\tilde{\gamma}}\|\widetilde{E_k'}(\check{x}^k) - E_k'(\check{x}^k)\|_{k,*}^2 - \frac{\tilde{\gamma}}{2}\|x - z\|_{k,\circ}^2 \\ &\ge -\frac{\tilde{\gamma}}{2}\|x - z\|_{k,\circ}^2 - \frac{5}{4\tilde{\gamma}}\varsigma_1^2\|\check{x}^k - x^k\|_{k,\circ}^2 - \frac{1}{2\tilde{\gamma}}\mathring{e}_{k-1}.\end{aligned} \tag{3.15}$$

Since E_k' is L_k'-Lipschitz between the semi-norm $\|\cdot\|_{k,\circ}$ and support function $\|\cdot\|_{k,*}$, and the two satisfy a Cauchy–Schwarz inequality [6, §2], also applying the triangle and Young's inequalities, we obtain

$$\begin{aligned}\langle E_k'(\check{x}^k) - E_k'(\hat{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} &= -\langle E_k'(\hat{x}^k) - E_k'(\check{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} \\ &\ge -\|E_k'(\hat{x}^k) - E_k'(\check{x}^k)\|_{k,*}\|x - \hat{x}^k\|_{k,\circ} \\ &\ge -L_k'\|\check{x}^k - \hat{x}^k\|_{k,\circ}\|x - \hat{x}^k\|_{k,\circ} \\ &\ge -L_k'\left(\|x - \check{x}^k\|_{k,\circ} + \|x - \hat{x}^k\|_{k,\circ}\right)\|x - \hat{x}^k\|_{k,\circ} \\ &= -\frac{L_k'}{2\tilde{\gamma}}\|x - \check{x}^k\|_{k,\circ}^2 - L_k'\left(1 + \frac{\tilde{\gamma}}{2}\right)\|x - \hat{x}^k\|_{k,\circ}^2.\end{aligned} \tag{3.16}$$

Hence,

$$\begin{aligned}\langle \widetilde{E_k'}(\check{x}^k) - E_k'(\hat{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} &= \langle \widetilde{E_k'}(\check{x}^k) - E_k'(\check{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} + \langle E_k'(\check{x}^k) - E_k'(\hat{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} \\ &\geq \langle \widetilde{E_k'}(\check{x}^k) - E_k'(\check{x}^k)|x - \hat{x}^k\rangle_{X_k^*,X_k} \\ &\quad - \frac{L_k'}{2\tilde{\gamma}}\|x - \check{x}^k\|_{k,\circ}^2 - L_k'\left(1 + \frac{\tilde{\gamma}}{2}\right)\|x - \hat{x}^k\|_{k,\circ}^2.\end{aligned}$$

Combining this with (3.15) applied to $x = x^k$ and $z = \hat{x}^k$ gives (3.12) for the stated coefficients.

(ii): To prove (3.13), let $\bar{x}^{1:k} \in \mathcal{X}_{1:k} \cap \prod_{n=1}^N B(\check{x}^n, \delta)$, and $x^k \in B(\bar{x}^k, \delta)$. Write $\langle \widetilde{E_k'}(\check{x}^k)|x^k - \bar{x}^k\rangle_{X_k^*,X_k} = \langle E_k'(\check{x}^k)|x^k - \bar{x}^k\rangle_{X_k^*,X_k} + \langle \widetilde{E_k'}(\check{x}^k) - E_k'(\check{x}^k)|x^k - \bar{x}^k\rangle_{X_k^*,X_k}$, apply (3.15) with $x = x^k$ and $z = \bar{x}^k$ and combine it with (3.10). This gives (3.13). Similarly, taking $x = x^k$ and $z = \hat{x}^k$ in (3.15) and combining it with (3.11) gives (3.14), finishing the proof. □

3.4 Quadratic energies

We prove (3.10) and (3.11) for E_k of the quadratic form (1.2), relevant in applications with Gaussian noise. The arguments rely only on local bounds on the Hessian, combined with with global boundedness of S_k and a global Lipschitz bound on S_k'.

Assumption 3.7 (Local strong convexity and smoothness). Fix $k \in \mathbb{N}$. Let X_k, Y_k be real Hilbert spaces, let $\|\cdot\|_{k,\circ}$ be a seminorm on X_k, and let $\hat{x} \in X_k$ and $S_{k,u} : X_k \to Y_k$ be C^2 on $B(\hat{x}, \delta)$ for some $\delta > 0$. Let $B_k : Y_k \to Y_k$ be bounded, self-adjoint, and positive semidefinite, and define

$$\|y\|_{Y_k,\circ}^2 := \langle B_k y, y\rangle_{Y_k} \quad \text{for all} \quad y \in Y_k.$$

Define

$$E_k(x) := \frac{1}{2}\|S_{k,u}(x)\|_{Y_k,\circ}^2.$$

Assume that $S_{k,u}$ is Fréchet differentiable on X_k and that there exist constants $\bar{L}_{S,k} \geq 0$, $\bar{M}_k \geq 0$, and $\bar{r}_k \geq 0$ such that

$$\|S_{k,u}(x)\|_{Y_k,\circ} \leq \bar{r}_k \quad \text{for all } x \in X_k,$$

$$\|S_{k,u}'(x)h\|_{Y_k,\circ} \leq \bar{L}_{S,k}\|h\|_{k,\circ} \quad \text{for all } x \in X_k,\ h \in X_k,$$

and

$$\|(S_{k,u}'(u) - S_{k,u}'(v))h\|_{Y_k,\circ} \leq \bar{M}_k\|u - v\|_{k,\circ}\|h\|_{k,\circ} \quad \text{for all } u, v \in X_k,\ h \in X_k.$$

Define

$$L_k' := \bar{L}_{S,k}^2 + \bar{M}_k\bar{r}_k.$$

Assume there exist constants $\mu_k > 0$, $L_{S,k} \geq 0$, $M_k \geq 0$, $r_k \geq 0$ such that for all $x \in B(\hat{x}, \delta)$:

$$\|S_{k,u}'(x)h\|_{Y_k,\circ} \geq \mu_k\|h\|_{k,\circ} \quad \text{for all} \quad h \in X_k, \tag{3.17a}$$

$$\|S_{k,u}'(x)h\|_{Y_k,\circ} \leq L_{S,k}\|h\|_{k,\circ} \quad \text{for all} \quad h \in X_k, \tag{3.17b}$$

$$\|S_{k,u}(x)\|_{Y_k,\circ} \leq r_k, \tag{3.17c}$$

$$\|S_{k,u}''(x)[h_1, h_2]\|_{Y_k,\circ} \leq M_k\|h_1\|_{k,\circ}\|h_2\|_{k,\circ} \quad \text{for all} \quad h_1, h_2 \in X_k. \tag{3.17d}$$

Define

$$m_k := \mu_k^2 - M_k r_k, \quad L_k'' := L_{S,k}^2 + M_k r_k,$$

and suppose that $m_k > 0$.

Corollary 3.8 (Global Lipschitz continuity of E_k'). *Suppose that Assumption 3.7 holds. Then E_k is Fréchet differentiable on X_k and*

$$\|E_k'(u) - E_k'(v)\|_{k,*} \le L_k'\|u - v\|_{k,\circ} \quad \textit{for all } u, v \in X_k.$$

Proof. Since $y \mapsto \frac{1}{2}\langle B_k y, y\rangle_{Y_k}$ is Fréchet differentiable and $S_{k,u}$ is Fréchet differentiable on X_k, the chain rule gives

$$E_k'(x)h = \langle B_k S_{k,u}(x), S_{k,u}'(x)h\rangle_{Y_k}.$$

Let $u, v, h \in X_k$. By Assumption 3.7,

$$\|S_{k,u}(u) - S_{k,u}(v)\|_{Y_k,\circ} \le \bar{L}_{S,k}\|u - v\|_{k,\circ}.$$

Therefore

$$\begin{aligned}|(E_k'(u) - E_k'(v))h| &\le \|S_{k,u}(u) - S_{k,u}(v)\|_{Y_k,\circ}\|S_{k,u}'(u)h\|_{Y_k,\circ} + \|S_{k,u}(v)\|_{Y_k,\circ}\|(S_{k,u}'(u) - S_{k,u}'(v))h\|_{Y_k,\circ}\\ &\le \left(\bar{L}_{S,k}^2 + \bar{M}_k\bar{r}_k\right)\|u - v\|_{k,\circ}\|h\|_{k,\circ} = L_k'\|u - v\|_{k,\circ}\|h\|_{k,\circ}.\end{aligned}$$

Taking the supremum over $\|h\|_{k,\circ} \le 1$ gives the claim. □

Remark 3.9 (Twice continous differentiability of the EIT solution operator). The assumption that $S_{k,u}$ be C^2 from the Hilbert space X_k to Y_k is somewhat strict. For $S_{k,u} = (I_i^{j,k})_{i=1,\dots,N_{\text{elec}};\, j=1,\dots,N_{\text{inj}}}$ the solution operator of the EIT problem (1.1), with $Y_k = \mathbb{R}^{N_{\text{inj}}N_{\text{elec}}}$, in [5] we prove that $S_{k,u}$ is C^2 on L^∞. Hence, if X_k is chosen so that $X_k \hookrightarrow L^\infty$ continuously, for instance $X_k = H^s(\Omega)$ with $s > d/2$, then the restriction of $S_{k,u}$ to X_k is Fréchet differentiable as a map from X_k to Y_k. Moreover, the L^∞-based derivative estimates imply corresponding X_k-based estimates through the embedding $X_k \hookrightarrow L^\infty$. In particular, a uniform bound for the second derivative gives the Lipschitz estimate for $S_{k,u}'$ required above.

The next lemma shows that local $\|\cdot\|_\circ$-strong convexity is induced by a uniform lower bound on the derivative $S_{k,u}'$, provided the residual $\|S_{k,u}(x)\|_{Y_k,\circ}$ is sufficiently small, i.e. if $m_k > 0$. This smallness condition corresponds to working near a critical point with small misfit.

Lemma 3.10 (Local strong convexity and smoothness for quadratic E_k). *Suppose that Assumption 3.7 holds. Then for all $x \in B(\hat{x}, \delta)$ and $h \in X_k$,*

$$m_k\|h\|_{k,\circ}^2 \le E_k''(x)[h,h] \le L_k''\|h\|_{k,\circ}^2. \tag{3.18}$$

Proof. Let $x \in B(\hat{x}, \delta)$. By the definition of E_k and the C^2-regularity of $S_{k,u}$ on $B(\hat{x}, \delta)$, E_k is C^2 on $B(\hat{x}, \delta)$ and

$$E_k'(x)h = \langle B_k S_{k,u}(x), S_{k,u}'(x)h\rangle_{Y_k}, \quad \text{and} \quad E_k''(x)[h,h] = \langle B_k S_{k,u}'(x)h, S_{k,u}'(x)h\rangle_{Y_k} + R_k(x)[h,h],$$

where $R_k(x)[h,h] := \langle B_k S_{k,u}(x), S_{k,u}''(x)[h,h]\rangle_{Y_k}$. By the Cauchy–Schwarz inequality for the positive semidefinite bilinear form $(u,v) \mapsto \langle B_k u, v\rangle_{Y_k}$, together with (3.17c) and (3.17d), we have

$$|R_k(x)[h,h]| \le \|S_{k,u}(x)\|_{Y_k,\circ}\|S_{k,u}''(x)[h,h]\|_{Y_k,\circ} \le M_k r_k\|h\|_{k,\circ}^2.$$

Also, by the definition of $\|\cdot\|_{Y_k,\circ}$ and (3.17a)–(3.17b),

$$\mu_k^2\|h\|_{k,\circ}^2 \le \|S_{k,u}'(x)h\|_{Y_k,\circ}^2 \le L_{S,k}^2\|h\|_{k,\circ}^2.$$

Summing yields the claim. □

The following lemma provides the standard upper and lower quadratic models for E_k. These inequalities form the basic building blocks for the three-point estimates.

Lemma 3.11 (Descent and support inequalities). *Suppose that Assumption 3.7 holds for some $\hat{x} \in X_k$. Then for all $x, z \in B(\hat{x}, \delta) \subset X_k$,*

$$E_k(x) \le E_k(z) + \langle E_k'(z)|x - z\rangle_{X_k^*,X_k} + \frac{L_k''}{2}\|x - z\|_{k,\circ}^2, \tag{3.19}$$

$$E_k(x) \ge E_k(z) + \langle E_k'(z)|x - z\rangle_{X_k^*,X_k} + \frac{m_k}{2}\|x - z\|_{k,\circ}^2. \tag{3.20}$$

Proof. Let $x, z \in B(\hat{x}, \delta)$ and set $\varphi(t) = E_k(z + t(x - z))$. Since $B(\hat{x}, \delta)$ is convex, $z + t(x - z) \in B(\hat{x}, \delta)$ for all $t \in [0, 1]$ and thus, since Assumption 3.7 holds, by Lemma 3.10, $\varphi(t)$ is $C^2(0, 1)$, $\varphi'(t) = \langle E_k'(z + t(x - z))|x - z\rangle_{X_k^*,X_k}$, and $\varphi''(t) = E_k''(z + t(x - z))[x - z, x - z]$. Moreover, (3.18) holds for every t and gives $m_k\|x - z\|_{k,\circ}^2 \le \varphi''(t) \le L_k''\|x - z\|_{k,\circ}^2$. Then

$$\begin{aligned}\frac{m_k}{2}\|x - z\|_{k,\circ}^2 = \int_0^1\int_0^s m_k\|x - z\|_{k,\circ}^2 dtds &\le \int_0^1\int_0^s \varphi''(t)dtds \\ &\le \int_0^1\int_0^s L_k''\|x - z\|_{k,\circ}^2 dtds = \frac{L_k''}{2}\|x - z\|_{k,\circ}^2\end{aligned}$$

Since $\int_0^1\int_0^s \varphi''(t)dtds = \varphi(1) - \varphi(0) - \varphi'(0) = E_k(x) - E_k(z) - \langle E_k'(z)|x - z\rangle_{X_k^*,X_k}$, (3.19) – (3.20) follow. ◻

The following lemma shows that combining the local support inequalities with the global Lipschitz bound on E_k' yields a family of three-point inequalities. These inequalities quantify how well the linearization at a possibly nonlocal point z approximates the energy difference between nearby points. The estimates allow for a trade-off between the curvature term and the error $\|x - z\|_{k,\circ}^2$ controlled by free parameters.

Lemma 3.12 (Three-point inequalities). *Suppose that Assumption 3.7 holds for some $\hat{x} \in X_k$. Then for all $z, x \in X_k$,*

$$\langle E_k'(z) - E_k'(\hat{x})|x - z\rangle_{X_k^*,X_k} \ge -L_k'\|z - \hat{x}\|_{k,\circ}\|x - z\|_{k,\circ}.$$

Moreover, for all $z \in X_k$, all $x, \bar{x} \in B(\hat{x}, \delta)$, and any $\theta > 0$,

$$\langle E_k'(z)|x - \bar{x}\rangle_{X_k^*,X_k} \ge E_k(x) - E_k(\bar{x}) + \frac{m_k}{2(1 + \theta)}\|x - \bar{x}\|_{k,\circ}^2 - \frac{(L_k')^2(1 + \theta)}{2m_k\theta}\|x - z\|_{k,\circ}^2. \tag{3.21}$$

Moreover, for all $z \in X_k$, all $x \in B(\hat{x}, \delta)$, and any $\varepsilon > 0$ such that $\varepsilon L_k' < m_k$,

$$\langle E_k'(z) - E_k'(\hat{x})|x - \hat{x}\rangle_{X_k^*,X_k} \ge \left(m_k - \varepsilon L_k'\right)\|x - \hat{x}\|_{k,\circ}^2 - \frac{L_k'}{4\varepsilon}\|x - z\|_{k,\circ}^2. \tag{3.22}$$

Proof. The first estimate follows from the definition of $\|\cdot\|_{k,*}$ and the global Lipschitz continuity of E_k', and the Cauchy–Schwarz inequality between $\|\cdot\|_{k,*}$ and $\|\cdot\|_{k,\circ}$ (see [6, §2])

$$\langle E_k'(z) - E_k'(\hat{x})|x - z\rangle_{X_k^*,X_k} \ge -\|E_k'(z) - E_k'(\hat{x})\|_{k,*}\|x - z\|_{k,\circ} \ge -L_k'\|z - \hat{x}\|_{k,\circ}\|x - z\|_{k,\circ}.$$

To prove (3.21), let $z \in X_k$ and $x, \bar{x} \in B(\hat{x}, \delta)$. Applying (3.20) to $(\bar{x}, x)$ gives

$$\langle E_k'(x)|x - \bar{x}\rangle_{X_k^*,X_k} \ge E_k(x) - E_k(\bar{x}) + \frac{m_k}{2}\|x - \bar{x}\|_{k,\circ}^2.$$

Hence, by the global Lipschitz continuity of E_k' proven in Corollary 3.8,

$$\begin{aligned}\langle E_k'(z)|x - \bar{x}\rangle_{X_k^*,X_k} &= \langle E_k'(x)|x - \bar{x}\rangle_{X_k^*,X_k} + \langle E_k'(z) - E_k'(x)|x - \bar{x}\rangle_{X_k^*,X_k} \\ &\ge E_k(x) - E_k(\bar{x}) + \frac{m_k}{2}\|x - \bar{x}\|_{k,\circ}^2 - L_k'\|x - z\|_{k,\circ}\|x - \bar{x}\|_{k,\circ}.\end{aligned}$$

For any $\theta > 0$, Young's inequality gives

$$L'_k \|x - z\|_{k,\circ} \|x - \bar{x}\|_{k,\circ} \le \frac{m_k \theta}{2(1+\theta)} \|x - \bar{x}\|^2_{k,\circ} + \frac{(L'_k)^2(1+\theta)}{2m_k\theta} \|x - z\|^2_{k,\circ}.$$

Substituting this yields (3.21).

To prove (3.22), apply (3.20) to $(x, \hat{x})$ and to $(\hat{x}, x)$ to obtain

$$\begin{aligned} E_k(x) - E_k(\hat{x}) - \langle E'_k(\hat{x}) | x - \hat{x} \rangle_{X_k^*, X_k} &\ge \frac{m_k}{2} \|x - \hat{x}\|^2_{k,\circ}, \\ E_k(\hat{x}) - E_k(x) - \langle E'_k(x) | \hat{x} - x \rangle_{X_k^*, X_k} &\ge \frac{m_k}{2} \|x - \hat{x}\|^2_{k,\circ}. \end{aligned}$$

Summing these gives

$$\langle E'_k(x) - E'_k(\hat{x}) | x - \hat{x} \rangle_{X_k^*, X_k} \ge m_k \|x - \hat{x}\|^2_{k,\circ}.$$

Therefore, by the global Lipschitz continuity of E'_k,

$$\begin{aligned} \langle E'_k(z) - E'_k(\hat{x}) | x - \hat{x} \rangle_{X_k^*, X_k} &= \langle E'_k(x) - E'_k(\hat{x}) | x - \hat{x} \rangle_{X_k^*, X_k} + \langle E'_k(z) - E'_k(x) | x - \hat{x} \rangle_{X_k^*, X_k} \\ &\ge m_k \|x - \hat{x}\|^2_{k,\circ} - L'_k \|x - z\|_{k,\circ} \|x - \hat{x}\|_{k,\circ}. \end{aligned}$$

Young's inequality gives

$$L'_k \|x - z\|_{k,\circ} \|x - \hat{x}\|_{k,\circ} \le \varepsilon L'_k \|x - \hat{x}\|^2_{k,\circ} + \frac{L'_k}{4\varepsilon} \|x - z\|^2_{k,\circ}.$$

Substituting this yields (3.22). ◻

We conclude with a corollary that explicitly proves the assumptions of Theorem 3.6 regarding E'_k. Thus combined with Theorem 3.6 , the next result completes the verification of Assumption 2.2 for the estimates $\widetilde{E'_k}(\breve{x}^k)$.

Corollary 3.13 (Instantiation of three-point inequalities). *Let $\delta > 0$ and suppose that Assumption 3.7 holds with center $\hat{x}^k$ for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$, and with center $\bar{x}^k$ for the kth component $\bar{x}^k$ of every comparison sequence $\bar{x}^{1:k} \in \mathcal{X}_{1:k}$ satisfying $\bar{x}^k \in B(\breve{x}^k, \delta)$, with the same constants m_k and L'_k. Then E'_k is globally Lipschitz continuous and the three-point inequalities (3.10) and (3.11) from Theorem 3.6 hold on $B(\hat{x}^k, \delta)$ and on $B(\bar{x}^k, \delta)$ with the following explicit constants:*

(i) For any $\theta > 0$, for every comparison sequence satisfying $\bar{x}^{1:N} \in \mathcal{X}_{1:N} \cap \prod_{i=1}^N B(\breve{x}^i, \delta)$, for all $x \in B(\bar{x}^k, \delta)$ and all $z \in X_k$,

$$\langle E'_k(z) | x - \bar{x}^k \rangle_{X_k^*, X_k} \ge E_k(x) - E_k(\bar{x}^k) + \frac{m_k}{2(1+\theta)} \|x - \bar{x}^k\|^2_{k,\circ} - \frac{(L'_k)^2(1+\theta)}{2m_k\theta} \|x - z\|^2_{k,\circ}.$$

In particular, if $x^k \in B(\bar{x}^k, \delta)$, choosing $x = x^k$ and $z = \breve{x}^k$ yields (3.10) with

$$\gamma_1 := \frac{m_k}{2(1+\theta)}, \quad L_1 := \frac{(L'_k)^2(1+\theta)}{2m_k\theta}, \tag{3.23}$$

i.e.

$$\langle E'_k(\breve{x}^k) | x^k - \bar{x}^k \rangle_{X_k^*, X_k} \ge E_k(x^k) - E_k(\bar{x}^k) + \gamma_1 \|x^k - \bar{x}^k\|^2_{k,\circ} - L_1 \|x^k - \breve{x}^k\|^2_{k,\circ}.$$

(ii) Let $\varepsilon > 0$ satisfy $\varepsilon L_k' < m_k$. Then for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$, for all $x \in B(\hat{x}^k, \delta)$ and all $z \in X_k$,

$$\langle E_k'(z) - E_k'(\hat{x}^k) | x - \hat{x}^k \rangle_{X_k^*, X_k} \geq \left(m_k - \varepsilon L_k'\right) \|x - \hat{x}^k\|_{k,\circ}^2 - \frac{L_k'}{4\varepsilon} \|x - z\|_{k,\circ}^2.$$

In particular, if $x^k \in B(\hat{x}^k, \delta)$, choosing $x = x^k$ and $z = \breve{x}^k$ yields (3.11) with

$$\gamma_2 := m_k - \varepsilon L_k', \quad L_2 := \frac{L_k'}{4\varepsilon}, \tag{3.24}$$

i.e.

$$\langle E_k'(\breve{x}^k) - E_k'(\hat{x}^k) | x^k - \hat{x}^k \rangle_{X_k^*, X_k} \geq \gamma_2 \|x^k - \hat{x}^k\|_{k,\circ}^2 - L_2 \|x^k - \breve{x}^k\|_{k,\circ}^2.$$

Proof. The Lipschitz continuity is proven by Corollary 3.8. Next, apply Lemma 3.12. For (i), fix a comparison sequence $\bar{x}^{1:k} \in \mathscr{X}_{1:k}$ whose kth component $\bar{x}^k$ satisfies $\bar{x}^k \in B(\breve{x}^k, \delta)$. Apply Lemma 3.12 with center $\bar{x}^k$, take $z = \breve{x}^k$ and $\bar{x} = \bar{x}^k$, and read off γ_1 and L_1 as in (3.23). For (ii), fix any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$. Apply Lemma 3.12 with center $\hat{x}^k$, take $z = \breve{x}^k$, and read off γ_2 and L_2 as in (3.24). □

4 CONSTRUCTION OF ONLINE INNER AND ADJOINT ALGORITHMS

In this section, we present inner and adjoint algorithms that satisfy Assumption 3.3. Let U_k and W_k be normed spaces, and let

$$d_{U_k}(u, \tilde{u}) = \|u - \tilde{u}\|_{U_k} \quad \text{and} \quad d_{W_k}(w, \tilde{w}) = \|w - \tilde{w}\|_{W_k}.$$

To prove Assumption 3.3 (i) and (ii), our strategy is to combine contractivity properties of the algorithms with stability properties of the predictors and solution mappings. More precisely, we use:

1. **Non-parametric inner contractivity:** The inner iterate u^{k+1} is computed from u^k by single or multiple steps of a standard optimisation or linear system splitting algorithm. This algorithm satisfies, for some $\bar{\kappa}_u > 1$, the contractivity

$$\bar{\kappa}_u \|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}} \leq \|\breve{u}^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}}. \tag{4.1}$$

   This ensures that each inner iteration advances towards the inner solution $S_{k+1,u}(\breve{x}^{k+1})$ for the parameter $\breve{x}^{k+1}$, and the current data embedded in the inner objective.

2. **Non-parametric adjoint contractivity:** Similarly, the adjoint variable w^{k+1} is computed from w^k by an algorithm. This algorithm satisfies for some $\bar{\kappa}_w > 1$ and $L_w \geq 0$ the contractivity

$$\bar{\kappa}_w \|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \leq \|\breve{w}^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} + L_w \|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}}. \tag{4.2}$$

3. **Bounded Lipschitz predictors:** Introduce the outer primal-dual, inner, anad adjoint predictor mappings

$$P_k : \Omega_{k-1} \to \Omega_k, \quad Q_k : U_{k-1} \to U_k, \quad \tilde{Q}_k : W_{k-1} \to W_k.$$

   Then

$$\breve{x}^k := P_k(x^{k-1}), \quad \breve{u}^k := Q_k(u^{k-1}), \quad \breve{w}^k := \tilde{Q}_k(w^{k-1}).$$

   These predictors describe how iterates and auxiliary variables are transported between successive time steps and data frames. We will generally assume these predictors to be bounded and Lipschitz. More precisely, the predictor Q_k is L_Q-Lipschitz:

$$\|Q_k(u) - Q_k(v)\|_{U_k} \leq L_Q \|u - v\|_{U_{k-1}}, \quad \text{for all} \quad u, v \in U_{k-1}. \tag{4.3}$$

   Also the adjoint predictor $\tilde{Q}_k : W_k \to W_{k+1}$ be $L_{\tilde{Q}}$-Lipschitz, i.e.,

$$\|\tilde{Q}_k(w_1) - \tilde{Q}_k(w_2)\|_{W_k} \leq L_{\tilde{Q}} \|w_1 - w_2\|_{W_{k-1}} \quad \text{for all} \quad w_1, w_2 \in W_{k-1}. \tag{4.4}$$

4. **Lipschitz solution mappings:** The inner problem solution mapping $S_{k,u}$ is L_{S_u}-Lipschitz on Ω_k, i.e.,

$$\|S_{k,u}(x) - S_{k,u}(\tilde{x})\|_{U_k} \leq L_{S_u}\|x - \tilde{x}\|_{k,\circ}, \quad \text{for all} \quad x, \tilde{x} \in \Omega_k. \tag{4.5}$$

Likewise, the adjoint solution mapping $S_{k,w}$ is L_{S_w}-Lipschitz on Ω_k, i.e.,

$$\|S_{k,w}(x) - S_{k,w}(\tilde{x})\|_{W_k} \leq L_{S_w}\|x - \tilde{x}\|_{k,\circ} \quad \text{for all} \quad x, \tilde{x} \in \Omega_k, \tag{4.6}$$

These conditions quantify the stability of the inner and adjoint solution maps with respect to perturbations in the parameter x.

We proceed with this plan by deriving general results in Sections 4.1 and 4.2, for inner and adjoint algorithms, respectively. In Section 4.3, we then treat the differential tarnsformation Assumption 3.3 (iii). Finally, in Section 4.4 we concentrate on the specific case linear system splitting methods for the inner and adjoint PDE, relevant to our EIT application.

4.1 COMPOSED ESTIMATES FOR INNER ALGORITHMS

The following lemma proves Assumption 3.3 (i) subject to the errors $\varepsilon_{u,k}$ still having an unquantified form, which will depend on properties of the predictor Q_k that we analyse afterwards.

Lemma 4.1. *For all $k \geq 1$, assume that (4.3) (Lipschitz inner predictor Q_k) and (4.5) (Lipschitz inner solution mapping $S_{k,u}$) hold. Also suppose we are given an inner algorithm that, given $\left((x^n, \check{x}^n, u^n)\right)_{n=0}^k \in \prod_{n=0}^k \Omega_n \times \Omega_n \times U_n$ produces $u^{k+1} \in U_{k+1}$ satisfying for a constant $\bar{\kappa}_u > 1$ the basic contractivity (4.1). Then Assumption 3.3 (i) holds, more precisely*

$$\frac{\bar{\kappa}_u}{L_Q}\|u^{k+1} - S_{k+1,u}(\check{x}^{k+1})\|_{U_{k+1}} \leq \|u^k - S_{k,u}(\check{x}^k)\|_{U_k} + L_{S_u}\|\check{x}^k - x^k\|_{k,\circ} + \varepsilon_{u,k}$$

for

$$\varepsilon_{u,k} := L_Q^{-1}\|Q_{k+1}(S_{k,u}(x^k)) - S_{k+1,u}(P_{k+1}(x^k))\|_{U_{k+1}}.$$

Proof. Using the definition of $\check{u}^{k+1}$, we get

$$\check{u}^{k+1} - S_{k+1,u}(\check{x}^{k+1}) = [Q_{k+1}(u^k) - Q_{k+1}(S_{k,u}(\check{x}^k))] + [Q_{k+1}(S_{k,u}(\check{x}^k)) - S_{k+1,u}(\check{x}^{k+1})].$$

Then

$$\begin{aligned}
\|\check{u}^{k+1} - S_{k+1,u}(\check{x}^{k+1})\|_{U_{k+1}} &\leq \|Q_{k+1}(u^k) - Q_{k+1}(S_{k,u}(\check{x}^k))\|_{U_{k+1}} \\
&\quad + \|Q_{k+1}(S_{k,u}(\check{x}^k)) - S_{k+1,u}(\check{x}^{k+1})\|_{U_{k+1}} \\
&\leq L_Q\|u^k - S_{k,u}(\check{x}^k)\|_{U_k} + \|Q_{k+1}(S_{k,u}(\check{x}^k)) - Q_{k+1}(S_{k,u}(x^k))\|_{U_{k+1}} \\
&\quad + \|Q_{k+1}(S_{k,u}(x^k)) - S_{k+1,u}(\check{x}^{k+1})\|_{U_{k+1}} \\
&\leq L_Q\|u^k - S_{k,u}(\check{x}^k)\|_{U_k} + L_Q L_{S_u}\|\check{x}^k - x^k\|_{k,\circ} \\
&\quad + \|Q_{k+1}(S_{k,u}(x^k)) - S_{k+1,u}(\check{x}^{k+1})\|_{U_{k+1}}.
\end{aligned} \tag{4.7}$$

Combining (4.1) and (4.7), and using $P_{k+1}(x^k) = \check{x}^{k+1}$, we obtain the claim. □

Remark 4.2 (Inner algorithms). Several standard algorithms satisfy the contractivity assumption (4.1) under appropriate conditions. As our focus is on PDEs, we treat linear system splitting in Section 4.4. For inner optimisation problems, forward-backward splitting is treated in [6].

Remark 4.3 (Remainder term). The error term $\varepsilon_{u,k}$ is a commutativity error between the solution mappings and the predictions of the parameter and the solutions. For Corollary 2.5 we want this term to be at least bounded, preferrably with a bounded sum over $k \in \mathbb{N}$. For identity predictions when $X_{k+1} = X_k$ and $U_{k+1} = U_k$, $\varepsilon_{u,k} = \|S_{k,u}(x^k) - S_{k+1,u}(x^k)\|_{U_{k+1}}$, so the term depends on the continuity properties of the solution mapping with respect to the embedded data, encoded in the index k.

More generally, for a conservative estimate when $X_k = X_{k+1}$ and $U_k = U_{k+1}$, if $S_{k,u}$ is $L_{S,u}$-Lipschitz for each k, we have

$$\begin{aligned}\varepsilon_{u,k} &= L_Q^{-1}\|[Q_{k+1} - \mathrm{Id}](S_{k,u}(x^k)) + S_{k,u}(x^k) - S_{k+1,u}(P_{k+1}(x^k))\|_{U_{k+1}} \\ &\le L_Q^{-1}\|[Q_{k+1} - \mathrm{Id}](S_{k,u}(x^k))\|_{U_{k+1}} + L_Q^{-1}\|S_{k,u}(x^k) - S_{k+1,u}(x^k)\|_{U_{k+1}} + \frac{L_{S,u}}{L_Q}\|(P_{k+1} - \mathrm{Id})(x^k)\|_{k+1,\circ}.\end{aligned}$$

The first and last term can be controlled and even made small if $P_{k+1} \approx \mathrm{Id}$, $Q_{k+1} \approx \mathrm{Id}$, and the solution mappings and iterates are bounded. Regarding the middle term, for parametrised linear systems $T_k(x,u) = A_{k,x}u - b_{k,x}$, we have $S_{k,u}(x) = A_{k,x}^{-1}b_{k,x}$, so can further estimate

$$\begin{aligned}\|S_{k,u}(x^k) - S_{k+1,u}(x^k)\| &= \|A_{k,x^k}^{-1}b_{k,x^k} - A_{k+1,x^k}^{-1}b_{k+1,x^k}\| \\ &= \|(A_{k,x^k}^{-1} - A_{k+1,x^k}^{-1})b_{k,x^k} - A_{k+1,x^k}^{-1}(b_{k+1,x^k} - b_{k,x^k})\| \\ &\le \|(A_{k,x^k}^{-1} - A_{k+1,x^k}^{-1})\|\|b_{k,x^k}\| + \|A_{k+1,x^k}^{-1}\|\|b_{k+1,x^k} - b_{k,x^k}\|.\end{aligned}$$

This can, again, be controlled if the dependence of A_{k,x^k}^{-1} and b_{k,x^k} on the data (encoded in the index k) is continuous.

4.2 COMPOSED ESTIMATES FOR ADJOINT ALGORITHMS

We now repeat the overall idea of the arguments of the previous subsection for Assumption 3.3 (ii).

Lemma 4.4. *For all $k \ge 1$, assume (4.4) (Lipschitz adjoint predictor $\tilde{Q}_k$) and (4.6) (Lipschitz adjoint solution mapping $S_{k,w}$). Also suppose that we are given an adjoint algorithm that, given $(x^n, \breve{x}^n, u^n)_{n=0}^k \subset \prod_{n=0}^k \Omega_n \times \Omega_n \times U_n$ and $(x^{k+1}, u^{k+1}) \in \Omega_{k+1} \times U_{k+1}$, produces a $w^{k+1} \in W_{k+1}$ that satisfies the basic adjoint contractivity condition (4.2) with constants $\bar{\kappa}_w > 1$ and $L_w \ge 0$. Then Assumption 3.3 (ii) holds, more precisely*

$$\begin{aligned}\frac{\bar{\kappa}_w}{L_{\tilde{Q}}}\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \le \|w^k - S_{k,w}(\breve{x}^k)\|_{W_k} &+ \frac{L_w}{L_{\tilde{Q}}}\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}} \\ &+ L_{S_w}\|\breve{x}^k - x^k\|_{X_k} + \varepsilon_{w,k}\end{aligned}$$

for

$$\varepsilon_{w,k} = L_{\tilde{Q}}^{-1}\|\tilde{Q}_{k+1}(S_{k,w}(x^k)) - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}}.$$

Proof. Using the definition of $\breve{w}^{k+1}$, we get

$$\breve{w}^{k+1} - S_{k+1,w}(\breve{x}^{k+1}) = [\tilde{Q}_{k+1}(w^k) - \tilde{Q}_{k+1}(S_{k,w}(\breve{x}^k))] + [\tilde{Q}_{k+1}(S_{k,w}(\breve{x}^k)) - S_{k+1,w}(\breve{x}^{k+1})].$$

Using that $\tilde{Q}_{k+1}$ is $L_{\tilde{Q}}$-Lipschitz, we obtain

$$
\begin{aligned}
(4.8)\quad \|\breve{w}^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} &\le \|\tilde{Q}_{k+1}(w^k) - \tilde{Q}_{k+1}(S_{k,w}(\breve{x}^k))\|_{W_{k+1}} \\
&\quad + \|\tilde{Q}_{k+1}(S_{k,w}(\breve{x}^k)) - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \\
&\le L_{\tilde{Q}}\|w^k - S_{k,w}(\breve{x}^k)\|_{W_k} + \|\tilde{Q}_{k+1}(S_{k,w}(\breve{x}^k)) - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \\
&\le L_{\tilde{Q}}\|w^k - S_{k,w}(\breve{x}^k)\|_{W_k} + \|\tilde{Q}_{k+1}(S_{k,w}(\breve{x}^k)) - \tilde{Q}_{k+1}(S_{k,w}(x^k))\|_{W_{k+1}} \\
&\quad + \|\tilde{Q}_{k+1}(S_{k,w}(x^k)) - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \\
&\le L_{\tilde{Q}}\|w^k - S_{k,w}(\breve{x}^k)\|_{W_k} + L_{\tilde{Q}}L_{S_w}\|\breve{x}^k - x^k\|_{k,\circ} \\
&\quad + \|\tilde{Q}_{k+1}(S_{k,w}(x^k)) - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}},
\end{aligned}
$$

where the third inequality follows by adding and subtracting the term $\tilde{Q}_{k+1}(S_{k,w}(x^k))$, the fourth inequality follows from the Lipschitz property of $\tilde{Q}_{k+1}$ and $S_{k,w}$. Combining (4.2) and (4.8) we obtain the claim. □

Remark 4.5 (Remainder term). Similarly to Remark 4.3, the error term $\varepsilon_{w,k}$ is a commutativity error between the adjoint solution mappings and the predictions of the parameter and the solutions. For Corollary 2.5, we want this term to be at least bounded, preferrably with a bounded sum over $k \in \mathbb{N}$. For identity predictions when $X_{k+1} = X_k$ and $U_{k+1} = U_k$, $\varepsilon_{w,k} = \|S_{k,w}(x^k) - S_{k+1,w}(x^k)\|_{W_{k+1}}$, so the term depends on the continuity properties of the solution mapping with respect to the embedded data, encoded in the index k. More generally, for a conservative estimate when $X_k = X_{k+1}$ and $U_k = U_{k+1}$, if $S_{k,w}$ is $L_{S,w}$-Lipschitz for each k, we have

$$
\begin{aligned}
\varepsilon_{w,k} &= L_{\tilde{Q}}^{-1}\|[\tilde{Q}_{k+1} - \mathrm{Id}](S_{k,w}(x^k)) + S_{k,w}(x^k) - S_{k+1,w}(P_{k+1}(x^k))\|_{W_{k+1}} \\
&\le L_{\tilde{Q}}^{-1}\|[\tilde{Q}_{k+1} - \mathrm{Id}](S_{k,w}(x^k))\|_{W_{k+1}} + L_{\tilde{Q}}^{-1}\|S_{k,w}(x^k) - S_{k+1,w}(x^k)\|_{W_{k+1}} \\
&\quad + L_{\tilde{Q}}^{-1}L_{S,w}\|(P_{k+1} - \mathrm{Id})(x^k)\|_{k+1,\circ}.
\end{aligned}
$$

The first and last term can be controlled and even made small if $\tilde{Q}_{k+1} \approx \mathrm{Id}$, $P_{k+1} \approx \mathrm{Id}$, and the solution mappings and iterates are bounded. Regarding the middle term, for parametrised linear systems $T_k(x,u) = A_{k,x}u - b_{k,x}$, we have $S_{k,w}(x) = (A_{k,x}^*)^{-1}J_k'(S_{k,u}(x))$, so can further estimate

$$
\begin{aligned}
&\|S_{k,w}(x^k) - S_{k+1,w}(x^k)\| = \|(A_{k,x}^*)^{-1}J_k'(S_{k,u}(x)) - (A_{k+1,x}^*)^{-1}J_k'(S_{k+1,u}(x))\| \\
&= \|((A_{k,x}^*)^{-1} - (A_{k+1,x}^*)^{-1})J_k'(S_{k,u}(x)) - (A_{k+1,x}^*)^{-1}\left(J_k'(S_{k+1,u}(x)) - J_k'(S_{k,u}(x))\right)\| \\
&\le \|(A_{k,x}^*)^{-1} - (A_{k+1,x}^*)^{-1}\|\|J_k'(S_{k,u}(x))\| + \|(A_{k+1,x}^*)^{-1}\|\|J_k'(S_{k+1,u}(x)) - J_k'(S_{k,u}(x))\|.
\end{aligned}
$$

This can, again, be controlled if the dependence of $(A_{k,x}^*)^{-1}$ and $J_k'(S_{k,u}(x))$ on the data (encoded in the index k) is continuous.

4.3 THE DIFFERENTIAL TRANSFORMATION

Finally, we provide conditions for the differential transformation component of Assumption 3.3,(iii) to hold for the reduced adjoint. As before, we omit subscripts in the norm notation and understand $\|.\|$ as the norm of the respective underlying space.

Lemma 4.6 (Differential transformation: reduced adjoint). *Continuing from Lemma 3.1, take $S_{k,w}(x) = w$ as a solution of (3.2). Suppose $T_k^{(x)}(\,\cdot\,,x)$ is $L_{T^{(x)};u}$-Lipshitz for all $x \in \Omega_k$ and $k \in \mathbb{N}$; that*

$$
\sup_{k,\in\mathbb{N},\, u\in U_k,\, x\in\Omega_k} \|T_k^{(x)}(u,x)\| =: M_{T^{(x)}} < \infty
$$

and that the adjoint solution mappings are bounded on Ω_k, i.e.,

$$N_{S_{k,w}} := \sup_{k\in\mathbb{N},\, x\in\Omega_k} \|S_{k,w}(x)\| < \infty.$$

Also let

$$\widetilde{E'_{k+1}}(\breve{x}^{k+1}) := w^{k+1}T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1}).$$

Then the differential transformation Assumption 3.3 (iii) holds with $\alpha_u = N_{S_{k,w}}L_{T^{(x)};u}$, $\alpha_w = M_{T^{(x)}}$, and $\tilde{\varepsilon}_k = 0$.

Proof. Write $\hat{w}^{k+1} := S_{k+1,w}(\breve{x}^{k+1})$. Since

$$E'_{k+1}(\breve{x}^{k+1}) = J'_{k+1}(S_{k+1,u}(\breve{x}^{k+1}))S'_{k+1,u}(\breve{x}^{k+1}) = \hat{w}^{k+1}T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1}),$$

it follows

$$\begin{aligned}
\|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\| &= \|w^{k+1}T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1}) - \hat{w}^{k+1}T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1})\| \\
&= \|[w^{k+1} - \hat{w}^{k+1}]T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1}) - \hat{w}^{k+1}[T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1}) - T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1})]\| \\
&\le \|T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1})\|\|w^{k+1} - \hat{w}^{k+1}\| \\
&\quad + \|\hat{w}^{k+1}\|\|T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1}) - T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1})\|
\end{aligned}$$

Therefore, recalling that $S_{k+1,w}(\breve{x}^{k+1}) = \hat{w}^{k+1}$, we prove Assumption 3.3 (iii):

$$\|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\| \le N_{S_{k+1,w}}L_{T^{(x)};u}\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\| + M_{T^{(x)}}\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|. \quad \square$$

4.4 linear system splitting

We now cover specific algorithms for PDEs as the inner problems. For U_k, X_k, and W_k^* normed spaces, let both $A_{k,x} \in \mathbb{L}(U_k; W_k^*)$, modelling a linear PDE parametrised x, and the right hand side $b_{k,x} \in W_k^*$ be Lipschitz in $x \in X_k$. Consider the inner constraint of $u = S_{k,u}(x)$ satisfying

$$A_{k,x}u = b_{k,x}. \tag{4.9}$$

This is again an instance of (1.5) when we set

$$T_k(u, x) = A_{k,x}u - b_{k,x}.$$

Two solve (4.9) inexactly and efficiently, we split $A_{k,x}$ into two components, as in standard Gauss–Seidel or Jacobi splittings.

Assumption 4.7 (Admissible splitting). Let X_k, U_k, and W_k^* be normed spaces.For all $k \ge 0$ and $x^k \in \Omega \subset X_k$, let $A_{k,x^k} \in \mathbb{L}(U_k; W_k^*)$. We assume to be given splittings $A_{k,\breve{x}^k} = N_{k,\breve{x}^k} + M_{k,\breve{x}^k}$, where $N_{k,\breve{x}^k}$ is invertible and

$$\zeta_u\|N^{-1}_{k,\breve{x}^k}M_{k,\breve{x}^k}\| \le 1$$

for some $\zeta_u > 1$.

Assumption 4.8 (Adjoint admissible splitting). Let X_k, U_k, and W_k^* be normed spaces.For all $k \ge 0$ and $x^k \in \Omega \subset X_k$, let $A_{u,k,\breve{x}^k} \in \mathbb{L}(W_k; U_k^*)$. We assume to be given splittings $A_{u,k,\breve{x}^k} = N_{u,k,\breve{x}^k} + M_{u,k,\breve{x}^k}$, where $N_{k,\breve{x}^k}$ is invertible and

$$\zeta_w\|N^{-1}_{u,k,\breve{x}^k}M_{u,k,\breve{x}^k}\| \le 1$$

for some $\zeta_w > 1$, and that $\|N^{-1}_{u,k,\breve{x}^k}\| \le \gamma$ for some $\gamma \ge 0$.

Example 4.9 (Gauss–Seidel splitting [23, Example 4.4]). Let $A_{k,x^k} \in \mathbb{R}^{n\times n}$, and take N_{k,x^k} as the upper triangle and diagonal of A_{k,x^k}. We have $\|N_{k,x^k}^{-1}M_{k,x^k}\| \le \zeta$ for some $\zeta \in [0,1)$ when A_{k,x^k} is either strictly diagonally dominant or symmetric and positive definite [7, §10.1], thus Assumption 4.7 holds. We also have $\|N_{k,x^k}^{-1}\| \le \gamma$ for some $\gamma \ge 0$ when N_{k,x^k} is invertible, i.e., has no zero on the main diagonal.

Theorem 4.10. *The following hold:*

(i) *Suppose Assumption 4.7 holds for $A_{k,\breve{x}^k} = N_{k,\breve{x}^k} + M_{k,\breve{x}^k}$, and that $S_{k,u}(x) = A_{k,x}^{-1}b_{k,x}$ is L_{S_u}-Lipschitz in Ω for all $k \ge 0$. Assume that Q_k is L_Q-Lipschitz for all $k \ge 0$. Then, the inner updates*

$$u^{k+1} = N_{k+1,\breve{x}^{k+1}}^{-1}(b_{k+1,\breve{x}^{k+1}} - M_{k+1,\breve{x}^{k+1}}\breve{u}^{k+1})$$

satisfy the contractivity condition (4.1). In particular, Assumption 3.3 (i) holds, i.e.,

$$L_Q^{-1}\zeta_u\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\| \le \|u^k - S_{k,u}(\breve{x}^k)\| + L_{S_u}\|\breve{x}^k - x^k\| + \varepsilon_{u,k}$$

for

$$\varepsilon_{u,k} := L_Q^{-1}\|Q_k(S_{k,u}(x^k)) - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}}.$$

(ii) *Suppose Assumption 4.8 holds for $A_{k,x}^* = N_{*,k,\breve{x}^k} + M_{*,k,\breve{x}^k}$. Assume that, for all $k \ge 0$, $\tilde{Q}_k$ is $L_{\tilde{Q}}$-Lipschitz, $S_{k,w}$ is L_{S_w}-Lipschitz, and J_{k+1}' is L_J-Lipschitz. Then, the adoint updates*

$$w^{k+1} = -N_{*,k,\breve{x}^{k+1}}^{-1}\left(J_{k+1}'(S_{k+1,u}(\breve{x}^{k+1})) + M_{*,k+1,\breve{w}^{k+1}}\breve{w}^{k+1}\right)$$

satisfy the contractivity condition (4.2). In particular, Assumption 3.3 (ii) holds, i.e.,

$$\begin{aligned}\frac{1}{L_{\tilde{Q}}\gamma L_J}\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \le \|w^k - S_{k,w}(\breve{x}^k)\|_{W_k} + \frac{1}{L_{\tilde{Q}}L_J\gamma\zeta_w}\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}} \\ + L_{S_w}\|\breve{x}^k - x^k\|_{X_k} + \varepsilon_{w,k}\end{aligned}$$

for

$$\varepsilon_{w,k} = L_{\tilde{Q}}^{-1}\|\tilde{Q}_k(S_{k,w}(x^k)) - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}}.$$

(iii) *Assume that $u \mapsto A_{k+1,\breve{x}^{k+1}}^{(x)}u - b_{k+1,\breve{x}^{k+1}}^{(x)}$ is $L_{T,u}$-Lipschitz, and $\|A_{k+1,\breve{x}^{k+1}}^{(x)}u^{k+1} - b_{k+1,\breve{x}^{k+1}}^{(x)}\| \le M_{T^{(x)}}$. Then, Assumption 3.3 (iii) holds:*

$$\|\widetilde{E_{k+1}'}(\breve{x}^{k+1}) - E_{k+1}'(\breve{x}^{k+1})\| \le M_{T^{(x)}}\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\| + L_{T,u}\|\hat{w}^{k+1}\|\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|.$$

Proof. (i): We have

$$\begin{aligned}u^{k+1} - S_{k+1,u}(\breve{x}^{k+1}) &= N_{k+1,\breve{x}^{k+1}}^{-1}(b_{k+1,\breve{x}^{k+1}} - M_{k+1,\breve{x}^{k+1}}\breve{u}^{k+1}) \\ &\quad - N_{k+1,\breve{x}^{k+1}}^{-1}(b_{k+1,\breve{x}^{k+1}} - M_{k+1,\breve{x}^{k+1}}S_{k+1,u}(\breve{x}^{k+1})) \\ &= -N_{k+1,\breve{x}^{k+1}}^{-1}M_{k+1,\breve{x}^{k+1}}(\breve{u}^{k+1} - S_{k+1,u}(\breve{x}^{k+1})).\end{aligned}$$

Hence,

$$\zeta_u\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\| \le \zeta_u\|N_{k+1,\breve{x}^{k+1}}^{-1}M_{k+1,\breve{x}^{k+1}}\|\|\breve{u}^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\| \le \|\breve{u}^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|.$$

Thus, using Lemma 4.1, Assumption 3.3 (i) holds.

(ii): For brevity, set $v_k = (*, k, \breve{x}^k)$. Hence, using the definition of w^{k+1}, and the fact that

$$(M_{v_{k+1}} + N_{v_{k+1}})S_{k+1,w}(\breve{x}^{k+1}) = -J'_{k+1}(u^{k+1}),$$

we obtain

$$\begin{aligned}
w^{k+1} - S_{k+1,w}(\breve{x}^{k+1}) &= -N_{v_{k+1}}^{-1}\left(J'_{k+1}(S_{k+1,u}(\breve{x}^{k+1})) + M_{v_{k+1}}\breve{w}^{k+1}\right) - S_{k+1,w}(\breve{x}^{k+1}) \\
&= -N_{v_{k+1}}^{-1}\left(J'_{k+1}(S_{k+1,u}(\breve{x}^{k+1})) - J'_{k+1}(u^{k+1})\right) - N_{v_{k+1}}^{-1}M_{v_{k+1}}\breve{w}^{k+1} \\
&\quad - N_{v_{k+1}}^{-1}J'_{k+1}(u^{k+1}) - S_{k+1,w}(\breve{x}^{k+1}) \\
&= -N_{v_{k+1}}^{-1}\left(J'_{k+1}(S_{k+1,u}(\breve{x}^{k+1})) - J'_{k+1}(u^{k+1})\right) - N_{v_{k+1}}^{-1}M_{v_{k+1}}\breve{w}^{k+1} \\
&\quad + N_{v_{k+1}}^{-1}\left((M_{v_{k+1}} + N_{v_{k+1}})S_{k+1,w}(\breve{x}^{k+1})\right) - S_{k+1,w}(\breve{x}^{k+1}) \\
&= -N_{v_{k+1}}^{-1}\left(J'_{k+1}(S_{k+1,u}(\breve{x}^{k+1})) - J'_{k+1}(u^{k+1})\right) + N_{v_{k+1}}^{-1}M_{v_{k+1}}\left(S_{k+1,u}(\breve{x}^{k+1}) - \breve{w}^{k+1}\right).
\end{aligned}$$

Thus, using the Lipschitz properties of J'_{k+1} and Assumption 4.8, we obtain

$$\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}} \le \gamma L_J\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|_{U_{k+1}} + \frac{1}{\zeta_w}\|\breve{w}^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\|_{W_{k+1}}.$$

This proves our contractivity condition (4.2) with $\bar{\kappa}_w = \frac{1}{L_J\gamma}$ and $L_w = \frac{1}{L_J\gamma\zeta_w}$. Thus, using Lemma 4.4, Assumption 3.3 (ii) holds.

(iii): Finally, we have

$$\begin{aligned}
&\|\widetilde{E'_{k+1}}(\breve{x}^{k+1}) - E'_{k+1}(\breve{x}^{k+1})\| = \|w^{k+1}T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1}) - S_{k+1,w}(\breve{x}^{k+1})T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1})\| \\
&= \|[w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})]T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1}) - \hat{w}^{k+1}[T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1}) - T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1})]\| \\
&\le \|T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1})\|\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\| \\
&\quad + \|\hat{w}^{k+1}\|\|T^{(x)}_{k+1}(S_{k+1,u}(\breve{x}^{k+1}), \breve{x}^{k+1}) - T^{(x)}_{k+1}(u^{k+1}, \breve{x}^{k+1})\| \\
&\le M_{T^{(x)}}\|w^{k+1} - S_{k+1,w}(\breve{x}^{k+1})\| + L_{T,u}\|\hat{w}^{k+1}\|\|u^{k+1} - S_{k+1,u}(\breve{x}^{k+1})\|.
\end{aligned}$$

Thus, Assumption 3.3 (iii) holds. □

Remark 4.11 (Full error expression). In the present setting, we can bound the error

$$e_N^{\Sigma}(u^{0:N-1}, \hat{u}^{0:N}) = \sum_{k=0}^{N-1}\left(\varepsilon^{\dagger}_{k+1}(u^k, \hat{u}^{k:k+1}) + \tau e_{k+1}\right).$$

given in Corollary 2.5, where the predicgtion error $\varepsilon^{\dagger}_{k+1}(u^k, \bar{u}^{k:k+1})$, defined in (2.5), depends on the predictors use, and is discussed in more detail in [4, 5]. From Theorem 3.6 (ii), the algorithm error satisfies $e_{k+1} = \frac{1}{2\tilde{\gamma}}\mathring{e}_k$, where $\mathring{e}_k$ is defined in Lemma 3.5. Moreover, $\mathring{e}_k$ can be bounded as

$$\begin{aligned}
\frac{1}{2\tilde{\gamma}}\sum_{k=0}^{N}\mathring{e}_k &:= \frac{1}{2\tilde{\gamma}}\sum_{k=0}^{N}\left(\frac{5}{2}\varsigma_1^2\left(\max\left(\frac{1}{\pi_u}\,\varepsilon_{u,k}, \frac{1}{\pi_w}\,\varepsilon_{w,k}\right)\right)^2 + e_{1,k}\right) \\
&\le \frac{1}{2\tilde{\gamma}}\sum_{k=0}^{N}\left(\frac{5}{2}\varsigma_1^2\left(\max\left(\frac{1}{\pi_u}\,\varepsilon_{u,k}, \frac{1}{\pi_w}\,\varepsilon_{w,k}\right)\right)^2\right) + \frac{1}{2\tilde{\gamma}}\Psi_1,
\end{aligned}$$

where Ψ_1 and ς_1 are bounded in (3.7) and (3.8) based on the first-step errors and the coefficients $\alpha_u = L_{T,u}$, $\alpha_w = M_{T^{(x)}}$, $\pi_u = L_{S_u}$, and $\pi_w = L_{S_w}$.

When $Q_k = \mathrm{Id}$ and $\tilde{Q}_k = \mathrm{Id}$, as in our numerical experiments of Section 5, we obtain from Remarks 4.3 and 4.5 that

$$
\begin{aligned}
\varepsilon_{u,k} &\le \|(A_{k,x^k}^{-1} - A_{k+1,x^k}^{-1})\|\|b_{k,x^k}\| + \|A_{k+1,x^k}^{-1}\|\|b_{k+1,x^k} - b_{k,x^k}\| + L_{S,u}\|(P_k - \mathrm{Id})(x^k)\|,\\
\varepsilon_{w,k} &\le \|(A_{k,x}^*)^{-1} - (A_{k+1,x}^*)^{-1}\|\|J_k'(A_{k,x}^{-1}b_{k,x})\|\\
&\qquad + \|(A_{k+1,x}^*)^{-1}\|\|J_k'(A_{k+1,x}^{-1}b_{k+1,x}) - J_k'(A_{k,x}^{-1}b_{k,x})\| + L_{S,w}\|(P_k - \mathrm{Id})(x^k)\|.
\end{aligned}
$$

These errors can be further bounded through the specific apperance of the frame-dependent data in $A_{k,x}$, $b_{k,x}$, and J_k, as well as the difference of the primal predictor P_k from the identity.

5 NUMERICAL EXPERIMENTS

In this section, in the context of application of the online Algorithm 2.1 to the dynamical electrical impedance tomography (EIT) problem, we compare the single-loop differential estimates of Sections 3 and 4 against the background solver of [5] for construting the gradient estimates $\widetilde{\nabla E}_k(\check{x}^k)$ for the EIT data term from (1.3) (Section 5.3).

In the dynamical EIT problem, the goal is to reconstruct a time-varying conductivity distribution containing moving inclusions. The evolution of the inclusions is governed by a transport equation corresponding to incompressible flow with constant density. In addition to this nominal setting, we deliberately consider scenarios that violate the modelling assumptions in order to assess the robustness of the method.

The numerical experiments, including the EIT forward model, data generation, discretisation choices, and noise model, closely follow the setup introduced in [5]. We therefore only summarise the key ingredients here and refer the reader to that work for full details. Our computations were performed on a 2020 MacBook Air M1 with 16GB memory. Our Julia implementation of the experiments and algorithms is available on Zenodo [14].

5.1 TEST SCENARIOS

All experiments are performed on synthetic data in a disk-shaped domain Ω equipped with $N_{\mathrm{elec}} = 16$ equally spaced boundary electrodes. The data fidelity term E_k defined in (1.3) is evaluated using a finite element approximation of the electrode potentials and the conductivity, both represented with piecewise linear basis functions. To avoid the inverse crime [20], the mesh used for the inverse problem is coarser (2917 nodes) than that used for data generation (5039 nodes). All remaining modelling and discretisation choices follow [5].

For each time instance, current injection is applied sequentially at a single electrode while all remaining electrodes are grounded, resulting in $N_{\mathrm{inj}} = 16$ excitation patterns. As is standard in EIT, the current measured at the excited electrode is excluded from the data, yielding $(N_{\mathrm{elec}} - 1)N_{\mathrm{inj}} = 240$ measurements per frame.

We consider the following four scenarios:

Constant Motion In this baseline experiment, a single inclusion moves with constant velocity in a homogeneous background. This serves as a reference case.

Circular Motion A single inclusion follows a circular trajectory, violating the constant-velocity assumption.

Halting Motion A single inclusion stops completely at frames 1000 and 2000, further challenging the prediction model.

Disappearing Inclusions Two inclusions move along circular paths. One inclusion disappears at frame 500 and the other at frame 1000; both reappear at frame 1500. This scenario violates the incompressibility assumption.

The *Constant Motion* experiment consists of 400 frames, while the remaining experiments contain 2000 frames. The background conductivity is fixed at $x_{\text{bg}} = 1\,\text{S}$, and all inclusions are resistive with conductivity $x_{\text{incl}} = 10^{-4}\,\text{S}$. Measurement data are simulated using a finite element approximation of the complete electrode model (1.1), with additive Gaussian noise at a relative level of 10^{-4}. Further details on the forward solver and derivative computations can be found in [16].

Unless stated otherwise, all experiments use identical algorithmic parameters. These are chosen in accordance with the guidelines of [5].

5.2 ALGORITHMS AND THEIR PARAMETERS

We compare the numerical performance of the following instances of Algorithm 2.1. The difference between the instances is how $\widetilde{\nabla E}_{k+1}(\check{x}^{k+1})$ is formed. All methods are based on the reduced adjoint equation of Lemma 3.1 (or, more precisely, a version with a change-of-basis as described in [6]), but differ in how the inner iterates u^{k+1} and adjoint iterates w^{k+1} are calculated.

Gauss–Seidel 7 + 1 This is our proposed single-loop primal-dual method, taking on each iteration of the outer optimisation method, 7 Gauss–Seidel iterations towards the solution the EIT forward PDE to form u^{k+1} from u^k, and 1 Gauss–Seidel iteration towards the solution of the reduced adjoint PDE to form w^{k+1} from w^k. This follows Theorem 4.10 and Example 4.9. The errors for Corollary 2.5 are expanded in Remark 4.11.

Exact In this variant of our method, we solve the PDE and its reduced adjoint exactly, taking (up to numerical precision) $\widetilde{\nabla E}_{k+1}(\check{x}^{k+1}) = \nabla E_k(\check{x}^{k+1})$, $u^{k+1} = S_u(x^k)$, and $w^{k+1} = S_w(x^k)$.

Background This is the method of [5] with the same step length parameters as we used with the above two methods. This method solves the PDE and the adjoint PDE in a background computational thread. While waiting for the solution to become available, the main thread proceeds with the iterations of the optimisation method with new data frames, using the modified data term

$$\check{E}(x) := \frac{1}{2}\|I(\check{x}) + \nabla I(\check{x})^*(x - \check{x}) - b\|^2_{\Sigma^{-1}}$$

that linearises the solution operator I at a past iterate (linearisation point) $\check{x}$. This model necessitates the full basic adjoint, instead of the reduced adjoint. Once the new solution becomes available, it updates the linearisation point, and asks the background solver to solve the PDE at its own current iterate.

Background, original step lengths This is the method used in the numerical experiments of [5], in which we take $\tau = 0.85/\check{L}^2$ and $\sigma = 1$, where $\check{x}$ is the current linearisation point of the background solver, and $\check{L} := \|\nabla^2 \check{E}(x)\| = \|\Sigma^{-1/2}\nabla I(\check{x})\|$.

The step length parameters for the three first algorithms are fixed $\tau = 0.005$. $\sigma = 6.0$. The scheme of [5] cannot be justified anymore, as we no longer linearise the solution operator. For $\check{E}$, $\check{L}$ above gives a valid step length for the convex PDPS. Without the linearisation, we could, in a heuristic fashion, adaptively updated the step lengths based on $L_k := \max_{j\le k} \|\nabla^2 \tilde{E}(x^j)\|$, however, due to the lack of linearisation, this would significantly increase the computational by demanding $\nabla^2 I(x^j)$.

Also, although we use the method and parametrisation of [5] to benchmark our own methods, the performance we obtain from the method is not exactly the same as in [5], as we significantly optimised the code. Due to its interleaved, threaded structure, the method is very sensitive to underlying hardware, other operating system events, and code performance. Although we have also optimised the PDE solver, especially the main computational thread is much faster than it was in [5], which causes much more frames to be processed while waiting for the PDE solution. This results in additional instability that we observe, in particular, in the *Halting Motion* and *Disappearing Inclusions* experiments. Slowing down the method—further limiting how far in history the linearisation point can be—could reduce the instability at additional computational cost.

5.3 RESULTS

The *Baseline* experiment features an inclusion moving at constant speed. Figures 5.1 to 5.4 and Table 5.1 show the relative objective values and iterate errors

The *Constant Motion* experiment features an inclusion moving at constant speed. Figures 5.1 to 5.4 and Table 5.1 show the relative objective values and iterate errors

$$J^k_{\mathrm{rel}} := \frac{J_k(x^k)}{J_k(x^0)} \quad \text{and} \quad e^k_{\mathrm{rel}} := \frac{\|x^k - x^k_{\mathrm{true}}\|}{\|x^k_{\mathrm{true}}\|}, \tag{5.1}$$

where x^k_{true} denotes the ground truth at iteration k, for the tested predictor configurations.

We illustrate the behaviour of the algorithm in terms of function values Figures 5.1 and 5.3 for no prediction and for the best-performing outer primal-dual predictor P_k from [5, 4]: optical flow for the primal variable x, and a scaling dual predictor for y. For the inner and adjoint predictor of the variables u and w we set, for simplicity, $Q_k = \mathrm{Id}$ and $\tilde{Q}_k = \mathrm{Id}$. Our numerical results could possibly be improved slightly by developing more fine-tuned predictors for these variables. Then the error terms of the regret Corollary 2.5 are given by Remark 4.11. The corresponding plots of relative error to the ground truth are in Figures 5.2 and 5.4.

To summarise the information shown in the graphs and tables, our proposed single-loop Gauss–Seidel approach is clearly superior to the alternatives. It has a 5-6 times lower CPU footprint than the approach with an exact PDE solver, and shaves off a third of the CPU footprint of the old background solver of [5]. While the latter might be an acceptable trade-off if the old approach were otherwise better, it, however, becomes very unstable in the Halting/Disappearing test scenarios well: the linearised model is not updated at a sufficient frequency compared to the nature of data evoluation. More frequent updates could be imposed at a higher computational cost, comparable to our proposed single-loop solvers that have a simpler and more reliable computational architecture that does not depend on synchronisation between computational threads. That being said, based on additional experiments, also the Gauss–Seidel approach becomes unstable if the number of adjoint steps is reduced from the 7 steps used here to 6 or less. The background solver also heavily depends on the specific adaptive step length estimation scheme of [5]; it does not perform well with the simple fixed step sizes that we use with the other approaches.

APPENDIX A TECHNICAL TRACKING ESTIMATES

Following the structure of [6, Appendix A], we prove here a simple scalar online tracking results, which are used to establish the results of Sections 3 and 4. The following assumption is a scalar variant of Assumption 3.3, and differs from [5, Assumption A.1] through the inclusion of the errors $\varepsilon_k, \tilde{\varepsilon}_k$.

Assumption A.1. For a given $k \geq 0$ and scalars $d^u_0, \ldots, d^u_{k+1}, d^w_0, \ldots, d^w_{k+1}, \varrho_1, \ldots, \varrho_k \geq 0$, and $\widetilde{d}_0, \ldots, \widetilde{d}_{k+1} \in$

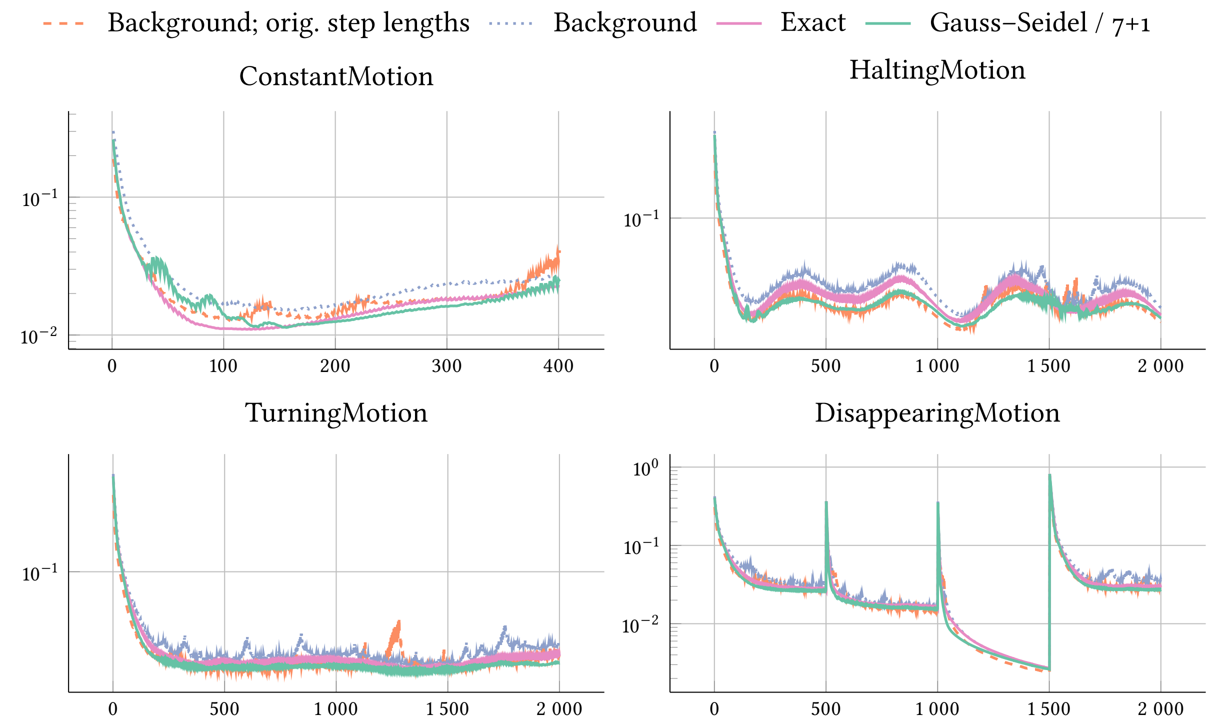

Figure 5.1: Relative function value for the *Greedy* predictor. The horizontal axis indicates the iteration, and the vertical axis the relative function value (5.1).

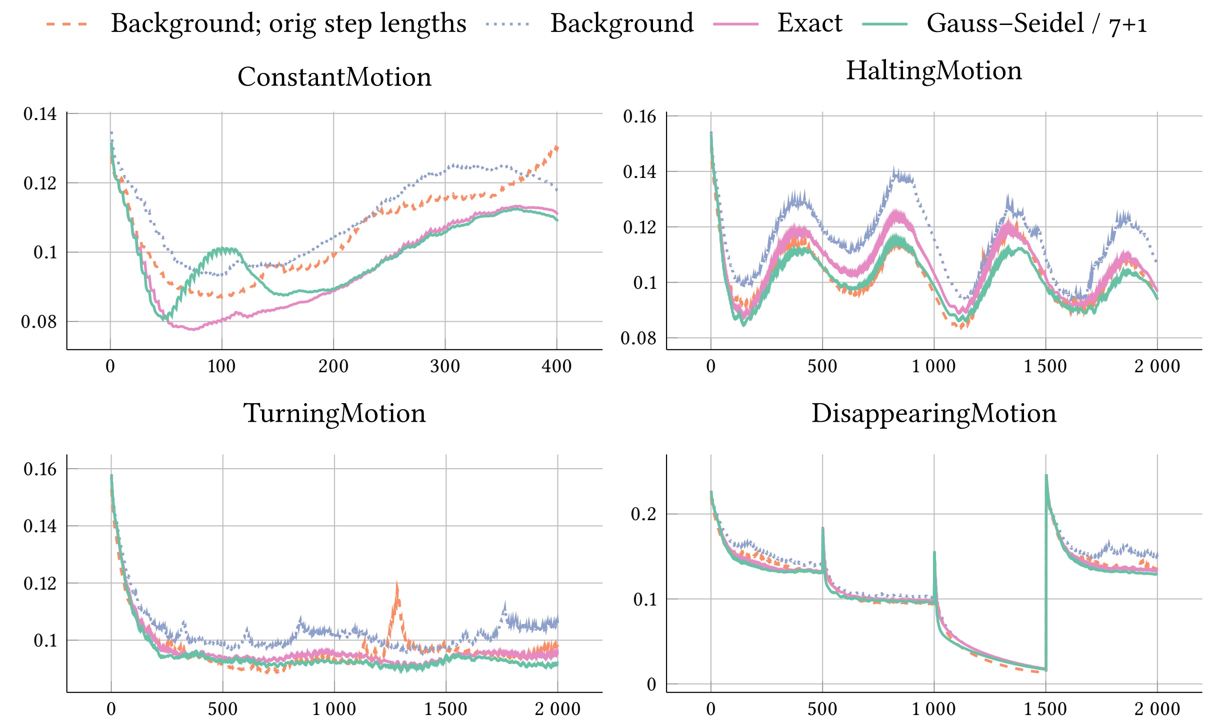

Figure 5.2: Ground truth relative error for the *Greedy* predictor. The horizontal axis indicates the iteration, and the vertical axis the error the relative error to the ground truth (5.1).

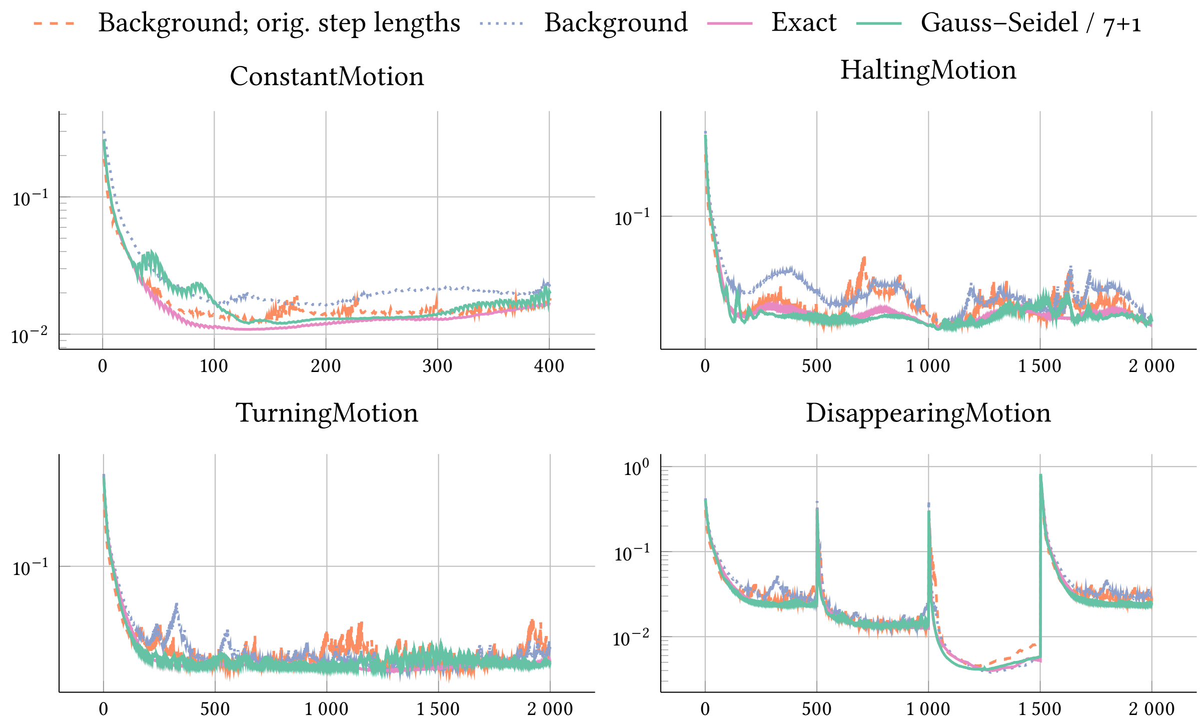

Figure 5.3: Relative function value for the *Affine* predictor. The horizontal axis indicates the iteration, and the vertical axis the relative function value (5.1).

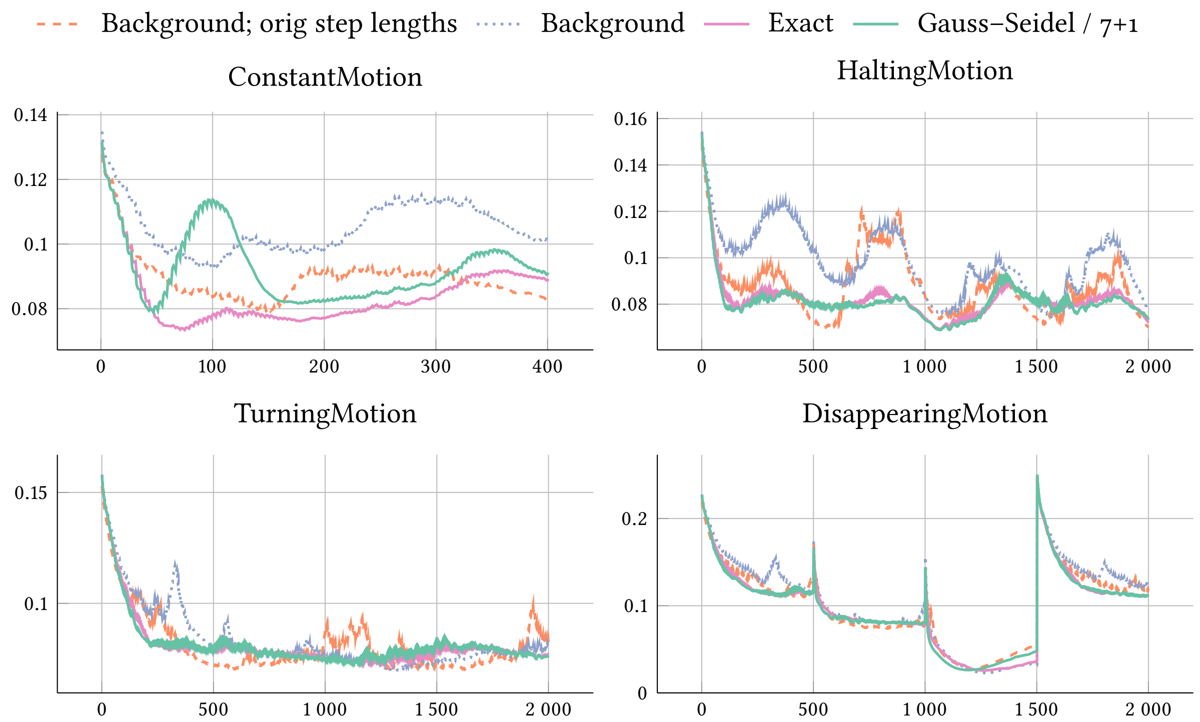

Figure 5.4: Ground truth relative error for the *Affine* predictor. The horizontal axis indicates the iteration, and the vertical axis the error the relative error to the ground truth (5.1).

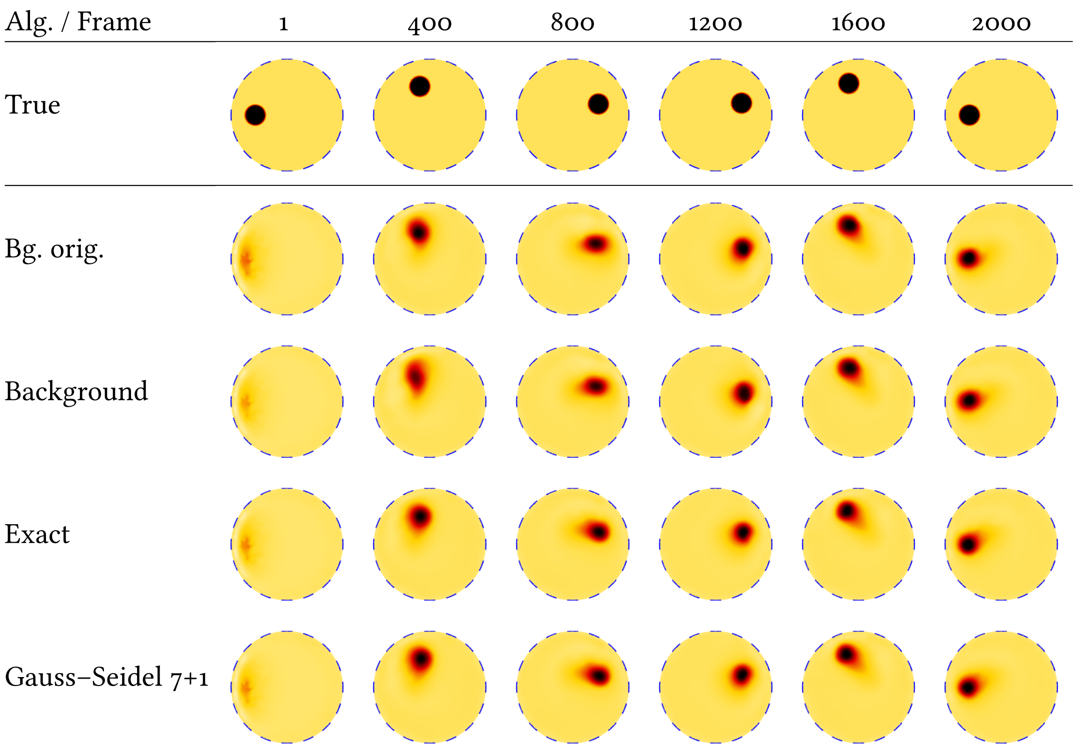

Figure 5.5: Reconstructions for the HaltingMotion experiment with the *Affine* predictor.

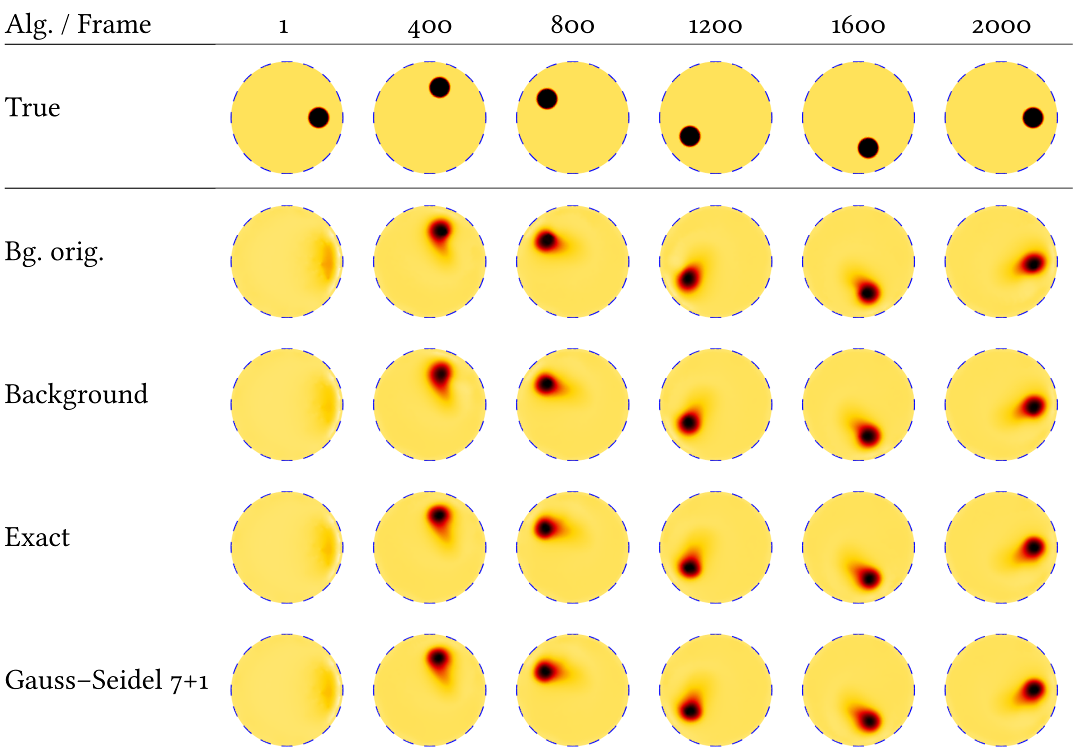

Figure 5.6: Reconstructions for the TurningMotion experiment with the *Affine* predictor.

Table 5.1: Summary statistics of all experiments. The experiments (Exp.), predictors (Pred.:Af = Affine; Gr = Greedy), and the algorithm (Alg.: GS = Gauss–Seidel 7 + 1d, Bg. = Background, new step lengths, Orig. = Background, original step lengths). Statistics marked with an asterisk* begin after the first 50 frames (Constant Motion) or 200 frames (other experiments). Mean, standard deviation (Std), and the 95% confidence interval (CI) are for the relative function value error J^k_{rel}, and the "GT" variants for the relative error to the ground truth, e^k_{rel}, both defined in(5.1). The average real and CPU times spent processing each data frame, are in seconds.

			Function value error*			Ground truth error*			Frame time	
Exp.	Pred.	Alg.	Mean	Std	CI	Mean	Std	CI	Real	CPU
Const.	Gr	GS	4.33	0.95	[4.23, 4.43]	0.099	0.009	[0.098, 0.1]	0.0055	0.0128
Const.	Gr	Exact	4.23	0.97	[4.13, 4.33]	0.095	0.013	[0.094, 0.097]	0.0153	0.0612
Const.	Gr	Bg.	5.54	1.03	[5.43, 5.65]	0.110	0.012	[0.108, 0.111]	0.0032	0.0154
Const.	Gr	Orig.	4.83	1.30	[4.7, 4.97]	0.105	0.013	[0.104, 0.107]	0.0035	0.0169
Const.	Af	GS	4.26	1.00	[4.15, 4.36]	0.091	0.009	[0.091, 0.092]	0.0055	0.0129
Const.	Af	Exact	3.53	0.44	[3.49, 3.58]	0.082	0.006	[0.081, 0.082]	0.0156	0.0621
Const.	Af	Bg.	5.37	0.58	[5.31, 5.43]	0.104	0.007	[0.103, 0.104]	0.0032	0.0155
Const.	Af	Orig.	4.05	0.37	[4.02, 4.09]	0.087	0.004	[0.087, 0.088]	0.0036	0.0172
Turn.	Gr	GS	3.76	0.13	[3.75, 3.76]	0.093	0.001	[0.093, 0.093]	0.0053	0.0110
Turn.	Gr	Exact	3.92	0.21	[3.91, 3.93]	0.094	0.001	[0.094, 0.094]	0.0154	0.0616
Turn.	Gr	Bg.	4.47	0.45	[4.45, 4.49]	0.101	0.003	[0.101, 0.101]	0.0038	0.0172
Turn.	Gr	Orig.	3.89	0.45	[3.87, 3.91]	0.095	0.004	[0.095, 0.095]	0.0038	0.0180
Turn.	Af	GS	3.45	0.21	[3.44, 3.46]	0.079	0.003	[0.079, 0.079]	0.0054	0.0111
Turn.	Af	Exact	3.18	0.11	[3.17, 3.18]	0.078	0.002	[0.078, 0.078]	0.0154	0.0617
Turn.	Af	Bg.	3.68	0.72	[3.64, 3.71]	0.080	0.008	[0.079, 0.08]	0.0035	0.0169
Turn.	Af	Orig.	3.68	0.61	[3.65, 3.71]	0.078	0.007	[0.077, 0.078]	0.0038	0.0181
Halt.	Gr	GS	4.16	0.47	[4.14, 4.19]	0.101	0.008	[0.1, 0.101]	0.0053	0.0110
Halt.	Gr	Exact	4.68	0.71	[4.65, 4.71]	0.106	0.009	[0.105, 0.106]	0.0154	0.0615
Halt.	Gr	Bg.	5.47	0.97	[5.43, 5.52]	0.115	0.012	[0.115, 0.116]	0.0035	0.0168
Halt.	Gr	Orig.	4.25	0.60	[4.23, 4.28]	0.102	0.009	[0.102, 0.103]	0.0038	0.0180
Halt.	Af	GS	3.41	0.30	[3.4, 3.43]	0.080	0.005	[0.08, 0.08]	0.0054	0.0111
Halt.	Af	Exact	3.39	0.23	[3.38, 3.4]	0.081	0.004	[0.08, 0.081]	0.0155	0.0620
Halt.	Af	Bg.	4.59	0.89	[4.54, 4.63]	0.096	0.013	[0.096, 0.097]	0.0037	0.0175
Halt.	Af	Orig.	3.99	0.79	[3.95, 4.02]	0.086	0.012	[0.086, 0.087]	0.0040	0.0188
Dis.	Gr	GS	7.59	14.54	[6.92, 8.26]	0.100	0.046	[0.098, 0.102]	0.0053	0.0110
Dis.	Gr	Exact	8.16	14.99	[7.47, 8.85]	0.104	0.046	[0.101, 0.106]	0.0154	0.0616
Dis.	Gr	Bg.	9.28	15.49	[8.56, 10]	0.110	0.051	[0.108, 0.113]	0.0036	0.0172
Dis.	Gr	Orig.	7.72	13.20	[7.11, 8.33]	0.102	0.049	[0.1, 0.104]	0.0038	0.0182
Dis.	Af	GS	7.09	14.37	[6.43, 7.76]	0.091	0.042	[0.089, 0.093]	0.0054	0.0111
Dis.	Af	Exact	7.21	14.87	[6.52, 7.9]	0.090	0.042	[0.088, 0.092]	0.0155	0.0621
Dis.	Af	Bg.	8.62	15.70	[7.9, 9.35]	0.097	0.048	[0.095, 0.1]	0.0036	0.0173
Dis.	Af	Orig.	7.73	13.37	[7.11, 8.35]	0.094	0.042	[0.092, 0.096]	0.0040	0.0188

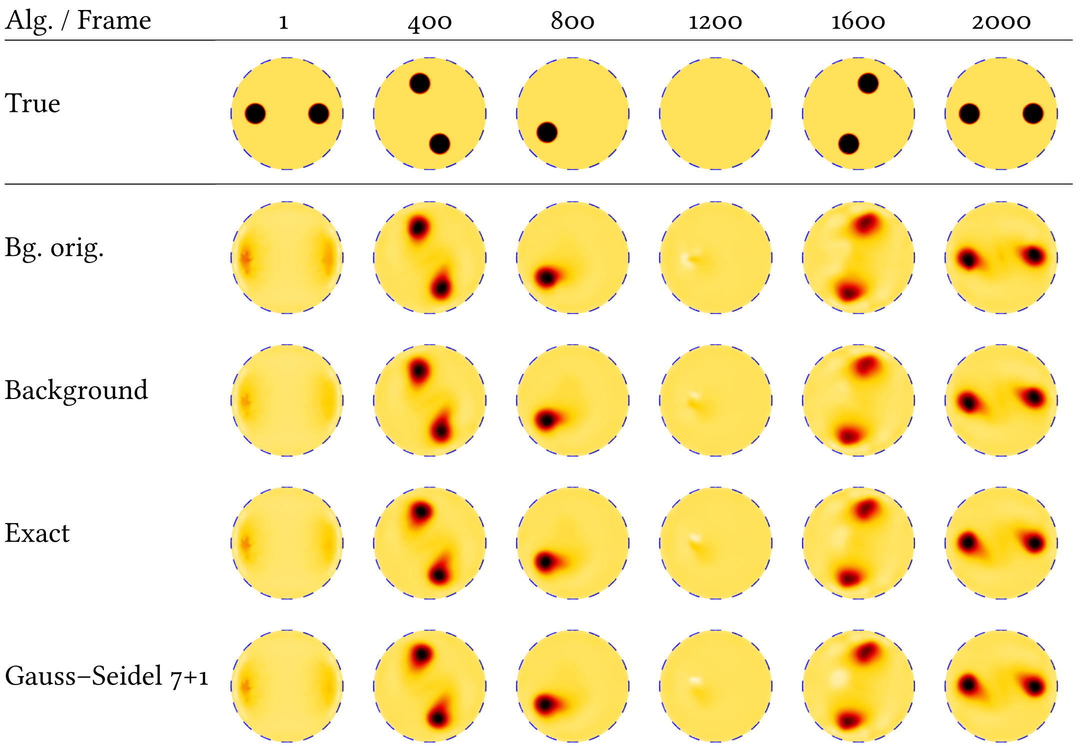

Figure 5.7: Reconstructions for the DisappearingMotion experiment with the *Affine* predictor.

$\mathbb{R}$, there exist $\pi_u, \pi_w, \mu_u \alpha_w, \alpha_u > 0$, and some errors $\varepsilon_k, \tilde{\varepsilon}_k \geq 0$, such that

(i) $\kappa_u d^u_{j+1} \leq d^u_j + \pi_u(\varrho_j + \varepsilon_j)$ for all $j = 1, \ldots, k$,

(ii) $\kappa_w d^w_{j+1} \leq d^w_j + \mu_u d^u_{j+1} + \pi_w(\varrho_j + \varepsilon_j)$ for all $j = 1, \ldots, k$, and

(iii) $\widetilde{d}_{j+1} \leq \alpha_u d^u_{j+1} + \alpha_w d^w_{j+1} + \tilde{\varepsilon}_j$ for all $j = 0, \ldots, k$.

Based on this recursive assumption, we aim to establish simple non-recursive bounds on $\widetilde{d}_{k+1}$. Note that we could at the level of the assumption combine ϱ_j and ε_j into the single variable

$$\check{\varrho}_j := \varrho_j + \varepsilon_j,$$

however, due to their different purposes, do not do this. This combination is, however, allows us use [6] to deduce the next lemma that unrolls the recursion of (i) and (ii).

Lemma A.2 ([6, Lemma A.2]). *Let Assumption A.1(i) and (ii) hold for a $k \geq 0$. Then, letting $\iota_k := \sum_{m=1}^k \kappa_u^{-m} \kappa_w^{-(k+1-m)}$ (understanding that $\iota_0 = 0$), we have*

$$\begin{aligned} R^{k+1} &:= \alpha_u d^u_{k+1} + \alpha_w d^w_{k+1} \\ &\leq (\alpha_u \kappa_u^{-k} + \alpha_w \iota_k \mu_u) d^u_1 + \alpha_w \kappa_w^{-k} d^w_1 \\ &\quad + \sum_{j=0}^{k-1} \big(\alpha_u \kappa_u^{-(k-j)} \pi_u + \alpha_w [\iota_{k-j} \mu_u \pi_u + \kappa_w^{-(k-j)} \pi_w]\big) \check{\varrho}_{j+1}. \end{aligned} \tag{A.1}$$

The next two lemmas form our core estimates. They extend the corresponding lemmas of [6] to k-dependent functions, with the corresponding additional errors from Assumption A.1. To obtain the estimates, recalling that $\kappa_u, \kappa_w > 1$, we observe that

$$p^k \iota_k \leq p^{-1} k (\kappa/p)^{-(k+1)} \quad \text{for} \quad \kappa := \min(\kappa_u, \kappa_w) > 1 \text{ and any } p \in (0, \kappa). \tag{A.2}$$

Thus, by sum formulae for arithmetic-geometric progressions [8, formula 0.113],

$$\sum_{k=0}^{n-1} p^k \iota_k \le \sum_{k=0}^{\infty} p^k \iota_k \le p^{-1}(\kappa/p-1)^{-2} = p(\kappa-p)^{-2} \quad \text{for all } n \in \mathbb{N}. \tag{A.3}$$

Lemma A.3. *Suppose Assumption A.1 holds for a* $k \ge 0$*. Then for any* $p \in (0,\kappa)$*, we have*

$$\tilde{d}_{k+1}^2 \le (\alpha_u d_{k+1}^u + \alpha_w d_{k+1}^w + \tilde{\varepsilon}_k)^2 \le \check{e}_{p,k}. \tag{A.4}$$

where, for $\psi_j := \alpha_u \kappa_u^{-j}\pi_u + \alpha_w[\iota_j \mu_u \pi_u + \kappa_w^{-j}\pi_w]$*, and* $\overline{\kappa} := \max\{\kappa_u, \kappa_w\}$*, we set*

$$\varsigma_p := \frac{\overline{\kappa}}{p}\sum_{j=0}^{\infty} p^j \psi_j \le \frac{(\alpha_u\pi_u + \alpha_w\pi_w)\kappa\overline{\kappa}}{p(\kappa-p)} + \frac{\alpha_w\mu_u\pi_u\overline{\kappa}}{p^2(\kappa-p)^2} \quad \textit{and} \tag{A.5}$$

$$\check{e}_{p,k} := \frac{5}{4}\left(\frac{\varsigma_p(\alpha_u\kappa_u^{-k} + \alpha_w\iota_k\mu_u)}{\pi_u p^k}(d_1^u)^2 + \frac{\varsigma_p\alpha_w\kappa_w^{-k}}{\pi_w p^k}(d_1^w)^2 + \sum_{j=0}^{k-1}\frac{\varsigma_p\psi_{k-j}^2}{p^{k-j}}\check{\varrho}_{j+1}^2 + 4\tilde{\varepsilon}_k^2\right). \tag{A.6}$$

Proof. Invoking the inner and adjoint tracking Assumption A.1 (i) and (ii) and Lemma A.2, we obtain

$$R^{k+1} := \alpha_u d_{k+1}^u + \alpha_w d_{k+1}^w \le (\alpha_u\kappa_u^{-k} + \alpha_w\iota_k\mu_u)d_1^u + \alpha_w\kappa_w^{-k}d_1^w + \sum_{j=0}^{k-1}\psi_{k-j}\check{\varrho}_{j+1}.$$

Thus, using Young's inequality several times, we deduce for any $\theta_k^u, \theta_k^w, \theta_{k,j}, s > 0$ that

$$\begin{aligned} 4sR^{k+1} \le{}& \frac{(\alpha_u\kappa_u^{-k} + \alpha_w\iota_k\mu_u)^2}{4\theta_k^u}(d_1^u)^2 + \frac{(\alpha_w\kappa_w^{-k})^2}{4\theta_k^w}(d_1^w)^2 \\ &+ \sum_{j=0}^{k-1}\frac{\psi_{k-j}^2}{\theta_{k,j}}\check{\varrho}_{j+1}^2 + 4\left(\theta_k^u + \theta_k^w + \sum_{j=0}^{k-1}\theta_{k,j}\right)s^2. \end{aligned} \tag{A.7}$$

Take $\theta_k^u = p^k\varsigma_p^{-1}\pi_u(\alpha_u\kappa_u^{-k} + \alpha_w\iota_k\mu_u)$, $\theta_k^w = p^k\varsigma_p^{-1}\pi_w\alpha_w\kappa_w^{-k}$, and $\theta_{k,j} = \varsigma_p^{-1}p^{k-j}\psi_{k-j}$. Obseve that $\iota_k \le \kappa_w\iota_{k+1}$ [6, proof of Lemma A.2]. Hence, we get $p^k\iota_k \le (\kappa_w/p)p^{k+1}\iota_{k+1}$, and thus $p^k\psi_k \le (\bar{\kappa}/p)p^{k+1}\psi_{k+1}$, where $\bar{\kappa}/p > 1$. Then we have

$$\theta_k^u + \theta_k^w + \sum_{j=0}^{k-1}\theta_{k,j} = \frac{1}{\varsigma_p}\left(p^k\psi_k + \sum_{j=1}^{k}p^j\psi_j\right) \le \frac{\bar{\kappa}}{p\varsigma_p}\sum_{j=0}^{k+1}p^j\psi_j \le 1.$$

By combining (A.7) with the previous bound and rearranging terms establishes

$$4sR^{k+1} - 4s^2 \le \frac{\varsigma_p(\alpha_u\kappa_u^{-k} + \alpha_w\iota_k\mu_u)}{\pi_u p^k}(d_1^u)^2 + \frac{\varsigma_p\alpha_w\kappa_w^{-k}}{\pi_w p^k}(d_1^w)^2 + \sum_{j=0}^{k-1}\frac{\varsigma_p\psi_{k-j}^2}{p^{k-j}}\check{\varrho}_{j+1}^2.$$

We have $4s\tilde{\varepsilon}_k \le s^2 + 4\tilde{\varepsilon}_k^2$, so it follows $4s(R^{k+1} + \tilde{\varepsilon}_k) - 5s^2 \le \frac{4}{5}\check{e}_{p,k}$. The left hand side is maximised by $s = \frac{2}{5}(R^{k+1} + \tilde{\varepsilon}_k)$. This gives $(R^{k+1} + \tilde{\varepsilon}_k)^2 \le \check{e}_{p,k}$ By Assumption A.1 (iii), $\widetilde{d}_{k+1} \le R^{k+1} + \tilde{\varepsilon}_k$, so (A.4) follows.

Finally, the bound in (A.5) on ς_p follows from (A.3) and $\sum_{j=0}^{\infty}(p/k)^j = 1/(1-p/k) = \kappa/(\kappa-p)$. ◻

Lemma A.4. *Suppose Assumption A.1 holds for a* $k \ge 0$*. Then, for any* $p \in [1,\kappa)$*, we have*

$$\widetilde{d}_{k+1}^2 \le \frac{5}{4}\varsigma_p^2\check{\varrho}_{k+1}^2 + e_{p,k}, \tag{A.8}$$

where, for $\check{e}_{p,k}$ defined in (A.6),

$$e_{p,k} := \check{e}_{p,k} - \frac{5}{4}\varsigma_p^2 \check{\varrho}_{k+1}^2 \tag{A.9}$$

satisfies

$$\sum_{n=0}^{k} p^n e_{p,n} \le \Psi_p := \frac{5}{4}\left(\frac{(d_1^u)^2}{\pi_u}\left(\frac{\varsigma_p \alpha_u \kappa}{\kappa - 1} + \frac{\varsigma_p \alpha_w \mu_u}{(\kappa-1)^2}\right) + \frac{(d_1^w)^2}{\pi_w}\left(\frac{\varsigma_p \alpha_w \kappa}{\kappa - 1}\right) + \sum_{n=0}^{k} p^n \tilde{\varepsilon}_n^2\right).$$

Proof. The proof is analogous to that of [6, Lemma A.4]. □

Finally we have the following result similar to [6, Lemma A.5].

Lemma A.5. *Suppose Assumption A.1 holds for a $k \in \mathbb{N}$. Then, for $\check{e}_{1,n}$ given in (A.6), we have $\sum_{n=0}^{k-1} \check{e}_{1,n} \le \Psi_1 + \frac{5}{4}\varsigma_1^2 \sum_{n=0}^{k-1} \check{\varrho}_{n+1}^2$.*

Proof. We have $\check{e}_{1,n} = e_{1,n} + \frac{5}{4}\varsigma_1^2 \check{\varrho}_{n+1}^2$, where (A.5) bounds $\sum_{n=0}^{k-1} e_{1,n} \le \Psi_1$. □

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] C. Clason and T. Valkonen, *Introduction to Nonsmooth Analysis and Optimization*, MOS-SIAM Series on Optimization, SIAM, 2026, doi:10.1137/1.9781611978995.

[3] N. D. Dizon, J. Jauhiainen, and T. Valkonen, Prediction Techniques for Dynamic Imaging with Online Primal–Dual Methods, *Journal of Mathematical Imaging and Vision* 66 (2024), 1109–1134.

[4] 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.

[5] N. Dizon, J. Jauhiainen, and T. Valkonen, Online optimisation for dynamic electrical impedance tomography, 2025, doi:10.1088/1361-6420/adcb66, arXiv:2412.12944.

[6] N. Dizon and T. Valkonen, Differential estimates for fast first-order multilevel nonconvex optimisation, 2024, arXiv:2412.01481. submitted.

[7] G. H. Golub and C. F. Van Loan, Matrix computations. Johns Hopkins studies in the mathematical sciences, 1996.

[8] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, Academic press, 2014.

[9] E. Hall and R. Willett, Dynamical models and tracking regret in online convex programming, in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester (eds.), volume 28 of Proceedings of Machine Learning Research, PMLR, Atlanta, Georgia, USA, 2013, 579–587, http://proceedings.mlr.press/v28/hall13.html.

[10] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, number 23 in Mathematical Modelling: Theory and Applications, Springer Netherlands, 2009, doi:10.1007/978-1-4020-8839-1.

[11] D. Holder and K. Aristovich, Electrical impedance tomography, in *The Physics of Medical Imaging*, CRC press, 2026, 381–391.

[12] D. J. Holland, D. M. Malioutov, A. Blake, A. J. Sederman, and L. F. Gladden, Reducing data acquisition times in phase-encoded velocity imaging u sing compressed sensing, *Journal of Magnetic Resonance* 203 (2010), 236–46.

[13] 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.

[14] J. Jauhiainen, N. Dizon, T. Valkonen, and Y. Nabou, Online optimisation codes for dynamic electrical impedance tomography, 2026, doi:10.5281/zenodo.19154746. Software.

[15] J. Jauhiainen, P. Kuusela, A. Seppanen, and T. Valkonen, Relaxed Gauss–Newton methods with applications to electrical impedance tomography, *SIAM Journal on Imaging Sciences* 13 (2020), 1415–1445.

[16] 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.

[17] J. Jauhiainen, Y. Nabou, and T. Valkonen, Dynamic inverse problems: Online regularisation theory, 2026, arXiv:2605.26022. submitted.

[18] 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.

[19] B. Jensen and T. Valkonen, A nonsmooth primal-dual method with interwoven PDE constraint solver, *Computational Optimization and Applications* 89 (2024), 115–149, doi:10.1007/s10589-024-00587-3, arXiv:2211.04807.

[20] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography, *Inverse Problems* 16 (2000), 1487.

[21] 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.

[22] T. Schuster, B. Hahn, and M. Burger, Dynamic inverse problems: modelling—regularization—numerics, *Inverse Problems* 34 (2018), 040301, doi:10.1088/1361-6420/aab0f5. Preface to special issue.

[23] E. Suonperä and T. Valkonen, General single-loop methods for bilevel parameter learning, 2024, arXiv:2408.08123. submitted.

[24] 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.

[25] 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.