ONLINE OPTIMISATION FOR DYNAMIC ELECTRICAL IMPEDANCE TOMOGRAPHY

Neil Dizon* Jyrki Jauhiainen† Tuomo Valkonen‡

Abstract Online optimisation studies the convergence of optimisation methods as the data embedded in the problem changes. Based on this idea, we propose a primal dual online method for nonlinear time-discrete inverse problems. We analyse the method through regret theory and demonstrate its performance in real-time monitoring of moving bodies in a fluid with Electrical Impedance Tomography (EIT). To do so, we also prove the second-order differentiability of the Complete Electrode Model (CEM) solution operator on L^∞.

1 INTRODUCTION

Electrical impedance tomography (EIT) is an imaging technique for inferring the electrical conductivity distribution within a body through boundary currents and potentials. While measurements in EIT can be performed in real time, reconstructing images from the data is computationally intensive. This challenge is critical in applications such as real-time monitoring of industrial processes, where immediate feedback is essential – for instance, in detecting blockages or leaks in pipelines.

Traditionally, inverse problems, including those arising in EIT, have been studied in a static context, where robust theoretical foundations and solution methods are available. However, the need for real-time reconstructions in dynamic settings has grown significantly [31, 19, 21, 6, 32]. Addressing this demand requires novel approaches capable of efficiently processing large data sets and capturing time-dependent changes in the imaged domain.

To this end, we introduce online optimization methods tailored to time-discrete nonlinear inverse problems, formulated as the conceptual problem

$$\min_{(x^0,x^1,x^2,\dots)\in\mathcal{X}} \sum_{k=0}^{\infty} J_k(x^k) \tag{1.1}$$

for a set $\mathcal{X} \subset \prod_{k=0}^{\infty} X_k$ that encodes the temporal couplings between the frame-wise variables x^k over time index k. With Tikhonov-type regularization, the frame-wise objective can be expressed as

$$J_k(x) := E(A_k(x) - b_k) + R(x), \tag{1.2}$$

where R is a regularization term, A_k a nonlinear forward operator, b_k the measurement data, and the data fidelity E models noise.

This research has been supported by the Research Council of Finland grants 338614, 314701, 353088, and 358944.

*School of Mathematics and Statistics, University of New South Wales, Sydney, Australia. n.dizon@unsw.edu.au, ORCID: 0000-0001-8664-2255

†Department of Technical Physics, University of Eastern Finland, Finland, jyrki.jauhiainen@uef.fi, ORCID: 0000-0001-6711-6997

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

In this work, we focus on EIT process monitoring using isotropic total variation regularization, $R(x) = \alpha\|Dx\|_{2,1}$. Here A_k represents the solution operator for the complete electrode model (CEM) [10]. For given electrode potentials, it maps the electrical conductivity x on a domain $\Omega \subset \mathbb{R}^n$ to boundary current measurements. The temporal couplings encoded by $\mathcal{X}$ arise from a partial differential equation (PDE) that models matter movement within the body. For simplicity in our numerical demonstrations, we employ the incompressible transport equation, though more sophisticated models, such as the Navier–Stokes equations, could be applied.

The problem (1.1) is formal: we cannot in practise solve an optimisation problem for an unbounded time segment. At each instant N, we can access at most the initial segment $\{(A_k, b_k)\}_{k=0}^N$ of forward operators and data. It is conceivable to solve the corresponding finite horizon problems $\min_{x^{0:N} \in \mathcal{X}_{0:N}} \sum_{k=0}^N J_k(x^k)$ where we use the slicing notation

$$x^{0:N} = (x^0, \ldots, x^N) \quad \text{and} \quad \mathcal{X}_{0:N} := \{x^{0:N} \mid (x^0, x^1, \ldots) \in \mathcal{X}\}. \tag{1.3}$$

However, even this is numerically unwieldy for large N, and hardly real-time with standard optimisation methods. In practise, in a long monitoring process, both CPU and memory requirements would also force us to work with a short time window of data.

Online optimisation [48] attempts to solve (1.1) in real time. The idea in most methods is to take a *single* step of a standard optimisation method for $\min J_k$ at each index k. For introductions we refer to [18, 5, 35, 37]. Basic online optimisation methods take $\mathcal{X}$ in (1.1) to constrain $x^0 = x^1 = x^2 = \ldots$, i.e., do not consider problems that evolve over time, only data that arrives gradually. *Dynamic online optimisation* methods [17, 42, 47, 8, 45, 34, 39, 3, 46, 7, 9] typically intersperse optimisation steps with *prediction* steps for some prediction operators W_k that attempt to model more general temporal constraint sets $\mathcal{X}$ subject to available, possibly noisy and corrupted, information. In particular, a dynamic online forward-backward (a.k.a. proximal gradient) method for (1.1)&(1.2) iterates

$$x^k := \operatorname{prox}_{\tau R}(\check{x}^k - \tau \nabla A_k(\check{x}^k)^*(A_k(\check{x}^k) - b_k)) \quad \text{with the predictions} \quad \check{x}^{k+1} := W_k(x^k). \tag{1.4}$$

Most earlier works on dynamic online optimisation concentrate on such forward-backward methods. They have limited applicability to inverse problems with total variation regularisation $R = \|\nabla \cdot\|_{2,1}$, as the proximal operator $\operatorname{prox}_{\tau R}(x) := \arg\min_z \frac{1}{2}\|z - x\|^2 + \tau R(z)$ is expensive to calculate: it corresponds to total variation denoising. Therefore, we developed in [42, 14] primal-dual dynamic online optimisation methods for (1.1) with a linear A_k in (1.2). The initial theory in [42] imposed severe restrictions on the dual component of the predictor. These were relaxed in [14], where several improved dual predictors were developed, for the primal predictor based on optical flow.

The main results of [42] could also only be interpreted through regularisation theory, as the *regret* that we were able to prove was non-symmetric. Indeed, *convergence* results are rarely available for online optimisation methods. Instead, one attempts to bound the *regret* of past updates with respect to all information available up to an instant N. For (1.4), one can bound the *dynamic regret* [17]

$$\operatorname{dynreg}(x^{0:N}) = \sup_{\bar{x}^{0:N} \in \mathcal{X}_{0:N}} \sum_{k=0}^N \left(J_k(x^k) - J_k(\bar{x}^k) \right).$$

This may be negative, if the *comparison set* $\mathcal{X}_{0:N}$ does not include all the possible "paths" $x^{k:N}$ that the iterates generated by the algorithm may take. For the primal-dual method of [42], the result was even weaker: we could only bound

$$\sup_{\bar{x}^{0:N} \in \mathcal{X}_{0:N}} \left(\check{J}_{0:N}(x^{0:N}) - \sum_{k=0}^N J_k(\bar{x}^k) \right)$$

for a temporal infimal convolution $\check{J}_{0:N}$ between the comparison set and the framewise objectives. In [14], at the cost of having to reduce the size of the comparison set, we managed to symmetrise these results to a bound on

$$\mathring{J}_{0:N}(x^{0:N}) - \sup_{\bar{x}^{0:N} \in \mathcal{X}_{0:N}} \mathring{J}_{0:N}(\bar{x}^{0:N})$$

for a different temporal (sub-)infimal convolution $\mathring{J}_{0:N}$.

The theory in [14] was still for convex functions. In Section 2, we will

(a) *extend the primal-dual method of [14] to non-convex objectives, including* (1.1)&(1.2) *with non-linear forward operators A_k, and*

(b) *allow for inexact gradients, in particular, inexact A_k and its Jacobian.*

The non-convexity adds significant new technical challenges to the proof. The inexact computations help to reduce the high expense of solving PDEs and can, fortunately, be incorporated with much less effort. While we do not yet treat interweaved PDE solvers as [27] did for static problems, that is our ultimate goal. Instead, we do delayed, intermittent solves of the PDEs "in the background" to obtain real-time performance. To prove that the EIT problem satisfies the conditions of the online method, in Section 3,

(c) *we will prove the second-order differentiability of the CEM solution operator for L^∞ conductivities.*

In the final Section 4, we evaluate the proposed method numerically on (nearly) real-time EIT reconstruction. Our approach to EIT process monitoring differs from the recent work in [2], where the authors achieved real-time performance using extensive multithreading and code optimisation with the D-bar method [33, 36]. In contrast, our online method attains real-time performance thanks to the inherently light computational cost of iterations of our algorithm, albeit dependent on meticulous predictor design. Finally, our optimisation-based approach is not the only possibility treat dynamic inverse problems in an online fashion. Especially if uncertainty quantification is required, full Bayesian modelling can be used with numerical techniques such as non-linear Kalman filters and Markov-Chain Monte Carlo [30, 16, 1]. Except in the simplest fully Gaussian cases (which would exclude a total variation prior), this can, however, come with much higher computational resource demands and complexity.

Additional notation In the slicing notation (1.3), we allow $N = \infty$ with $x^{k:\infty} = (x^k, x^{k+1}, \dots)$ and also set $\mathcal{X}_k := \mathcal{X}_{k:k}$. We write $\mathbb{L}(X; Y)$ for the space of bounded linear operators between normed spaces X and Y, and $\mathrm{Id} \in \mathbb{L}(X; X)$ for the identity operator. The block-diagonal operator consisting of M and Γ reads $\mathrm{diag}(M,\ \Gamma)$. When X is Hilbert, we abbreviate $\langle x, y\rangle_M = \langle Mx, y\rangle$ for $M \in \mathbb{L}(X; X)$. The notation $[\![x]\!]_M^2$ mimics norm notation when M may not be positive semi-definite. If it is, written $M \geq 0$, we set $\|x\|_M := \sqrt{\langle x, x\rangle_M}$. For a Borel $g : \mathbb{R}^n \supset \Omega \to \mathbb{R}^n$, we set $\|g\|_{2,1} := \int_\Omega \|g(\xi)\|_2 \, d\xi$. Applied to the gradient of a differentiable function, this produces the isotropic total variation.

For $A \subset X$ and $x \in X$, we write $\langle A, x\rangle = \{\langle z, x\rangle \mid z \in A\}$. For any $B \subset \mathbb{R}$ (especially $B = \langle A, x\rangle$), $B \geq 0$ means $t \geq 0$ for all $t \in B$. For a convex function $F : X \to \overline{\mathbb{R}}$, ∂F denotes the subdifferential, and F^* the Fenchel conjugate. The $\{0, \infty\}$-valued indicator function of A is written δ_A. We refer to [12] for more details on convex analysis.

2 ONLINE PRIMAL-DUAL PROXIMAL SPLITTING FOR NONCONVEX PROBLEMS

We consider the conceptual problem

$$\min_{(x^0, x^1, x^2, \dots) \in \mathcal{X}} \sum_{k=0}^{\infty} F_k(x^k) + E_k(x^k) + G_k(K_k x^k), \tag{2.1}$$

Algorithm 1 Nonconvex predictive online primal-dual proximal splitting (POPD-N).

Require: For all $k \in \mathbb{N}$, on Hilbert spaces X_k and Y_k, convex, proper, lower semi-continuous $E_{k+1} : X_{k+1} \to \mathbb{R}$, $F_{k+1} : X_{k+1} \to \overline{\mathbb{R}}$ and $G^*_{k+1} : Y_{k+1} \to \overline{\mathbb{R}}$, primal-dual predictor $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$, and $K_{k+1} \in \mathbb{L}(X_{k+1}; Y_{k+1})$. Estimates $\widetilde{\nabla E}_{k+1}(\check{x}^{k+1})$ of the gradients $\nabla E_{k+1}(\check{x}^{k+1})$. Step length parameters $\tau_{k+1}, \sigma_{k+1} > 0$.

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

where G_k and F_k are convex and possibly nonsmooth, but E_k is smooth but possibly nonconvex. The operator K_k is linear and bounded. The *primal comparison set* $\mathcal{X}$ encodes temporal couplings between the framewise variables x^k. This clearly includes (1.1)&(1.2), with non-linear forward-operators A_k accommodated by setting $E_k(x) = E(A_k(x) - b_k)$.

We present in Algorithm 1 our proposed primal-dual online method for (2.1). It involves the Fenchel conjugate G^*_k of G_k and a dual variable y^k introduced through writing $G_k(Kx) = \sup_y \langle K_k x, y\rangle - G^*_k(y)$, as well as the predictor P_k and the predictions

$$(\check{x}^{k+1}, \check{y}^{k+1}) := P_k(x^k, y^k).$$

We have also replaced $\nabla E_k(\check{x}^k)$ by an estimate $\widetilde{\nabla E}_k(\check{x}^k)$. This allows for the inexact computation of A_k and ∇A_k, as discussed in the Introduction.

To derive the algorithm, note that a single term in (2.1) is minimised when we solve

$$\min_x \max_y \; F_k(x) + E_k(x) + \langle K_k x, y\rangle - G^*_k(y). \tag{2.2}$$

Taking forward-backward steps alternatingly with respect to x and y, we obtain the primal step of the algorithm, as well as the dual step subject to a small modification. The predictor transfers iterates generated this way between time frames and spaces.

In Section 2.1, we provide essential definitions and outline the formal assumptions needed to prove regret bounds for the method in the subsequent Section 2.2.

2.1 ASSUMPTIONS

Write $u^k = (x^k, y^k)$. Then $u^{0:\infty}$ generated by Algorithm 1 from an initial u^0, lies in

$$U_{0:\infty} := (X_0 \times Y_0) \times (X_1 \times Y_1) \times (X_2 \times Y_2) \cdots .$$

Since Algorithm 1 involves the dual variable y^k, we need to expand the set $\mathcal{X} \subset X_{0:\infty}$ of primal comparison sequences in (2.1) to $\mathcal{U} \subset U_{0:\infty}$. We then define the sets of primal and dual comparison sequences by projection as

$$\mathcal{X} := \left\{\bar{x}^{0:\infty} \in X_{0:\infty} \mid (\bar{x}^{0:\infty}, \bar{y}^{0:\infty}) \in \mathcal{U}\right\} \quad \text{and} \quad \mathcal{Y} := \left\{\bar{y}^{0:\infty} \in Y_{0:\infty} \mid (\bar{x}^{0:\infty}, \bar{y}^{0:\infty}) \in \mathcal{U}\right\}.$$

To algorithmically track sequences in these sets, we need the primal-dual predictor P_k. We next formalise this predictor, the comparison sequences, and the functions of (2.1).

Assumption 2.1 (Basic structural assumptions). Given $N \in \mathbb{N}$, on Hilbert spaces X_k and Y_k, $(0 \le k \le N)$, we are provided with:

(i) For all $0 \le k \le N$, convex, proper, and lower semi-continuous $F_k : X_k \to \overline{\mathbb{R}}$, and $G_k^* : Y_k \to \overline{\mathbb{R}}$ (with respective strong convexity factors $\gamma_k \ge 0$ and $\rho_k \ge 0$), as well as $K_k \in \mathbb{L}(X_k; Y_k)$, and a possibly non-convex but finite-valued $E_k : X_k \to \mathbb{R}$.
(ii) A bounded set $\mathcal{U}_{0:N} \subset U_{0:N}$ of primal-dual comparison sequences.
(iii) Primal-dual predictors $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$.

Example 2.2 (Primal-dual predictors). In the numerical experiments of Section 4, for the primal variable x, we will use the incompressible flow predictor $W_k : X_k \to X_{k+1}$, $W_k(x^k) : \xi \mapsto x^k(\xi + h^k(\xi))$, where $h^k(\xi)$ is the predicted displacement at $\xi \in \Omega$ between the frames k and $k+1$. For the dual variable y, we use two different predictions: The first one, $T_k^1 : Y_k \to Y_{k+1}$, seeks to maintain $\langle \nabla_\xi \check{x}^{k+1}, \check{y}^{k+1} \rangle = \langle \nabla_\xi x^k, y^k \rangle$. This is beneficial for preserving total variation [14]: With $G_k(z) = \alpha \|z\|_{2,1}$, at a solution $(\hat{x}^k, \hat{y}^k)$ to (2.2), we have $\langle \nabla_\xi \hat{x}^k, \hat{y}^k \rangle = \alpha \|\nabla_\xi \hat{x}^k\|_{2,1}$. The second one, $T_k^2 : Y_k \to Y_{k+1}$, is defined by the affine update $\check{y}^{k+1} = y^k + c \nabla_\xi \check{x}^{k+1}$. The spatial parameter c is designed to promote sparsity in desired areas. We take c to be inversely proportional to the magnitude of the flow h^k, meaning that sparsity is promoted in areas with less inter-frame movement.

Necessary first-order optimality conditions for the static problem $\min F_k + E_k + G_k \circ K_k$, equivalently (2.2), can be expressed with the general notation $u = (x, y)$ as [12, 11]

$$0 \in H_k(\hat{u}^k) \quad \text{for} \quad H_k(u) := \begin{pmatrix} \partial F_k(x) + \nabla E_k(x) + K_k^* y \\ \partial G_k^*(y) - K_k x \end{pmatrix}, \tag{2.3}$$

Likewise, Algorithm 1 reads in implicit as iteratively solving for u^k the inclusion

$$0 \in \widetilde{H}_k(u^k) + M_k(u^k - \check{u}^k), \tag{2.4}$$

where

$$\widetilde{H}_k(u) := \begin{pmatrix} \partial F_k(x) + \widetilde{\nabla E}_k(\check{x}^k) + K_k^* y \\ \partial G_k^*(y) - K_k x \end{pmatrix} \quad \text{and} \quad M_k = \begin{pmatrix} \tau_k^{-1} \operatorname{Id} & -K_k^* \\ -K_k & \sigma_k^{-1} \operatorname{Id} \end{pmatrix}. \tag{2.5}$$

By avoiding explicit proximal maps, this formulation facilitates convergence analysis [40] and, by extension, regret analysis.

Since we will be taking forward steps with respect to E_k, we will require it to be smooth in an appropriate sense. The next assumption introduces a global online version of the *three-point smoothness* inequality; for the corresponding static inequality, see [41, Appendix B] and [12] for exact gradients, and [15] for inexact gradients based on single-loop splitting methods. For EIT with exact forward operator computations, we reformulate the condition in terms of the PDE itself in Section 3, with some technical results relegated to Appendix B.1. The assumption also introduces compatibility conditions on the *step length parameters* τ_k and σ_k from Algorithm 1, and so-called *testing parameters* that encode both primal and dual convergence rates; see [41, 12]. In this paper, while providing a general theory, we will only apply unaccelerated methods with constant step length and testing parameters.

Assumption 2.3 (Global three-point growth and smoothness inequality). Given $N \in \mathbb{N}$, Assumption 2.1 holds, and for some $\gamma_{E,k}, \lambda_{E,k} \ge 0$, and an error $e_k \ge 0$, for all $1 \le k \le N$:

(i) E_k satisfies for all $\bar{x}^k \in \mathcal{X}_k$ and $x \in X_k$ the "erroneous" three-point smoothness

$$\langle \widetilde{\nabla E}_k(\check{x}^k), x - \bar{x}^k \rangle \ge E_k(x) - E_k(\bar{x}^k) + \frac{\gamma_{E,k}}{2} \|x - \bar{x}^k\|^2 - \frac{\lambda_{E,k}}{2} \|x - \check{x}^k\|^2 - e_k.$$

(ii) We are given step length and testing parameters $\tau_k, \sigma_k > 0$ and $\eta_k, \varphi_k, \psi_k > 0$ with

$$\eta_k = \varphi_k \tau_k = \psi_k \sigma_k, \qquad \text{(primal-dual coupling)} \tag{2.6a}$$

$$1 > \lambda_{E,k} \tau_k + \tau_k \sigma_k \|K_k\|^2 \qquad \text{(metric positivity).} \tag{2.6b}$$

Example 2.4. For an unaccelerated method, we take $\tau_k \equiv \tau$ and $\sigma_k \equiv \sigma$ for some $\tau, \sigma > 0$, along with $\eta_k \equiv \tau$, $\varphi_k \equiv 1$ and $\psi_k \equiv \frac{\tau}{\sigma}$. For accelerated choices for standard (non-online) primal-dual proximal splitting under strong convexity, see [41, 12].

The next assumption presents a local version of the previous global assumption. In what follows, we will assume *either* of these assumptions to hold.

Assumption 2.5 (Local three-point growth and smoothness inequality). Given $N \in \mathbb{N}$, Assumption 2.1 holds, and for all $1 \le k \le N$:

(i) E_k satisfies for some $\bar{\gamma}_{E,k} \in \mathbb{R}$, $\bar{\lambda}_{E,k} \ge 0$, and $\bar{e}_k \ge 0$, for any $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$, and for all $x \in X_k$ the "erroneous" three-point monotonicity-like property

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

(ii) E_k satisfies for some factors $\lambda_{E,k} \ge 0$ and $\hat{\gamma}_{E,k} \ge \gamma_{E,k} \ge 0$, errors $e_k, \hat{e}_k \ge 0$, and a radius $\delta > 0$, for all $\bar{x}^k \in \mathcal{X}_k \cap B(\check{x}^k, \delta)$ and $x \in B(\bar{x}^k, \delta)$, the inequality

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

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 $x \in B(\hat{x}^k, \delta)$ that

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

(iii) We are given step length and testing parameters $\tau_k, \sigma_k > 0$ and $\eta_k, \varphi_k, \psi_k > 0$ as well as a $\kappa_k \in (0, 1)$ with

$$\eta_k = \varphi_k \tau_k = \psi_k \sigma_k, \quad \text{and} \tag{2.7a}$$

$$1 \ge \max \left\{ \lambda_{E,k}, -2(\gamma_k + \bar{\gamma}_{E,k}), \frac{2}{1 - \kappa_k} \bar{\lambda}_{E,k}, \frac{2}{1 - \kappa_k} \hat{\lambda}_{E,k} \right\} \tau_k + \tau_k \sigma_k \|K_k\|^2. \tag{2.7b}$$

Assumption 2.5 (i) is global, while (ii) is local. The former is completely analogous to the second part of the latter, however, the factor $\bar{\gamma}_{E,k}$ is allowed to be negative.

To use the local version of the three-point growth and smoothness inequality, we need additional assumptions on the comparison sequences. Before stating these assumptions, we formalise the predictors P_k of Algorithm 1. To do so, we introduce

$$\Gamma_k := \operatorname{diag}\big((\gamma_k + \gamma_{E,k}) \operatorname{Id},\ \rho_k \operatorname{Id}\big) \quad \text{and} \quad \Omega_k := \operatorname{diag}\big(\lambda_{E,k} \operatorname{Id},\ 0\big).$$

For use with Assumption 2.5 (i) and the first part of (ii), we also define

$$\bar{\Gamma}_k := 2 \operatorname{diag}\big((\gamma_k + \bar{\gamma}_{E,k}) \operatorname{Id},\ \rho_k \operatorname{Id}\big), \qquad \bar{\Omega}_k := 2 \operatorname{diag}\big(\bar{\lambda}_{E,k} \operatorname{Id},\ 0\big),$$

$$\hat{\Gamma}_k := 2 \operatorname{diag}\big((\gamma_k + \hat{\gamma}_{E,k}) \operatorname{Id},\ \rho_k \operatorname{Id}\big), \quad \text{and} \qquad \hat{\Omega}_k := 2 \operatorname{diag}\big(\hat{\lambda}_{E,k} \operatorname{Id},\ 0\big).$$

The first part of the next assumption ensures that the set of comparison sequences is large enough to have primal-dual critical points of the static objectives in its proximity. The second part is a more technical condition on the uniformity of bounds. To state the assumption, recalling that $\check{u}^{k+1} := P_k(u^k)$, we define the *prediction errors*

$$\varepsilon_{k+1}^{\dagger}(u^k, \bar{u}^{k:k+1}) := \frac{1}{2} [\![\check{u}^{k+1} - \bar{u}^{k+1}]\!]^2_{\eta_{k+1} M_{k+1}} - \frac{1}{2} [\![u^k - \bar{u}^k]\!]^2_{\eta_k (M_k + \Gamma_k)} \quad \text{for all} \quad \bar{u}^{k:k+1} \in \mathcal{U}_{k:k+1}. \tag{2.8}$$

They measure the difference of deviation from a chosen comparison sequence between the current iterate and its prediction. Since u^k is always the algorithm-generated iterate, when $\bar{u}^{k:k+1}$ is clear from the context, we write for brevity $\varepsilon_{k+1}^{\dagger} := \varepsilon_{k+1}^{\dagger}(u^k, \bar{u}^{k:k+1})$. For brevity, we also write

$$H_{0:N}(u^{0:N}) := H_0(u^0) \times \cdots \times H_N(u^N) \subset U_{0:N}.$$

Assumption 2.6 (Critical point proximity). Let $N \in \mathbb{N}$ be given, and Assumption 2.5 to be assumed (instead of Assumption 2.3). Then:

(i) For some $\hat{r}_k > 0$, $(1 \le k \le N)$, we have

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

(ii) For some $\xi_k, \Delta > 0$ as well as $\tilde{\delta} \in (0, \delta)$, $(1 \le k \le N)$, for

$$\theta_k = \varphi_k(1 + 2\tau_k \min\{\gamma_k + \bar{\gamma}_{E,k}, \bar{\lambda}_{E,k}\} - \tau_k \sigma_k \|K_k\|^2) > 0$$

(where the positivity is a consequence of (2.7b)) for all $\bar{u}^{0:N} \in \mathcal{U}_{0:N}$

$$0 < d_N(\bar{u}^{0:N}) := \inf_{0 \le n \le N} \left(\frac{\theta_{n+1}(\delta - \tilde{\delta})^2}{\xi_{n+1}} - \frac{1+\Delta}{\xi_{n+1}} \eta_{n+1} \hat{r}_{n+1}^2 - 2\varepsilon_{n+1}^{\dagger} - \sum_{k=1}^{n} \left(\frac{1+\Delta}{2\kappa_k} \eta_k \hat{r}_k^2 + \varepsilon_k^{\dagger} + \eta_k \hat{e}_k \right) \right) \tag{2.9a}$$

as well as

$$(2 - \kappa_k)(\xi_k^{-1} + 1)(\delta - \tilde{\delta})^2 + 2\theta_k^{-1} \eta_k \bar{e}_k \le \tilde{\delta}^2. \tag{2.9b}$$

Remark 2.7. We must have $\theta_{n+1}(\delta - \tilde{\delta})^2 \ge (1+\Delta)\eta_{n+1}\hat{r}_{n+1}^2$ and $2\theta_k^{-1}\eta_k \bar{e}_k \le \tilde{\delta}^2$ for (2.9) to hold. Therefore, given these bounds, the optimal choice of ξ_k is to solve it from (2.9b) as an equality, and insert the result into (2.9a).

Remark 2.8. Similarly to θ_k, also $\hat{r}_k$ depends on the testing parameter φ_k, so its mere increase is not sufficient to satisfy $d_N(\bar{u}^{0:N}) > 0$. For constant step lengths, $\varphi_k \equiv 1$, $\eta_k \equiv \tau$, and $\xi_k \equiv \kappa_k \equiv \kappa$, provided that $\gamma_{n+1} + \bar{\gamma}_{E,n+1} \ge 0$ for all $1 \le n \le N$, taking and ensuring

$$\bar{\theta}_N := 1 - \tau\sigma \sup_{1 \le k \le N} \|K_k\|^2 > 0,$$

(2.9) hold when

$$\bar{\theta}_N(\delta - \tilde{\delta})^2 > \sum_{k=1}^{N+1} \left((1+\Delta)\tau \hat{r}_k^2 + 2\kappa \varepsilon_k^{\dagger} + \tau\kappa \hat{e}_k \right) \quad \text{and}$$
$$(2 - \kappa)(\kappa^{-1} + 1)(\delta - \tilde{\delta})^2 + 2\bar{\theta}_N^{-1} \tau \bar{e}_k \le \tilde{\delta}^2.$$

That is, we can satisfy the conditions by taking τ and $\tau\sigma$ small enough, $\tilde{\delta} < \delta$ large enough, given that the prediction errors $\varepsilon_k^{\dagger}$ have small sums compared to the radius $\delta > 0$. However, large critical point proximities $\hat{r}_k$ and gradient errors $\bar{e}_k$ and $\hat{e}_k$ can be compensated for by small step lengths, as long as their sums are bounded.

2.2 REGRET ANALYSIS

We now analyse the regret of Algorithm 1. The main work is with the local Assumption 2.5 for the nonconvex function E_k. Readers only interested in the global Assumption 2.3, or the general ideas, may skip Lemmas 2.10 and 2.11.

We first verify the positive semi-definiteness of operators central to our analysis.

Lemma 2.9. *Let $N \ge 1$ and suppose Assumption 2.1 holds for any $0 \le k \le N$ and $u^{1:N}$ generated by Algorithm 1 for an initial $u^0 \in X_0 \times Y_0$. Then the following statements hold:*

(i) The operators Γ_k, $\hat{\Gamma}_k$, Ω_k, and $\hat{\Omega}_k$ are positive semi-definite.

(ii) If Assumption 2.3 holds, then M_k and $M_k - \Omega_k$ are positive semi-definite.

(iii) If Assumption 2.5 holds, then M_k and $M_k - \Omega_k$ are positive semi-definite, as well as

$$(1-\kappa_k)M_k \geq \bar{\Omega}_k, \qquad \eta_k(M_k + \bar{\Omega}_k) \geq \operatorname{diag}\big(\theta_k \operatorname{Id},\ 0\big), \qquad \hat{\Gamma}_k \geq 2\Gamma_k,$$
$$(1-\kappa_k)M_k \geq \hat{\Omega}_k, \qquad \eta_k(M_k + \bar{\Gamma}_k) \geq \operatorname{diag}\big(\theta_k \operatorname{Id},\ 0\big)$$

for $\theta_k = \varphi_k(1 + 2\tau_k \min\{\gamma_k + \bar{\gamma}_{E,k}, \bar{\lambda}_{E,k}\} - \tau_k\sigma_k\|K_k\|^2) > 0$ as in Assumption 2.6.

Proof. The positive semi-definiteness of Γ_k, $\hat{\Gamma}_k$, Ω_k, and $\hat{\Omega}_k$ follows from $\gamma_k, \rho_k, \lambda_{E,k}, \gamma_{E,k}, \hat{\gamma}_{E,k}, \hat{\lambda}_{E,k} \geq 0$. This establishes (i).

For (ii), using Assumption 2.3 and Young's inequality, we have

$$\eta_k(M_k - \Omega_k) = \eta_k \begin{pmatrix} (\tau_k^{-1} - \lambda_{E,k}) \operatorname{Id} & -K_k^* \\ -K_k & \sigma_k^{-1} \operatorname{Id} \end{pmatrix} \geq \varphi_k \begin{pmatrix} \operatorname{Id} - \tau_k\lambda_{E,k} - \tau_k\sigma_k K_k^* K_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus, (2.6b) establishes the positive semi-definiteness of M_k and $M_k - \Omega_k$.

Finally, for (iii), using (2.7b) in Assumption 2.5, and Young's inequality, we estimate

$$\begin{aligned}(1-\kappa_k)M_k - \bar{\Omega}_k &= (1-\kappa_k) \begin{pmatrix} \big(\tau_k^{-1} - \frac{2\bar{\lambda}_{E,k}}{1-\kappa_k}\big) \operatorname{Id} & -K_k^* \\ -K_k & \big(\sigma_k^{-1} + \frac{2\rho_k}{1-\kappa_k}\big) \operatorname{Id} \end{pmatrix} \\ &\geq (1-\kappa_k) \begin{pmatrix} \big(\tau_k^{-1} - \frac{2\bar{\lambda}_{E,k}}{1-\kappa_k}\big) \operatorname{Id} - \sigma_k K_k^* K_k & 0 \\ 0 & 0 \end{pmatrix} \geq 0,\end{aligned}$$

which also shows that $M_k \geq 0$. Likewise, we show $(1-\kappa_k)M_k \geq \hat{\Omega}_k$. Moreover,

$$\begin{aligned}\eta_k(M_k + \bar{\Gamma}_k) &= \eta_k \begin{pmatrix} \Big(\tau_k^{-1} + 2(\gamma_k + \bar{\gamma}_{E,k})\Big) \operatorname{Id} & -K_k^* \\ -K_k & \Big(\sigma_k^{-1} + 2\rho_k\Big) \operatorname{Id} \end{pmatrix} \\ &\geq \varphi_k \begin{pmatrix} \big(1 + 2\tau_k(\gamma_k + \bar{\gamma}_{E,k})\big) \operatorname{Id} - \tau_k\sigma_k K_k^* K_k & 0 \\ 0 & 0 \end{pmatrix} \geq \begin{pmatrix} \theta_k \operatorname{Id} & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

The claim regarding $\eta_k(M_k + \bar{\Omega}_k)$ follows by a similar argument, as does the positive semi-definiteness of $M_k - \Omega_k$. Finally, $\hat{\Gamma}_k - 2\Gamma_k = \operatorname{diag}\big((\hat{\gamma}_{E,k} - \gamma_{E,k}) \operatorname{Id},\ 0\big)$ by the definitions of Γ_k and $\hat{\Gamma}_k$. Since $\hat{\gamma}_{E,k} \geq \gamma_{E,k}$ by Assumption 2.5 (ii), this proves $\hat{\Gamma}_k \geq 2\Gamma_k$. □

Lemma 2.9 justifies the norm notation $\|\cdot\|_M$ for $M = M_k, M_k + \Gamma_k, M_k - \Omega_k$. Now, using the critical point proximity Assumption 2.6 and the global Lipschitz-like bound of Assumption 2.5 (i), and assuming sufficient proximity of the previous iterate u^{k-1} to the corresponding time index of a comparison sequence, the next "a priori" lemma shows that the current primal iterate x^k is in the ball where the local three-point inequalities of Assumption 2.5 (ii) hold. The "a posteriori" lemma that follows, will then show that, in fact, x^k is also in the proximity of a comparison sequence. An inductive argument will then easily establish regret estimates as in the convex case [14].

Lemma 2.10 (A priori estimate). *Let $N \geq 1$ and suppose Assumptions 2.1, 2.5 and 2.6 hold, and that $u^{1:N}$ and $\check{u}^{1:N}$ are generated by Algorithm 1 for an initial $u^0 \in X_0 \times Y_0$. Further, let $\bar{u}^{0:N} \in \mathcal{U}_{0:N}$ and $\Delta > 0$, and for a $1 \leq k \leq N$, suppose that*

$$\|u^{k-1} - \bar{u}^{k-1}\|^2_{\eta_{k-1}(M_{k-1}+\Gamma_{k-1})} \leq \frac{\theta_k(\delta - \tilde{\delta})^2}{\xi_k} - \frac{1+\Delta}{\xi_k}\eta_k \hat{r}_k^2 - 2\varepsilon_k^\dagger \tag{2.10}$$

for the prediction error $\varepsilon_k^\dagger = \varepsilon_k^\dagger(u^{k-1}, \bar{u}^{k-1:k})$ of (2.8), the constants $c_k, \kappa_k, \theta_k, \xi_k, \hat{r}_k, \tilde{\delta}$ and radius $\delta > 0$ of Assumptions 2.5 and 2.6. Then there exists $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$ with

$$\|\bar{u}^k - \hat{u}^k\|^2_{M_k + \hat{\Gamma}_k} \leq (1+\Delta)\hat{r}_k^2, \tag{2.11}$$

as well as

$$\|x^k - \hat{x}^k\| < \delta, \quad \|\check{x}^k - \hat{x}^k\| < \delta, \quad \|\check{x}^k - \bar{x}^k\| < \delta, \quad \textit{and} \quad \|x^k - \bar{x}^k\| < \delta. \tag{2.12}$$

Proof. By Assumption 2.6 (i), (2.11) holds for some $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$. By the definition of H_k, we have $\hat{q}^k := -\nabla E_k(\hat{x}^k) - K_k^* \hat{y}^k \in \partial F_k(\hat{x}^k)$ and $\hat{p}^k := K_k \hat{x}^k \in \partial G_k^*(\hat{y}^k)$. Now Cauchy-Schwartz inequality, Young's inequality and the (strong) monotonicity of ∂F_k and ∂G_k^* together with the erroneous three-point monotonicity-like property of E_k and $\widetilde{\nabla E}_k$ in Assumption 2.5 (i) yield

$$\begin{aligned}
\langle \widetilde{H}_k(x^k, y^k), u^k - \hat{u}^k \rangle &= \langle \partial F_k(x^k) - \hat{q}^k, x^k - \hat{x}^k \rangle + \langle \partial G_k^*(y^k) - \hat{p}^k, y^k - \hat{y}^k \rangle \\
&\quad + \langle \widetilde{\nabla E}_k(\check{x}^k) - \nabla E_k(\hat{x}^k), x^k - \hat{x}^k \rangle \\
&\quad + \langle K_k^*(y^k - \hat{y}^k), x^k - \hat{x}^k \rangle - \langle K_k(x^k - \hat{x}^k), y^k - \hat{y}^k \rangle \\
&\geq (\gamma_k + \bar{\gamma}_{E,k}) \|x^k - \hat{x}^k\|^2 + \rho_k \|y^k - \hat{y}^k\|^2 - \bar{\lambda}_{E,k} \|\check{x}^k - \hat{x}^k\|^2 - \bar{e}_k \\
&= \frac{1}{2} [\![u^k - \hat{u}^k]\!]_{\bar{\Gamma}_k}^2 - \frac{1}{2} \|\check{u}^k - \hat{u}^k\|_{\bar{\Omega}_k}^2 - \bar{e}_k.
\end{aligned}$$

Applying here (2.4) and the Pythagoras' identity

$$\langle u^k - \check{u}^k, u^k - \hat{u}^k \rangle_{M_k} = \frac{1}{2} \|u^k - \check{u}^k\|_{M_k}^2 + \frac{1}{2} \|u^k - \hat{u}^k\|_{M_k}^2 - \frac{1}{2} \|\hat{u}^k - \check{u}^k\|_{M_k}^2,$$

we get

$$\frac{1}{2} \|\check{u}^k - \hat{u}^k\|_{M_k + \bar{\Omega}_k}^2 \geq \frac{1}{2} \|u^k - \hat{u}^k\|_{M_k + \bar{\Gamma}_k}^2 + \frac{1}{2} \|u^k - \check{u}^k\|_{M_k}^2 - \bar{e}_k. \tag{2.13}$$

The operators M_k, $M_k + \bar{\Gamma}_k$, and $M_k + \bar{\Omega}_k$ are positive semi-definite by Lemma 2.9.

Since $\xi_k > 0$ by Assumption 2.6 (ii), Young's inequality gives

$$\|\check{u}^k - \hat{u}^k\|_{M_k + \bar{\Omega}_k}^2 \leq (1 + \xi_k) \|\check{u}^k - \bar{u}^k\|_{M_k + \bar{\Omega}_k}^2 + (1 + \xi_k^{-1}) \|\bar{u}^k - \hat{u}^k\|_{M_k + \bar{\Omega}_k}^2.$$

By Lemma 2.9, $M_k + \bar{\Omega}_k \leq (2 - \kappa_k) M_k$. By the definition of the prediction errors in (2.8)

$$\frac{1}{2} \|\check{u}^k - \bar{u}^k\|_{\eta_k M_k}^2 \leq \frac{1}{2} \|u^{k-1} - \bar{u}^{k-1}\|_{\eta_{k-1}(M_{k-1} + \Gamma_{k-1})}^2 + \varepsilon_k^\dagger$$

Using the above inequalities and $\Gamma_k \geq 0$ from Lemma 2.9, we thus obtain

$$\frac{1}{2} \|\check{u}^k - \hat{u}^k\|_{\eta_k (M_k + \bar{\Omega}_k)}^2 \leq \frac{(2 - \kappa_k)(1 + \xi_k)}{2} \left(\|u^{k-1} - \bar{u}^{k-1}\|_{\eta_{k-1}(M_{k-1} + \Gamma_{k-1})}^2 + 2\varepsilon_k^\dagger + \xi_k^{-1} \|\bar{u}^k - \hat{u}^k\|_{\eta_k M_k}^2 \right).$$

We multiply (2.13) by η_k and use it on the left hand side. Then we use (2.10) and (2.11) on the right hand side. This yields

$$\frac{1}{2} \|u^k - \hat{u}^k\|_{\eta_k (M_k + \bar{\Gamma}_k)}^2 + \frac{1}{2} \|u^k - \check{u}^k\|_{\eta_k M_k}^2 \leq \frac{1}{2} \|\check{u}^k - \hat{u}^k\|_{\eta_k (M_k + \bar{\Omega}_k)}^2 \leq \frac{1}{2} \alpha_k \theta_k \tag{2.14}$$

for $\alpha_k := (2 - \kappa_k)(\xi_k^{-1} + 1)(\delta - \tilde{\delta})^2 + 2\theta_k^{-1} \eta_k \bar{e}_k$. By Assumption 2.6 (ii), we have $\alpha_k \leq \tilde{\delta}^2 < \delta^2$. Thus Lemma 2.9 and (2.14) show

$$\begin{aligned}
&\delta^2 \theta_k \geq \alpha_k \theta_k \geq \|\check{u}^k - \hat{u}^k\|_{\eta_k (M_k + \bar{\Omega}_k)}^2 \geq \theta_k \|\check{x}^k - \hat{x}^k\|^2 \quad \text{as well as} \\
&\delta^2 \theta_k \geq \alpha_k \theta_k \geq \|u^k - \hat{u}^k\|_{\eta_k (M_k + \bar{\Gamma}_k)}^2 \geq \theta_k \|x^k - \hat{x}^k\|^2.
\end{aligned} \tag{2.15}$$

This shows the first and second parts of of (2.12).

By Lemma 2.9 and (2.10), we have $\|u\|^2_{\eta_k M_k} \geq \theta_k \|x\|^2$, $\hat{\Gamma}_k \geq 0$, and $\eta_k \hat{r}_k^2 < (1+\Delta)\eta_k \hat{r}_k^2 < \theta_k(\delta - \tilde{\delta})^2$. Thus using Assumption 2.5 (i), we establish

$$\theta_k \|\bar{x}^k - \hat{x}^k\|^2 \leq \|\bar{u}^k - \hat{u}^k\|^2_{\eta_k M_k} \leq \|\bar{u}^k - \hat{u}^k\|^2_{\eta_k(M_k + \hat{\Gamma}_k)} \leq \eta_k \hat{r}_k^2 < \theta_k(\delta - \tilde{\delta})^2.$$

Hence $\|\bar{x}^k - \hat{x}^k\| \leq \delta - \tilde{\delta}$. By Assumption 2.6 (ii), $\alpha_k < \tilde{\delta}^2$. Minding (2.15), it follows that

$$\|\check{x}^k - \bar{x}^k\| \leq \|\check{x}^k - \hat{x}^k\| + \|\bar{x}^k - \hat{x}^k\| \leq \sqrt{\alpha_k} + \delta - \tilde{\delta} < \delta,$$

and completely analogously $\|x^k - \bar{x}^k\| < \delta$, proving last two inequalities of (2.12). □

Now that we know that the curent primal iterate belongs to a ball where the estimates of Assumption 2.5 (ii) hold, we can continue with the aforementioned a posteriori estimate.

Lemma 2.11 (A posteriori estimate). *Let $N \geq 1$ and suppose Assumptions 2.1, 2.5 and 2.6 hold, $\bar{u}^{0:N} \in \mathcal{U}_{0:N}$. Let $u^{1:N}$ and $\check{u}^{1:N}$ be generated by Algorithm 1 for an initial $u^0 \in X_0 \times Y_0$ satisfying for d_N given in Assumption 2.6 the local initialisation condition*

$$\frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0(M_0+\Gamma_0)} \leq d_N(\bar{u}^{0:N}). \tag{2.16}$$

Then for all $1 \leq n \leq N$, the three-point smoothness and growth inequality

$$E_n(x^n) - E_n(\bar{x}^n) + \frac{\gamma_{E,k}}{2}\|\bar{x}^n - x^n\|^2 - \frac{\lambda_{E,k}}{2}\|x^n - \check{x}^n\|^2 \leq \langle \widetilde{\nabla E}_n(\check{x}^n), x^n - \bar{x}^n\rangle + e_k \tag{2.17}$$

holds, as does the tracking inequality

$$\frac{1}{2}\|u^n - \bar{u}^n\|^2_{\eta_n(M_n+\Gamma_n)} < \frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0(M_0+\Gamma_0)} + \sum_{k=1}^{n}\left(\varepsilon_k^\dagger + \frac{1}{\kappa_k}\hat{r}_k^2\right). \tag{2.18}$$

Proof. The estimate (2.17) follows directly from Assumption 2.5 if we show that the iterates and predictions $x^k, \check{x}^k \in B(\bar{x}^k, \delta)$. This follows from Lemma 2.10, if we prove its assumptions, i.e., (2.10) for all $1 \leq k \leq N$. We do this by induction, and in the course of it, also prove (2.18). That is, we prove by induction, for all $1 \leq n \leq N$ that

$$\frac{1}{2}\|u^{k-1} - \bar{u}^{k-1}\|^2_{\eta_{k-1}(M_{k-1}+\Gamma_{k-1})} \leq \frac{\theta_k(\delta - \tilde{\delta})^2}{\xi_k} - \frac{1+\Delta}{\xi_k}\eta_k \hat{r}_k^2 - 2\varepsilon_k^\dagger \quad \text{for all } 1 \leq k \leq n. \tag{2.19}$$

The inductive basis, i.e., (2.19) for for $n = 1$, follows directly from (2.16) and the definition of d_N in Assumption 2.6. For the inductive step, suppose that (2.19) holds for some $1 \leq n < N$. We will prove that the same holds for $n+1$. Towards this end, let $k = 1, \ldots, n$ be arbitrary. The inductive assumption (2.19) directly proves the assumption (2.10) of Lemma 2.10 up to $k = n$. By the lemma, $x^k, \check{x}^k \in B(\hat{x}^k, \delta)$ for some $\hat{u}^k = (\hat{x}^k, \hat{y}^k) \in H_k^{-1}(0)$, so the latter inequality of Assumption 2.5 (ii) gives

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

Recall the expression for $\widetilde{H}_k$ from (2.5) and H_k from (2.3). Since $0 \in H_k(\hat{x}^k, \hat{y}^k)$, we have $\hat{q}^k := -\nabla E_k(\hat{x}^k) - K_k^* \hat{y}^k \in \partial F_k(\hat{x}^k)$ and $\hat{p}^k := K_k \hat{x}^k \in \partial G_k^*(\hat{y}^k)$. These together with the above three-point inequality and the (strong) monotonicity of ∂F_k and ∂G_k^* yield

$$\begin{aligned}
\langle \widetilde{H}_k(x^k, y^k), u^k - \hat{u}^k\rangle &= \langle \partial F_k(x^k) - \hat{q}^k, x^k - \hat{x}^k\rangle + \langle \widetilde{\nabla E}_k(\check{x}^k) - \nabla E_k(\hat{x}^k), x^k - \hat{x}^k\rangle \\
&\quad + \langle \partial G_k^*(y^k) - \hat{p}^k, y^k - \hat{y}^k\rangle \\
&\quad + \langle K_k^*(y^k - \hat{y}^k), x^k - \hat{x}^k\rangle - \langle K_k(x^k - \hat{x}^k), y^k - \hat{y}^k\rangle \\
&\geq \left(\gamma_k + \hat{\gamma}_{E,k}\right)\|x^k - \hat{x}^k\|^2 + \rho_k\|y^k - \hat{y}^k\|^2 - \hat{\lambda}_{E,k}\|x^k - \check{x}^k\|^2 - \hat{e}_k \\
&= \frac{1}{2}\|u^k - \hat{u}^k\|^2_{\hat{\Gamma}_k} - \frac{1}{2}\|u^k - \check{u}^k\|^2_{\hat{\Omega}_k} - \hat{e}_k.
\end{aligned} \tag{2.20}$$

Applying the linear "testing operator" $\langle \cdot, u^k - \hat{u}^k \rangle$ to both sides of (2.4) then yields

$$0 \geq \langle u^k - \check{u}^k, u^k - \hat{u}^k \rangle_{M_k} + \frac{1}{2}\|u^k - \hat{u}^k\|^2_{\hat{\Gamma}_k} - \frac{1}{2}\|u^k - \check{u}^k\|^2_{\hat{\Omega}_k} - \hat{e}_k.$$

Since M_k is positive semi-definite by Lemma 2.9, as are $M_k + \hat{\Gamma}_k$ and $M_k - \hat{\Omega}_k$, they all induce semi-norms. Similarly to the proof of Lemma 2.10, the Pythagoras' identity yields

$$\frac{1}{2}\|\check{u}^k - \hat{u}^k\|^2_{M_k} \geq \frac{1}{2}\|u^k - \hat{u}^k\|^2_{M_k+\hat{\Gamma}_k} + \frac{1}{2}\|u^k - \check{u}^k\|^2_{M_k-\hat{\Omega}_k} - \hat{e}_k$$

Using here the Pythagoras' identitities

$$\begin{aligned}\|u^k - \hat{u}^k\|^2_{M_k+\hat{\Gamma}_k} &= \|u^k - \bar{u}^k\|^2_{M_k+\hat{\Gamma}_k} + \|\bar{u}^k - \hat{u}^k\|^2_{M_k+\hat{\Gamma}_k} + 2\langle u^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{M_k+\hat{\Gamma}_k} \quad \text{and} \\ \|\check{u}^k - \hat{u}^k\|^2_{M_k} &= \|\check{u}^k - \bar{u}^k\|^2_{M_k} + \|\bar{u}^k - \hat{u}^k\|^2_{M_k} + 2\langle \check{u}^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{M_k},\end{aligned}$$

we obtain

$$\begin{aligned}&\frac{1}{2}\|\check{u}^k - \bar{u}^k\|^2_{M_k} + \langle \check{u}^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{M_k} + \hat{e}_k \\ &\geq \frac{1}{2}\|u^k - \bar{u}^k\|^2_{M_k+\hat{\Gamma}_k} + \frac{1}{2}\|\bar{u}^k - \hat{u}^k\|^2_{\hat{\Gamma}_k} + \frac{1}{2}\|u^k - \check{u}^k\|^2_{M_k-\hat{\Omega}_k} + \langle u^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{M_k+\hat{\Gamma}_k}.\end{aligned}$$

Splitting $\langle \check{u}^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{M_k} - \langle u^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{M_k} = \langle \check{u}^k - u^k, \bar{u}^k - \hat{u}^k \rangle_{M_k}$, multiplying by η_k, using the definition (2.8) of the prediction errors, and rearranging, we obtain

$$\begin{aligned}&\frac{1}{2}\|u^{k-1} - \bar{u}^{k-1}\|^2_{\eta_{k-1}(M_{k-1}+\Gamma_{k-1})} + (\varepsilon_k^\dagger + \eta_k \hat{e}_k) + \langle \check{u}^k - u^k, \bar{u}^k - \hat{u}^k \rangle_{\eta_k M_k} \\ &\geq \frac{1}{2}\|u^k - \bar{u}^k\|^2_{\eta_k(M_k+\hat{\Gamma}_k)} + \frac{1}{2}\|\bar{u}^k - \hat{u}^k\|^2_{\eta_k\hat{\Gamma}_k} + \frac{1}{2}\|u^k - \check{u}^k\|^2_{\eta_k(M_k-\hat{\Omega}_k)} + \langle u^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{\eta_k\hat{\Gamma}_k}.\end{aligned} \tag{2.21}$$

Summing both sides of (2.21) over $k = 1, \ldots, n$, telescoping, and rearranging gives

$$\frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0(M_0+\Gamma_0)} + \sum_{k=1}^{n}\big((\varepsilon_k^\dagger + \eta_k \hat{e}_k) + A_k\big) \geq \frac{1}{2}\|u^n - \bar{u}^n\|^2_{\eta_n(M_n+\hat{\Gamma}_n)}, \tag{2.22}$$

for

$$\begin{aligned}A_k := &\langle \check{u}^k - u^k, \bar{u}^k - \hat{u}^k \rangle_{\eta_k M_k} - \langle u^k - \bar{u}^k, \bar{u}^k - \hat{u}^k \rangle_{\eta_k\hat{\Gamma}_k} \\ &- \frac{1}{2}\|\bar{u}^k - \hat{u}^k\|^2_{\eta_k\hat{\Gamma}_k} - \frac{1}{2}\|u^k - \bar{u}^k\|^2_{\eta_k(\hat{\Gamma}_k-\Gamma_k)} - \frac{1}{2}\|u^k - \check{u}^k\|^2_{\eta_k(M_k-\hat{\Omega}_k)},\end{aligned}$$

where we used the notation $\|\cdot\|_{\hat{\Gamma}_k-\Gamma_k}$ since Lemma 2.9 guarantees $\hat{\Gamma}_k \geq \Gamma_k + \hat{\Gamma}_k/2 \geq \Gamma_k$ and both operators are by definition self-adjoint. Using the first of these inequalities, $M_k - \hat{\Omega}_k \geq \kappa_k M_k$ (which also holds by Lemma 2.9), and $0 < \kappa_k < 1$, applying of Young's inequality twice, and rearranging, gives

$$\begin{aligned}A_k \leq &\frac{\kappa_k}{2}\|u^k - \check{u}^k\|^2_{\eta_k M_k} + \frac{1}{2\kappa_k}\|\bar{u}^k - \hat{u}^k\|^2_{\eta_k M_k} + \frac{1}{4}\|u^k - \bar{u}^k\|^2_{\eta_k\hat{\Gamma}_k} + \frac{1}{2}\|\bar{u}^k - \hat{u}^k\|^2_{\eta_k\hat{\Gamma}_k} \\ &- \frac{1}{2}\|u^k - \bar{u}^k\|^2_{\eta_k(\hat{\Gamma}_k-\Gamma_k)} - \frac{1}{2}\|u^k - \check{u}^k\|^2_{\eta_k(M_k-\hat{\Omega}_k)} \\ \leq &\frac{1}{2\kappa_k}\|\bar{u}^k - \hat{u}^k\|^2_{\eta_k(M_k+\hat{\Gamma}_k)} \leq \frac{1+\Delta}{2\kappa_k}\eta_k \hat{r}_k^2.\end{aligned}$$

In the final step we have used (2.11) from Lemma 2.10. Inserting this result in (2.22) and using $\Gamma_N \leq \hat{\Gamma}_N$ from Lemma 2.9, we obtain

$$\frac{1}{2}\|u^n - \bar{u}^n\|^2_{\eta_n(M_n+\Gamma_n)} \leq \frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0(M_0+\Gamma_0)} + \sum_{k=1}^{n}\left(\varepsilon_k^\dagger + \eta_k \hat{e}_k + \frac{1+\Delta}{2\kappa_k}\eta_k \hat{r}_k^2\right).$$

Since $\Delta < 1$, this directly proves the tracking estimate (2.18). Using the assumed (2.16), with the definition of d_N from Assumption 2.6, we also obtain (2.19) for $k = n + 1$, finishing the induction step. As mentioned in the beginning of the proof, it now follows from Lemma 2.10 that $x^k, \check{x}^k \in B(\bar{x}^k, \delta)$ for all $1 \le k \le N$, so that (2.17) follows from Assumption 2.5 (ii). □

With this, we are now ready to show the main theorems. For brevity, we define the initialisation and cumulative prediction and gradient error

$$e_N^{\Sigma}(u^{0:N-1}, \bar{u}^{0:N}) := \sum_{k=0}^{N-1} \left(\varepsilon_{k+1}^{\dagger}(u^k, \bar{u}^{k:k+1}) + \eta_{k+1} e_{k+1} \right) \tag{2.23}$$

and the Lagrangian duality gap

$$\mathcal{G}_k^H(u, \bar{u}) := \big([F_k + E_k](x) + \langle K_k x, \bar{y} \rangle - G_k^*(\bar{y}) \big) - \big([F_k + E_k](\bar{x}) + \langle K_k^* y, \bar{x} \rangle - G_k^*(y) \big).$$

The next theorem shows that the sum of the Lagrangian duality gaps is bounded by the initialisation and the cumulative prediction error.

Theorem 2.12. *Let $N \ge 1$ and suppose Assumption 2.1 holds for $u^{1:N}$ generated by Algorithm 1 for an initial $u^0 \in X_0 \times Y_0$. Let $\bar{u}^{0:N} \in \mathcal{U}^{0:N}$, and suppose that either*

(i) The global Assumption 2.3 holds, or

(ii) The local Assumptions 2.5 and 2.6 hold, and, for d_N given in Assumption 2.6, we have the local initialisation bound

$$\frac{1}{2} \|u^0 - \bar{u}^0\|^2_{\eta_0(M_0 + \Gamma_0)} \le d_N(\bar{u}^{0:N}). \tag{2.24}$$

Then $M_k + \Gamma_k$ and $M_k - \Omega_k$ are positive semi-definite and with $\check{u}^k := P_k(u^{k-1})$,

$$\sum_{k=1}^{N} \left(\mathcal{G}_k^H(u^k, \bar{u}^k) + \frac{1}{2} \|u^k - \check{u}^k\|^2_{\eta_k(M_k - \Omega_k)} \right) \le \frac{1}{2} \|u^0 - \bar{u}^0\|^2_{\eta_0(M_0 + \Gamma_0)} + e_N^{\Sigma}(u^{0:N-1}, \bar{u}^{0:N}).$$

Proof. Lemma 2.9 proves the positive semi-definiteness of $\eta_k M_k$ and $\eta_k(M_k - \Omega_k)$. Consequently, also $\eta_k M_k + \Gamma_k$ is positive semi-definite.

Fix, for now, $k \in \{1, \dots, N-1\}$. Recall the expression for $\widetilde{H}_k$ from (2.5). In the case (ii), conditions of Lemma 2.11 hold, so we have that

$$\langle \widetilde{\nabla E}_k(\check{x}^k), x^k - \bar{x}^k \rangle \ge E_k(x^k) - E_k(\bar{x}^k) - \frac{\lambda_{E,k}}{2} \|x^k - \check{x}^k\|^2 + \frac{\gamma_{E,k}}{2} \|\bar{x}^k - x^k\|^2 - e_k, \tag{2.25}$$

whenever $\bar{x}^k$ is a component of $\bar{u}^{1:N} = ((\bar{x}^1, \bar{y}^1), \dots, (\bar{x}^N, \bar{y}^N)) \in \mathcal{U}_{1:N}$. The (strong) convexity of F_k and G_k^* with (2.25) yield

$$\begin{aligned} \langle \widetilde{H}_k(x^k, y^k), u^k - \bar{u}^k \rangle &= \langle \partial F_k(x^k), x^k - \bar{x}^k \rangle + \langle \widetilde{\nabla E}_k(\check{x}^k), x^k - \bar{x}^k \rangle \\ &\quad + \langle \partial G_k^*(y^k), y^k - \bar{y}^k \rangle + \langle K_k^* y^k, x^k - \bar{x}^k \rangle - \langle K_k x^k, y^k - \bar{y}^k \rangle \\ &\ge \left(F_k(x^k) - F_k(\bar{x}^k) + \frac{\gamma_k + \gamma_{E,k}}{2} \|x^k - \bar{x}^k\|^2 \right) + \left(E_k(x^k) - E_k(\bar{x}^k) - \frac{\lambda_{E,k}}{2} \|\check{x}^k - x^k\|^2 \right) \\ &\quad + \left(G_k^*(y^k) - G_k^*(\bar{y}^k) + \frac{\rho_k}{2} \|y^k - \bar{y}^k\|^2 \right) - \langle K_k^* y^k, \bar{x}^k \rangle + \langle K_k x^k, \bar{y}^k \rangle - e_k \\ &= \frac{1}{2} \|u^k - \bar{u}^k\|^2_{\Gamma_k} + \mathcal{G}_k^H(x^k, \bar{u}^k) - \frac{1}{2} \|u^k - \check{u}^k\|^2_{\Omega_k} - e_k. \end{aligned} \tag{2.26}$$

The case (i) assumes (2.25) directly, thus (2.26) holds globally.

In both cases, for all $k = 1, \dots, N$, we apply the linear "testing operator" $\langle \cdot, u^k - \bar{u}^k \rangle$ to both sides of (2.4). This followed by (2.26) yields

$$e_k \geq \langle u^k - \check{u}^k, u^k - \bar{u}^k \rangle_{M_k} + \frac{1}{2}\|u^k - \bar{u}^k\|^2_{\Gamma_k} + \mathcal{G}^H_k(u^k, \bar{u}^k) - \frac{1}{2}\|u^k - \check{u}^k\|^2_{\Omega_k}.$$

Multiplying this by η_k, using the Pythagoras' identity to convert the inner product into norms, then continuing with the definition (2.8) of the prediction errors, we obtain

$$\begin{aligned}\frac{1}{2}\|u^{k-1} - \bar{u}^{k-1}\|^2_{\eta_{k-1}(M_{k-1}+\Gamma_{k-1})} + \varepsilon^\dagger_k(\bar{u}^{k-1:k}) + \eta_k e_k &\geq \frac{1}{2}\|\check{u}^k - \bar{u}^k\|^2_{\eta_k M_k} + \eta_k e_k \\ &\geq \frac{1}{2}\|u^k - \bar{u}^k\|^2_{\eta_k(M_k+\Gamma_k)} + \eta_k \mathcal{G}^H_k(u^k, \bar{u}^k) + \frac{1}{2}\|u^k - \check{u}^k\|^2_{\eta_k(M_k-\Omega_k)}.\end{aligned}$$

Summing over $k = 1, \dots, N$ gives

$$\begin{aligned}\frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0(M_0+\Gamma_0)} + e^\Sigma_N(u^{0:N-1}, \bar{u}^{0:N}) \geq{}& \frac{1}{2}\|u^N - \bar{u}^N\|^2_{\eta_N(M_N+\Gamma_N)} \\ &+ \sum_{k=1}^N \Big(\eta_k \mathcal{G}^H_k(u^k, \bar{u}^k) + \frac{1}{2}\|u^k - \check{u}^k\|^2_{\eta_k(M_k-\Omega_k)}\Big).\end{aligned}$$

Since $\frac{1}{2}\|u^N - \bar{u}^N\|^2_{\eta_N M_N+\Gamma_N} \geq 0$, the claim follows. □

The next corollary derives function value estimates from the preceding gap estimates. Its proof is exactly as the proof of [14, Theorem 2.6]. The estimates are with respect to

$$\mathring{G}_{1:N}(z^{1:N}) := \sum_{k=1}^N \sup_{\tilde{y}^k \in \mathcal{Y}_k} \big[\langle z^k, \tilde{y}^{1:N}\rangle - \eta_k G^*_k(\tilde{y}^k)\big] \quad \text{in place of} \quad G_{1:N}(y^{1:N}) := \sum_{k=1}^N \eta_k G_k(y^k),$$

We also denote

$$Q_{1:N}(x^{1:N}) := \sum_{k=1}^N \eta_k [F_k + E_k](x^k), \qquad K_{1:N}x^{1:N} := (\eta_1 K_1 x^1, \dots, \eta_N K_N x^N).$$

If the dual comparison sets $\mathcal{Y}_{1:N}$ were convex, then, recalling the formula $(f_1 + f_2)^* = f_1^* \,\square\, f_2^*$ for infimal convolutions (denoted $\square$) of convex functions f_1 and f_2, we would have $\mathring{G}_{1:N} = G_{1:N} \,\square\, \delta^*_{\mathcal{Y}_{1:N}}$. In general, $\mathring{G}_{1:N} \leq G_{1:N} \,\square\, \delta^*_{\mathcal{Y}_{1:N}}$. That is, $\mathring{G}_{1:N}$ is a "sub-infimal" convolution of $G_{1:N}$ and the temporal coupling. As $G_k \circ K_k$ is typically a total variation type regularisation functional, $\mathring{G}_{1:N} \circ K_{1:N}$ becomes a spatiotemporal regulariser with aspects of spatial total variation, and the temporal properties of the problem at hand.

Corollary 2.13. *Suppose that the assumptions of Theorem 2.12 hold, with the initialisation bound (2.24) for all $\bar{u}^{0:N} \in \mathcal{U}^{0:N}$ in the local convergence case (ii). Then*

$$\begin{aligned}[Q_{1:N}(x^{1:N}) + \mathring{G}_{1:N}(K_{1:N}x^{1:N})] - \sup_{\bar{x}^{1:N} \in \mathcal{X}_{1:N}} [Q_{1:N}(\bar{x}^{1:N}) + \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N})] \\ \leq \sup_{\bar{u}^{0:N} \in \mathcal{U}_{0:N}} \left(\frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0 M_0+\Gamma_0} + c_N(\bar{x}^{1:N}, y^{1:N}) + e^\Sigma_N(u^{0:N-1}, \bar{u}^{0:N})\right),\end{aligned}$$

where $e^\Sigma_N(u^{0:N-1}, \bar{u}^{0:N})$ is given by (2.23), and the comparison set solution discrepancy

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

Proof. The claim follows from Theorem 2.12, by the exact same proof as that of [14, Theorem 2.6], with $e_N^{\Sigma}(u^{0:N-1}, \bar{u}^{0:N})$ in place of $e_N(u^{0:N-1}, \bar{u}^{0:N})$ of the latter. □

Remark 2.14 (Comparison set solution discrepancy). According to [14, Remark 2.7 and Section 3], $c_N \leq 0$ when $G_k \circ K_k = \alpha\|\nabla \cdot \|_{2,1}$ is the total variation, the dual initialisation achieves the total variation (i.e., $\langle y^0, x^0\rangle = \alpha\|\nabla x^0\|_{2,1}$ with $\|y^0\|_{2,\infty} \leq \alpha$), and the dual predictor is total variation preserving. Examples of total variation preserving predictors are provided in [14].

3 THE ONLINE EIT PROBLEM

We now treat the online EIT problem, which we describe in Section 3.1. To apply Theorem 3.6, we need to prove the smoothness inequalities of Assumptions 2.3 and 2.5. Based on auxiliary results from Appendix A, we do this in Section 3.3 after first proving the necessary second-order differentiability of the CEM solution operator in Section 3.2.

3.1 PROBLEM DESCRIPTION

To model the dynamic EIT problem as the optimisation problem (2.1), we now take

$$E_k(x) := \frac{1}{2}\sum_{j=1}^{N_2} \|\Sigma^{-1/2}(I(x, U^{j,k}) - \mathcal{J}^{j,k})\|_2^2, \tag{3.1}$$
$$F_k(x) := \delta_{[x_m, x_M]}(x), \quad \text{and} \quad G_k(K_k x) := \alpha\|x\|_{2,1},$$

where $\Sigma^{-1/2} \in \mathbb{R}^{(N_1-1)\times(N_1-1)}$ is a data precision matrix, modelling noise characteristics, and $\alpha > 0$ is the regularisation parameter. The conductivity x is bounded between $0 < x_m < x_M$. The term $\mathcal{J}^{j,k} \in \mathbb{R}^{N_1}$ is a vector of measurements corresponding to the electrode potentials $U^{j,k} \in \mathbb{R}^{N_1}$ at time instance k. The currents $I_i^{j,k} := I_i(x, U^{j,k})$ for $i = 1, \ldots, N_1$ are obtained by solving $(u^{j,k}, I_1, \ldots, I_{N_1})$ from weak versions of the Complete Eletrode Model (CEM) equations

$$\nabla_\xi \cdot (x^k(\xi)\nabla_\xi u^{j,k}(\xi)) = 0 \qquad \text{for } \xi \in \Omega, \tag{3.2a}$$
$$u^{j,k}(\xi) + \zeta_i x^k(\xi)\nabla_\xi u^{j,k}(\xi)\cdot\nu(\xi) = U_i^{j,k} \qquad \text{for } \xi \in \partial\Omega_{e_i},\ i = 1, \ldots, N_1, \tag{3.2b}$$
$$x^k(\xi)\nabla_\xi u^{j,k}(\xi)\cdot\nu(\xi) = 0 \qquad \text{for } \xi \in \partial\Omega \setminus (\partial\Omega_{e_1} \cup \ldots \cup \partial\Omega_{e_{N_1}}), \tag{3.2c}$$
$$\int_{\partial\Omega_{e_i}} x^k(\xi)\nabla_\xi u^{j,k}(\xi)\cdot\nu(\xi)\ \mathrm{d}s = -I_i^{j,k} \qquad \text{for } i = 1, \ldots, N_1, \tag{3.2d}$$

where $u^{j,k}$ is the inner potential and ζ_i is the contact impedance between the electrode i and the medium inside the domain $\Omega \subset \mathbb{R}^d$. We focus on this "potential-to-current" model, as it aligns with state-of-the-art EIT measurement devices [23]. This model was originally introduced in [43], and subsequently used in e.g. [44, 24, 23]. The conventional "current-to-potential" model solves the potentials given the electric currents.

We assume that $x \in L^\infty(\Omega)$ for a bounded Lipschitz domain[1] $\Omega \subset \mathbb{R}^n$. Given the discretisation of x in the numerical experiments of Section 4, the results of this section remain compatible with the theory of Section 2. For a given x, we write

$$w_x = (v_x, V_x) \in \mathcal{H} := H^1(\Omega) \oplus \mathbb{R}^{N_1} \tag{3.3}$$

[1]The domain Ω is unrelated to the operator $\bar{\Omega}_k$ of Section 2.

for the electrical potential and electrode currents (u, I) that solve (3.2) weakly, i.e.,

$$B_x(w_x, w) = L(w) \quad \text{for all} \quad w = (v, V) \in \mathcal{H}. \tag{3.4}$$

where the bilinear form B_x reads

$$B_x(w_x, w) = \int_\Omega x\nabla_\xi v_x \cdot \nabla_\xi v \,\mathrm{d}\xi + \sum_{i=1}^{N_1} \frac{1}{\zeta_i} \int_{\partial\Omega_{e_i}} v_x(v - V_i)\,\mathrm{d}s + \sum_{i=1}^{N_1} (V_x)_i V_i,$$

and the linear form L is

$$L(w) = \sum_{i=1}^{N_1} \frac{1}{\zeta_i} \int_{\partial\Omega_{e_i}} U_i(v - V_i)\,\mathrm{d}s.$$

We equip $\mathcal{H}$ with the norm

$$\|w\|_{\mathcal{H}}^2 := \|v\|_{H^1}^2 + \|V\|_2^2 \quad \text{for} \quad w \in \mathcal{H}. \tag{3.5}$$

3.2 DIFFERENTIABILITY OF THE EIT SOLUTION OPERATOR

The problem (3.4) is well-posed [38]: see the "potential-to-current" model in [25]. The first order differentiability of the reversed "current-to-potential" model has, moreover, been extensively discussed in earlier works, e.g., [28, 13]. We will prove the first order differentiability of the solution operator of the potential-to-current model. This, to our knowledge, has not been previously proven. We then show the Lipschitz boundedness of this derivative and use it to prove second order differentiability in L^∞. This has only been shown in the finite dimensional case for the current-to-potential model.

For notational clarity, we write $w'_{x,\cdot}$, i.e., $h \mapsto w'_{x,h}$ for the (Fréchet) derivative of $x \mapsto w_x$ at $x \in L^\infty(\Omega)$, when it exists. The notation for v_x is analogous. Thus, $h \mapsto w'_{x,h} \in \mathbb{L}(L^\infty(\Omega); \mathcal{H})$. Likewise, $w''_{x,\cdot,\cdot}$, i.e., $(h_1, h_2) \mapsto w''_{x,h_1,h_2}$ is the second (Fréchet) derivative of $x \mapsto w_x$ at x, when it exists. For clarity, we write $\nabla_\xi w_x$ for the weak gradient of w_x, and $\nabla_\xi w'_{x,h}$ for $\nabla_\xi[\xi \mapsto w'_{x,h}(\xi)]$, i.e., ∇_ξ is always a (spatial) gradient with respect to $\xi \in \Omega$. Given $x \in [x_m, x_M]$, a.e., and that the domain Ω is Lipschitz, the solutions $w_x : X \to \mathcal{H}$ are continuous in L^p for any $p \geq 1$ [26, Remark 2.8]. Moreover, under the domain scaling condition $\zeta_k^{-1}|\partial\Omega_{e_k}| \leq 1$, we have

$$C_1 B_x(w, w) \geq \|w\|_{\mathcal{H}}^2 \quad \text{for all} \quad w \in \mathcal{H} \quad \text{and} \tag{3.6}$$

$$\|w_x\|_{\mathcal{H}} \leq C_2 \|U\|_2 \quad \text{for any } x \in [x_m, x_M]. \tag{3.7}$$

For solutions $w_x \in \mathcal{H}$ to (3.4), the scaling condition $\zeta_k^{-1}|\partial\Omega_{e_k}| \leq 1$ can be removed (see [26, Lemma 2.6 and Theorem 2.7]), and the coefficients then have values

$$C_1 = 2C_\Omega(\Lambda \min\{1, x_m\})^{-1} \quad \text{and} \quad C_2 = \sqrt{2}C_1 \quad \text{for} \quad C_\Omega = (\zeta_m/e_M)^{-\frac{N-2}{N-1}}$$

with $\zeta_m = \min_k \zeta_k$ and $e_M = \max_k |\partial\Omega_{e_k}|$.[2] Throughout, we work with:

Assumption 3.1. $0 < x_m < x_M < \infty$ and $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain.

The next corollary is a simple application of this and the earlier results of [26].

Corollary 3.2. *Let Assumption 3.1 hold. Then for $C_3 := C_1 C_2 \|U\|_2$ we have*

$$\|w_{x_2} - w_{x_1}\|_{\mathcal{H}} \leq C_3 \|x_2 - x_1\|_\infty \quad \textit{for all} \quad x_1, x_2 \in [x_m, x_M].$$

[2] We know that Λ is a finite positive constant but the exact value is unknown. For more details, see [38, Lemma 3.2] for current-to-potential model and potential-to-current model [24, Lemma 2].

Proof. Using Hölder inequality, (3.6) and (3.7), recalling that we write $w_{x_1} = (v_{x_1}, V_{x_1})$ and $w_{x_2} = (v_{x_2}, V_{x_2})$, we estimate

$$\|w_{x_2} - w_{x_1}\|_{\mathcal{H}}^2 \le C_1 B_{x_1}(w_{x_2} - w_{x_1}, w_{x_2} - w_{x_1}) \le C_1 \int_\Omega (x_1 - x_2)(\nabla_\xi v_{x_2} \cdot \nabla_\xi (v_{x_2} - v_{x_1}))\, \mathrm{d}\xi$$
$$\le C_1 \|x_1 - x_2\|_\infty \|w_{x_2}\|_{\mathcal{H}} \|w_{x_2} - w_{x_1}\|_{\mathcal{H}} \le C_1 C_2 \|U\|_2 \|x_2 - x_1\|_\infty \|w_{x_2} - w_{x_1}\|_{\mathcal{H}}.$$

For a detailed derivation of the second inequality, see [26, equation (20)]) □

As promised, we can now show first-order differentiability in $L^\infty(\Omega)$.

Lemma 3.3. *Let Assumption 3.1 hold. Then the solution map $x \mapsto w_x : L^\infty(\Omega) \to \mathcal{H}$ of (3.4) is Fréchet differentiable at any $x \in [x_m, x_M]$ with the Fréchet derivative at x, i.e., the map $(h \mapsto w'_{x,h}) \in \mathbb{L}(L^\infty(\Omega); \mathcal{H})$ norm-bounded by $C_3 = C_1 C_2 \|U\|_2$ and satisfying*

$$B_x(w'_{x,h}, w) = -\int_\Omega h \nabla_\xi v_x \cdot \nabla_\xi v \, d\xi \tag{3.8}$$

and

$$\|w_{x+h} - w_x - w'_{x,h}\|_{\mathcal{H}} \le C_1 C_3 \|h\|_\infty^2 \quad \text{for all} \quad h \in L^\infty(\Omega) \text{ and } w = (v, V) \in \mathcal{H}.$$

Proof. Suppose that w_x is a solution to (3.4) with x. Now, with respect to $w = (v, V)$, the left-hand side of (3.8) is coercive by (3.6) and the right-hand side is clearly linear and bounded. Thus the Lax-Milgram theorem establishes the existence of $w'_{x,h}$. We will show that the mapping $h \mapsto w'_{x,h}$ is bounded and linear with respect to h, and then proceed to confirm that it is, indeed, the Fréchet derivative, i.e., that

$$\lim_{\|h\|_\infty \to 0} \frac{\|w_{x_0+h} - w_{x_0} - w'_{x_1,h}\|_{\mathcal{H}}}{\|h\|_{L^\infty(\Omega)}} = 0. \tag{3.9}$$

To see linearity, take $a, b \in \mathbb{R}$ and notice from (3.8) that, due to the linearity of the right-hand side with respect to h, both w'_{x,ah_1+bh_2} and $aw'_{x,h_1} + bw'_{x,h_2}$ solve (3.8). Hence, due the well-posedness (3.8), w'_{x,ah_1+bh_2}, w and $aw'_{x,h_1} + bw'_{x,h_2}$ must be equal.

To show boundedness, we observe that by the Hölder inequality and (3.6)–(3.8),

$$\|w'_{x,h}\|_{\mathcal{H}}^2 \le C_1 |B_x(w'_{x,h}, w'_{x,h})| = \left| \int_\Omega h \nabla_\xi v_x \cdot \nabla_\xi w'_{x,h} \, \mathrm{d}\xi \right| \le C_1 C_2 \|U\|_2 \|h\|_\infty \|w'_{x,h}\|_{\mathcal{H}}. \tag{3.10}$$

Finally, to confirm (3.9), suppose that $\|h\|_\infty \le x_m/2$. This ensures that the solution w_{x+h} to (3.4) at $x + h$ exists since $0 < x_m/2 \le x + h \le x_M + x_m/2$. By (3.6)

$$\|w_{x+h} - w_x - w'_{x,h}\|_{\mathcal{H}}^2 \le C_1 B_x(w_{x+h} - w_x - w'_{x,h}, w_{x+h} - w_x - w'_{x,h}). \tag{3.11}$$

Moreover, for any $w \in \mathcal{H}$,

$$B_x(w_{x+h} - w_x, w) = B_x(w_{x+h}, w) - B_x(w_x, w) = B_x(w_{x+h}, w) - L(w)$$
$$= B_x(w_{x+h}, w) - B_{x+h}(w_{x+h}, w) = -\int_\Omega h \nabla_\xi v_{x+h} \cdot \nabla_\xi v \, \mathrm{d}\xi,$$

hence

$$B_x(w_{x+h} - w_x - w'_{x,h}, w) = B_x(w_{x+h} - w_x, w) - B_x(w'_{x,h}, w) = -\int_\Omega h \nabla_\xi (v_{x+h} - v_x) \cdot \nabla_\xi v \, \mathrm{d}\xi.$$

Taking $w = w_{x+h} - w_x - w'_{x,h}$ and using Corollary 3.2 we thus estimate

$$
\begin{aligned}
B_x(w,w) &= -\int_\Omega h\nabla_\xi(v_{x+h}-v_x)\cdot\nabla_\xi(v_{x+h}-v_x-v'_{x,h})\,\mathrm{d}\xi \\
&\le \|h\|_\infty \left|\int_\Omega \nabla_\xi(v_{x+h}-v_x)\cdot\nabla_\xi(v_{x+h}-v_x-v'_{x,h})\,\mathrm{d}\xi\right| \\
&\le \|h\|_\infty \|w_{x+h}-w_x\|_{\mathcal{H}}\|w_{x+h}-w_x-w'_{x,h}\|_{\mathcal{H}} \\
&\le \|h\|_\infty C_3\|h\|_\infty\|w_{x+h}-w_x-w'_{x,h}\|_{\mathcal{H}}.
\end{aligned}
\tag{3.12}
$$

Combining (3.11) and (3.12) yields $\|w_{x+h}-w_x-w'_{x,h}\|_{\mathcal{H}} \le C_1C_3\|h\|_\infty^2$, which proves (3.9). □

The following lemma shows that the Fréchet derivative of the solution map is Lipschitz. This will be needed to show the second-order differentiability.

Lemma 3.4. *Let Assumption 3.1 hold. Then, for any given $h \in L^\infty(\Omega)$, the map $x \mapsto w'_{x,h} : L^\infty(\Omega) \to \mathcal{H}$ is Lipschitz with constant $2C_1C_3\|h\|_\infty$, that is*

$$
\|w'_{x_1,h} - w'_{x_2,h}\|_{\mathcal{H}} \le 2C_1C_3\|h\|_\infty\|x_2-x_1\|_\infty \quad \textit{for all} \quad x_1,x_2 \in [x_m,x_M].
$$

Proof. For any $w_1, w_2 \in \mathcal{H}$ we have

$$
B_{x_1}(w_1,w_2) = B_{x_2}(w_1,w_2) + \int_\Omega (x_1-x_2)\nabla_\xi v_1\cdot\nabla_\xi v_2\,\mathrm{d}\xi.
\tag{3.13}
$$

Furthermore, by Lemma 3.3, $w'_{x_1,h}$ and $w'_{x_2,h}$ satisfy

$$
B_{x_1}(w'_{x_1,h},w) = -\int_\Omega h\nabla_\xi v_{x_1}\cdot\nabla_\xi v\,\mathrm{d}\xi \quad \text{and} \quad B_{x_2}(w'_{x_2,h},w) = -\int_\Omega h\nabla_\xi v_{x_2}\cdot\nabla_\xi v\,\mathrm{d}\xi
$$

for any $w \in \mathcal{H}$. Using these and (3.13) yields

$$
\begin{aligned}
A &:= B_{x_1}(w'_{x_1,h}-w'_{x_2,h}, w'_{x_1,h}-w'_{x_2,h}) \\
&= B_{x_1}(w'_{x_1,h}, w'_{x_1,h}-w'_{x_2,h}) - B_{x_2}(w'_{x_2,h}, w'_{x_1,h}-w'_{x_2,h}) \\
&\quad - \int_\Omega (x_1-x_2)\nabla_\xi v'_{x_2,h}\cdot\nabla_\xi(v'_{x_1,h}-v'_{x_2,h})\,\mathrm{d}\xi \\
&= \int_\Omega [h\nabla_\xi(v_{x_2}-v_{x_1}) + (x_2-x_1)\nabla_\xi v'_{x_2,h}]\cdot\nabla_\xi(v'_{x_1,h}-v'_{x_2,h})\,\mathrm{d}\xi \\
&\le \left(\|h\nabla_\xi(v_{x_2}-v_{x_1})\|_2 + \|(x_2-x_1)\nabla_\xi v'_{x_2,h}\|_2\right)\|\nabla_\xi(v'_{x_1,h}-v'_{x_2,h})\|_2.
\end{aligned}
$$

Using the Cauchy-Schwartz inequality and $\|\nabla_\xi v\|_2 \le \|w\|_{\mathcal{H}}$ (see (3.5)) again followed by Corollary 3.2 and $\|w'(x_2)\|_{\mathcal{H}} \le C_3\|h\|_\infty\|U\|_2$ allows us to continue

$$
\begin{aligned}
A &\le \|h\|_\infty\|w_{x_2}-w_{x_1}\|_{\mathcal{H}}\|w'_{x_1,h}-w'_{x_1,h}\|_{\mathcal{H}} + \|x_2-x_1\|_\infty\|w'_{x_2,h}\|\|w'_{x_2,h}-w'_{x_1,h}\|_{\mathcal{H}} \\
&\le \|h\|_\infty C_3\|x_2-x_1\|_\infty\|w'_{x_1,h}-w'_{x_1,h}\|_{\mathcal{H}} + \|x_2-x_1\|_\infty C_3\|h\|_\infty\|w'_{x_1,h}-w'_{x_1,h}\|_{\mathcal{H}}.
\end{aligned}
\tag{3.14}
$$

Finally, (3.6) and (3.14) establish

$$
\begin{aligned}
\|w'_{x_1,h}-w'_{x_2,h}\|_{\mathcal{H}}^2 &\le C_1B_{x_1}(w'_{x_1,h}-w'_{x_2,h}, w'_{x_1,h}-w'_{x_2,h}) \\
&\le 2C_1C_3\|h\|_\infty\|x_2-x_1\|_\infty\|w'_{x_1,h}-w'_{x_2,h}\|_{\mathcal{H}}.
\end{aligned}
$$

Dividing by $\|w'_{x_1,h}-w'_{x_2,h}\|_{\mathcal{H}}$ establishes the claimed Lipschitz constant. □

The next lemma confirms the second-order differentiability of the solution mapping.

Lemma 3.5. *Let Assumption 3.1 hold and $x \in [x_m, x_M]$. Then $h_1 \mapsto w'_{x,h_1} : L^\infty(\Omega) \to \mathcal{H}$ is Fréchet differentiable at all $h_1 \in L^\infty(\Omega)$, the Fréchet derivative $h_2 \to v''_{x,h_1,h_2}$ is norm-bounded in $\mathbb{L}(L^\infty(\Omega) \times L^\infty(\Omega); \mathcal{H})$ by $C_4 = 2C_1C_3\|U\|_2$, and satisfies the linear equation*

$$B_x(w''_{x,h_1,h_2}, w) = -\left(\int_\Omega h_1 \nabla_\xi v'_{x,h_2} \cdot \nabla_\xi v \, d\xi + \int_\Omega h_2 \nabla_\xi v'_{x,h_1} \cdot \nabla_\xi v \, d\xi\right) \tag{3.15}$$

for all $h_2 \in L^\infty(\Omega)$ and $w = (v, V) \in \mathcal{H}$.

Proof. The right-hand side of (3.15) is linear and bounded with respect to w. Hence by the Lax-Milgram theorem there exists a solution w''_{x,h_1,h_2} to (3.15). We will show that $h_2 \mapsto w''_{x,h_1,h_2}$ is the Fréchet derivative of $h_1 \mapsto w'_{x,h_1}$. To see that the former is a candidate for the Fréchet derivative, we first establish the boundedness and the bilinearity of $(h_1, h_2) \mapsto w''_{x,h_1,h_2}$. Indeed, due to (3.6) and (3.15),

$$\begin{aligned}
\|w''_{x,h_1,h_2}\|^2_{\mathcal{H}} &\le C_1 B_x(w''_{x,h_1,h_2}, w''_{x,h_1,h_2}) \\
&\le C_1 \left(\left|\int_\Omega h_1 \nabla_\xi v'_{x,h_2} \cdot \nabla_\xi v''_{x,h_1,h_2} \, \mathrm{d}\xi\right| + \left|\int_\Omega h_2 \nabla_\xi v'_{x,h_1} \cdot \nabla_\xi v''_{x,h_1,h_2} \, \mathrm{d}\xi\right|\right) \\
&\le C_1 \big(\|h_1\|_\infty \|w'_{x,h_2}\|_{\mathcal{H}} + \|h_2\|_\infty \|w'_{x,h_1}\|_{\mathcal{H}}\big) \|w''_{x,h_1,h_2}\|_{\mathcal{H}} \\
&\le 2C_1C_3\|U\|_2 \|h_1\|_\infty \|h_2\|_\infty \|w''_{x,h_1,h_2}\|_{\mathcal{H}}.
\end{aligned}$$

Clearly then $\|w''_{x,h_1,h_2}\|_{\mathcal{H}} \le 2C_1C_3\|U\|_2\|h_1\|_\infty\|h_2\|_\infty$.

To see bilinearity, first observe that the first integral left-hand side of (3.15) is linear with respect to h_1 and the second integral is linear with respect to h_2. Then take $h_2 = af + bg$, for $a, b \in \mathbb{R}$ and $f, g \in L^\infty(\Omega)$. Since $h \mapsto v'_{x,h}$ is linear, we have

$$\begin{aligned}
\int_\Omega h_1 \nabla_\xi v'_{x,af+bg} \cdot \nabla_\xi v \, \mathrm{d}\xi &= \int_\Omega h_1 \nabla_\xi (a v'_{x,f} + b v'_{x,g}) \cdot \nabla_\xi v \, \mathrm{d}\xi \\
&= \int_\Omega h_1 (a \nabla_\xi v'_{x,f} \cdot \nabla_\xi v) \, \mathrm{d}\xi + \int_\Omega h_1 (b \nabla_\xi v'_{x,g} \cdot \nabla_\xi v) \, \mathrm{d}\xi,
\end{aligned}$$

This show sthe linearity of the first term with respect to h_2. The same reasoning shows the linearity of the second term with respect to h_1, confirming the bilinearity of the both terms. Arguing similarly to Lemma 3.3, we conclude that $(h_1, h_2) \mapsto w''_{x,h_1,h_2}$ is bilinear.

Finally, we confirm that $h_2 \mapsto w''_{x,h_1,h_2}$ is the Fréchet derivative of $h_1 \mapsto w'_{x,h_1}$. Let $\|h_2\|_\infty \le x_m/2$ and $h_1 \in L^\infty(\Omega)$. Then

$$\begin{aligned}
&B_x(w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}, w) \\
&= B_x(w'_{x+h_2,h_1}, w) - B_x(w'_{x,h_1}, w) - B_x(w''_{x,h_1,h_2}, w) \\
&= B_{x+h_2}(w'_{x+h_2,h_1}, w) - B_x(w'_{x,h_1}, w) - \int_\Omega h_2 \nabla_\xi v'_{x+h_2,h_1} \cdot \nabla_\xi v \, \mathrm{d}\xi - B_x(w''_{x,h_1,h_2}, w) \\
&= -\int_\Omega h_1 \nabla_\xi (v_{x+h_2} - v_x) \cdot \nabla_\xi v \, \mathrm{d}\xi - \int_\Omega h_2 \nabla_\xi v'_{x+h_2,h_1} \cdot \nabla_\xi v \, \mathrm{d}\xi - B_x(w''_{x,h_1,h_2}, w) \\
&= -\int_\Omega h_1 \nabla_\xi (v_{x+h_2} - v_x - v'_{x,h_2}) \cdot \nabla_\xi v \, \mathrm{d}\xi - \int_\Omega h_2 \nabla_\xi (v'_{x+h_2,h_1} - v'_{x,h_1}) \cdot \nabla_\xi v \, \mathrm{d}\xi,
\end{aligned} \tag{3.16}$$

where on the fourth line we used Lemma 3.3 on the first two bilinear terms, and on the last line we used (3.15). Observe for any $w_1 = (v_1, V_1), w_2 = (v_2, V_2) \in \mathcal{H}$ that

$$\int_\Omega h \nabla_\xi v_1 \cdot \nabla_\xi v_2 \, \mathrm{d}\xi \le \|h\|_\infty \|\nabla_\xi v_1\|_2 \|\nabla_\xi v_2\|_2 \le \|h\|_\infty \|w_1\|_{\mathcal{H}} \|w_2\|_{\mathcal{H}}.$$

Using this, (3.6) and (3.16) with $w = w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}$, as well as the triangle inequality, we obtain

$$
\begin{aligned}
(3.17)\qquad & \|w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}\|_{\mathcal{H}}^2 \\
&\le C_1 |B_x(w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}, w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2})| \\
&\le C_1 \Bigg(\left| \int_\Omega h_1 \nabla_\xi (v_{x+h_2} - v_x - v'_{x,h_2}) \cdot \nabla_\xi (v'_{x+h_2,h_1} - v'_{x,h_1} - v''_{x,h_1,h_2}) \, \mathrm{d}\xi \right| \\
&\qquad + \left| \int_\Omega h_2 \nabla_\xi (v'_{x+h_2,h_1} - v'_{x,h_1}) \cdot \nabla_\xi (v'_{x+h_2,h_1} - v'_{x,h_1} - v''_{x,h_1,h_2}) \, \mathrm{d}\xi \right| \Bigg) \\
&\le C_1 \big(\|h_1\|_\infty \|w_{x+h_2} - w_x - w'_{x,h_2}\|_{\mathcal{H}} + \|h_2\|_\infty \|w'_{x+h_2,h_1} - w'_{x,h_1}\|_{\mathcal{H}} \big) \\
&\qquad \cdot \|w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}\|_{\mathcal{H}}.
\end{aligned}
$$

By Lemmas 3.3 and 3.4, we have, respectively

$$
\|w_{x+h_2} - w_x - w'_{x,h_2}\|_{\mathcal{H}} \le C_1 C_3 \|h_2\|_\infty^2 \quad \text{and} \quad \|w'_{x,h_1} - w'_{x+h_2,h_1}\|_{\mathcal{H}} \le 2 C_1 C_3 \|h_1\|_\infty \|h_2\|_\infty.
$$

Dividing (3.17) by $\|w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}\|_{\mathcal{H}}$ and using these estimates establishes

$$
\|w'_{x+h_2,h_1} - w'_{x,h_1} - w''_{x,h_1,h_2}\|_{\mathcal{H}} \le C_1^2 C_3 (1 + 2\|h_1\|_\infty) \|h_1\|_\infty \|h_2\|_\infty^2.
$$

Dividing by $\|h_2\|_\infty$ and letting $\|h_2\|_\infty \searrow 0$ shows second-order Fréchet differentiability. □

3.3 THREE-POINT INEQUALITIES

Next, we discuss how Assumption 2.5 can be satisfied. We assume that $\mathcal{X}_k$ is a finite dimensional space, as will be the case in the numerical realisation of the next section, so that the results of Section 3.2 are compatible with the results of Section 2. We write $w_x(U^{j,k})$ for the interior potential–electrode current tuples (3.3) corresponding to multiple electrode potentials $U^{j,k}$ and the conductivity x. Define $S : L^\infty(\Omega) \to R^{N_1 N_2}$, by

$$
(3.18)\qquad S(x) = (S_1(x), \dots, S_{N_1}(x)) \quad \text{for} \quad S_k(x) = \big(\Sigma^{-1/2} P w_x(U^{1,k}), \dots, \Sigma^{-1/2} P w_x(U^{N_2,k})\big),
$$

where P is a projection from $\mathcal{H}$ to $\mathbb{R}^{N_1}$ that extracts the electrode currents, i.e., $P w_x(U^{j,k}) = I^{j,k}$. Then Lemmas 3.3 and 3.5 show that S' and S'' are norm-bounded by

$$
(3.19)\qquad S'_{\max} := N_2 C_1 C_2 \max_{j,k} \|U^{j,k}\|_2 \|\Sigma^{-1/2}\|_2 \text{ and } S''_{\max} := 2 N_2 C_1^2 C_2^2 \max_{j,k} \|U^{j,k}\|_2^2 \|\Sigma^{-1/2}\|_2.
$$

Writing $\mathcal{I}^k$ for the measurement vector corresponding to (3.18)[3], E_k of (3.1) then reads

$$
(3.20)\qquad E_k(x) := \frac{1}{2} \|S_k(x) - \mathcal{I}^k\|^2.
$$

We define $\bar{B}_x w$ as the Riesz representation of $B_x(w, \cdot)$. Then, minding (3.6), $\bar{B}_x \in \mathbb{L}(\mathcal{H}; \mathcal{H})$ is invertible with eigenvalues bounded from below by C_1^{-1}. Also write $l_{x,h} \in \mathcal{H}$ for the Riesz representation of the right hand side of (3.8) as a functional of w: $\langle l_{x,h}, w \rangle = -\int_\Omega h \nabla_\xi v_x \cdot \nabla_\xi v \, \mathrm{d}\xi$ for any $w = (v, V) \in \mathcal{H}$. Note that here v_x is a component of the solution $w_x = (v_x, W_x) \in \mathcal{H}$. Then, by Lemma 3.3, $\bar{B}_x w'_{x,h} = l_{x,h}$.

[3]The entries of $\mathcal{I}^k$ are also multiples of $\Sigma^{-1/2}$.

By the invertibility of $\bar{B}_x$, we can write $w'_{x,h} = \bar{B}_x^{-1} l_{x,h}$. Also write $l_{x,h} = L_x h$, where $L \in \mathbb{L}(L^p(\Omega); \mathcal{H})$, with $1 < p < \infty$. It follows that

$$\|\Sigma^{-1/2} P w'_{x,h}\|^2 = \langle \Sigma^{-1} P^* \bar{B}_x^{-1} l_{x,h}, P\bar{B}_x^{-1} l_{x,h}\rangle = \langle L_x^* \bar{B}_x^{-*} P^* \Sigma^{-1} P \bar{B}_x^{-1} L_x h, h\rangle.$$

Writing $w'_{x,h}(U^{j,k})$ for the differential of $w_x(U^{j,k})$ in the direction h from x, and $L_x^{j,k}$ for L_x corresponding to $w_x(U^{j,k}) = (v_x(U^{j,k}), V_x(U^{j,k}))$, we thus get

$$\|S'_k(x,h)\|^2_{\mathbb{R}^{N_1 N_2}} = \sum_{j=1}^{N_2} \|\Sigma^{-1/2} P w'_{x,h}(U^{j,k})\|^2_{\mathbb{R}^{N_1}} = \langle A_k h, h\rangle$$

for

$$A_k := \sum_{j=1}^{N_2} (L_x^{j,k})^* \bar{B}_x^{-*} P^* \Sigma^{-1} P \bar{B}_x^{-1} L_x^{j,k}.$$

The next proof relies on a lower bound on A_k to model the idea that for a potential measurement setup $U^{j,k}$, $(j = 1, \dots, N_2)$, some electrode, indexed by $i = 1, \dots, N_1$, should react to a change h in the conductivity x. If x and h are discretised to a finite grid of n nodes, then it seems reasonable that this can be achieved as long as $N_1 N_2 > n$. If the latter does not hold, the condition could still be achieved at specific x or for specific directions h. Practically, for the condition to hold, the comparison point $\bar{x}^k$ (e.g., ground-truth) data fit $\|S_k(\bar{x}^k) - b_k\|$ has to be good enough, i.e., for the noise level to be low enough, and the radius $\delta >$ where we seek to satisfy Assumption 2.5, has to be small enough. This radius affects the closeness requirement of the initial iterate to $\bar{x}^0$ through Assumption 2.6 and (2.24).

Theorem 3.6. *Let Assumption 3.1 hold and $\mathcal{X}_k$ be finite dimensional. Define E_k by (3.20), and suppose $\bar{x}^k \in \mathcal{X}_k \cap B(\check{x}^k, \delta)$, $\hat{x}^k \in B(\check{x}^k, \delta)$, and $x \in B(\bar{x}^k, \delta)$ for a $\delta > 0$, and*

$$A_k \geq \max\{c_1, c_2\} \operatorname{Id} \tag{3.21}$$

for

$$c_1 := 4\theta + S''_{\max}(4\|S_k(\bar{x}^k) - b_k\|_{R^{N_1 N_2}} + (1 + \sqrt{2} + 2S''_{max})\delta^2) \quad \textit{and}$$
$$c_2 := 2\theta + 2S''_{\max}\left(\left(\frac{\delta}{8} + S'_{\max}\right)\delta + \|S_k(\bar{x}^k) - b_k\|_{R^{N_1 N_2}}\right).$$

Then Assumption 2.5 (i) and (ii) hold with $\gamma_{E,k} = 4\theta$, $\hat{\gamma}_{E,k} = 3\theta$,$\hat{e}_k = e_k = 0$, and

$$\lambda_{E,k} = 2\hat{\lambda}_{E,k} = S''_{\max}\left(S'_{\max}\delta + 3\|S_k(\bar{x}^k) - b_k\|_{R^{N_1 N_2}} + \frac{\delta^2 S''_{max}}{2}\right) + \left(2 + \sqrt{2}\right)(S'_{\max})^2.$$

Proof. Since S_k is twice differentiable, the bounds (3.19) on $\|S'\|$ and $\|S''\|$ guarantee Assumption A.1. By the preceding discussion, (3.21) guarantees

$$\|S'_k(\bar{x}^k)(x - \bar{x}^k)\|^2_{Z_k} \geq \max\{c_1, c_2\}\|x - \bar{x}^k\|^2_{X_k}$$

Now Corollary A.5 with $\varepsilon = 2\tilde{\varepsilon} = 2\theta$ and $\beta = 1 - 2^{-1/2}$ shows for $D = \frac{1}{2}\lambda_{E,k} = \hat{\lambda}_{E,k}$ that

$$\langle \nabla E_k(\check{x}^k), x - \bar{x}^k\rangle_{X_k} \geq E_k(x) - E_k(\bar{x}^k) + 2\theta\|x - \bar{x}^k\|^2_{X_k} - D\|x - \check{x}^k\|^2_{X_k}$$

and

$$\langle \nabla E(\check{x}^k) - \nabla E(\hat{x}^k), x - \hat{x}^k\rangle_{X_k} \geq 3\theta\|x - \bar{x}^k\|^2_{X_k} - D\|x - \check{x}^k\|^2_{X_k}. \qquad \square$$

Since we enforce $x \in [x_m, x_M]$, a.e., the previous theorem can also be used to prove the global Assumption 2.3 by taking δ large enough.

4 NUMERICAL EXPERIMENTS

We numerically assess Algorithm 1 in dynamic EIT imaging of a solid object moving in a fluid assumed to follow the incompressible constant-density transport equation. Our evaluation extends to scenarios that challenge the constant speed and incompressibility assumptions. We evaluate several dual predictors, and compare the results against static reconstructions with the Relaxed Inexact Gauss–Newton method (RIPGN) [23]. Our software implementation is available on Zenodo [22].

4.1 TEST SCENARIOS

The test scenarios take place in a disk-shaped domain denoted by Ω. We use $N_1 = 16$ evenly placed boundary electrodes. We set the electrode potentials $U^{j,k}$ so that an electrode j is set to a potential $U_j^{j,k} = 1\,\mathrm{V}$ while all others are grounded, $U_i^{j,k} = 0\,\mathrm{V}$ for $i \neq j$. This pattern repeats for all electrodes, leading to $N_2 = 16$ sets of electrode potentials. To mimic the typical EIT measurements, we exclude the currents at the excited electrode j from both measurements and the forward operator, as these currents are often not measured by EIT devices. Thus, each time instance k yields a total of $(N_1 - 1)N_2 = 240$ measurements.

The four experiments are as follows:

Baseline Features an inclusion moving at constant speed on a homogeneous background. Serves as a validation scenario to ensure the algorithm works as expected.

Circular Motion Features an inclusion following a circular motion path, challenging the constant movement assumption.

Halting Motion Features an inclusion that comes to a halt at frames 1000 and 2000, further challenging the constant movement assumption.

Disappearing Inclusions Features two inclusions moving in circular path. First inclusions disappears at frame 500 and the second at frame 1000. Both reappear at frame 1500. This case challenges the incompressibility assumption.

The *Baseline* experiment has 400 time frames, while the other three have 2000. The background conductivity is $x_{\mathrm{bg}} = 1\,\mathrm{S}$ and all inclusions are resistive with $x_{\mathrm{incl}} = 10^{-4}\,\mathrm{S}$.

We simulate the measurement data by approximating (3.2) with the Galerkin finite element method (FEM). We use piecewise linear basis. The simulation mesh has 5039 nodes and 9852 elements. Each simulated measurement $\mathcal{I}_i^{j,k}$ has added Gaussian noise with standard deviation $\mathrm{std} = 10^{-4}|\mathcal{I}_i^{j,k}|$, in the standard range of EIT. For the specifics of how to solve (3.2) and its Fréchet derivative with FEM, see [23].

4.2 NUMERICAL SETUP

Recall the definition of E_k from (3.1). The currents $S_k : X^k \times \mathbb{R}^{N_1} \to \mathbb{R}^{N_1 N_2}$ in E_k are obtained by approximating the potential functions $u^{j,k}$ in (3.2) through FEM, using piecewise linear basis functions. The conductivity x is also represented in the same basis. To prevent the 'inverse crime' [29], we use a less dense mesh for the forward problem than we used for simulation, featuring 2917 nodes and 5430 mesh elements.

4.2.1 *BACKGROUND PROCESSING*

Note that the computation of $S_k(x^k)$ and $\nabla S_k(x^k)$ is highly resource-intensive. To optimise the computational speed of Algorithm 1, we implement the following background processing strategy: we

approximate $\nabla S_k(x^k)$ by $\nabla S_k(\check{x})$, initially $\check{x} = x^0$. We then approximate $S_k(x^k)$ by a first-order Taylor expansion around $\check{x}$. In the background, we compute $S_k(\tilde{x})$ and $\nabla S_k(\tilde{x})$, at a linearisation point $\tilde{x} = x^j$. After completing these background computations, we update $\check{x} = \tilde{x}$ as well as $\tilde{x} = x^k$ for the current iterate, and start computing new values of the operators in the background. In Appendix B.1 we show that the background processing scheme satisfies Assumption 2.5.

4.2.2 *PREDICTORS*

We construct the primal component of the predictor P_k by assuming constant velocity and incompressibility in the moving objects. Denoting by $h^k(\xi) \in \mathbb{R}^2$ the displacement at $\xi \in \Omega$ between frames k and $k-1$, we define the primal prediction referred to as Flow, as $\check{x}^{k+1}(\xi) = W_k x^k(\xi) := x^k(\xi + h^k(\xi))$. Since h^k represents the estimated displacement between the current and previous frames and is used to predict the next frame, this effectively assumes a constant velocity between frames.

We estimate displacement h^k from the incompressible transport equation

$$\frac{\partial x^k}{\partial t} - \nabla_\xi x^k \cdot v = 0. \tag{4.1}$$

Here, $v = (v_1, v_2) \in \mathbb{R}^2$ represents the velocity responsible for the time-dependent displacement h at time t. Using a fixed time step of $\Delta t = 1$, we approximate $\frac{\partial x^k}{\partial t}$ as $(x^k - x^{k-1})/\Delta t = x^k - x^{k-1}$. This leads us to a problem akin to optical flow, where the solution for v is subject to non-uniqueness due to the aperture problem [4]. Similar to the well-established Horn-Schunck method [20], we approximate the solution of (4.1) by

$$v^k = \arg\min_v \frac{1}{2}\left\|\frac{\partial x^k}{\partial t} - \nabla_\xi x^{k-1} \cdot v\right\|^2 + \frac{\beta_1}{2}\left(\|\nabla_\xi v_1\|^2 + \|\nabla_\xi v_2\|^2\right) + \frac{\beta_2}{2}\|v\|^2. \tag{4.2}$$

Here, ∇_ξ denotes the spatial gradient operator, and β_1 and β_2 are positive regularisation parameters. The components v_1^k and v_2^k share the same piecewise linear basis as the conductivity x^k, and the displacement h^k corresponds to v^k due to the chosen time step. To expedite the algorithm, we only update h^k every fourth iteration.

For the dual variable, we consider two distinct predictors: *Greedy* preserves $\langle \nabla_\xi \check{x}^{k+1}, \check{y}^{k+1}\rangle = \langle \nabla_\xi x^k, y^k\rangle$ element-wise, while *Affine* sets $\check{y}^{k+1} := y^k + c\nabla_\xi \check{x}^{k+1}$. Through the choice of $c > 0$, it seeks to promote sparsity in regions where we expect constant values, as discussed in Example 2.2. To bound the prediction component of the error $e_N^\Sigma(u^{0:N-1}, \bar{u}^{0:N})$ in Corollary 2.13 for these predictors, see [14, Lemma 3.11 and Lemma 3.13]. The gradient estimate error we treat in Appendix B.1. We also perform experiments with the *Identity* predictor $W_k = \mathrm{Id}$ and $T_k = \mathrm{Id}$.

We evaluate the algorithm across four predictor configurations. *No Prediction* is fully uninformed having identity predictors for both primal and dual variables. *Primal Only* uses incompressible flow prediction for the primal combined with uninformed identity prediction for the dual. The remaining two configurations feature the incompressible flow prediction for the primal, accompanied by either *Greedy* or *Affine* dual prediction.

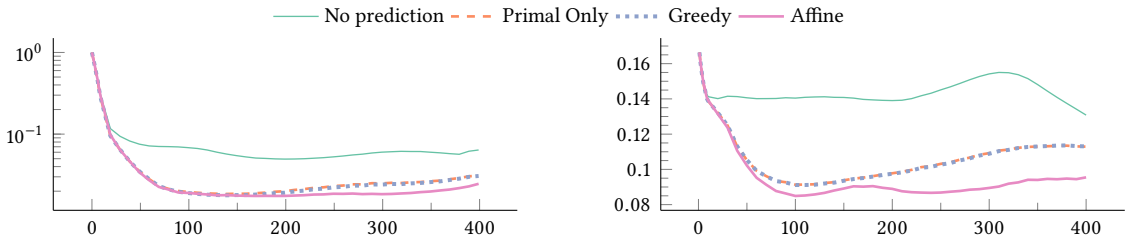
Note that only the Baseline experiment fully satisfies the constant velocity assumption. Additionally, the fourth test scenario violates the incompressibility assumption. However, we expect the algorithm to perform correctly, as the optimisation steps should compensate for small errors introduced in the prediction step. Furthermore, v is only an approximate solution to (4.1), meaning some degree of compression may still be present.

4.2.3 *ALGORITHM PARAMETERS*

In all experiments, we use the following parameters:

Table 1: Average relative errors and confidence intervals (CI). The latter are calculated using Student's t-distribution with the relative errors for each iteration from 1 to 500 as samples.

Baseline				**Circular motion**			
	Average RE				Average RE		
Predictor	iter 1	iter 50	95% CI	Predictor	iter 1	iter 50	95% CI
Affine	9.56	9.0725	9.0311 - 9.1139	Affine	9.0218	8.8642	8.8424 - 8.886
Greedy	10.7881	10.4691	10.3895 - 10.5487	Greedy	10.5676	10.4494	10.4213 - 10.4775
No predict	14.4899	14.4774	14.4204 - 14.5344	No predict	14.4643	14.4275	14.3768 - 14.4782
Primal only	10.844	10.53	10.4523 - 10.6077	Primal only	10.632	10.5153	10.4868 - 10.5438
Halting motion				**Halting motion**			
	Average RE				Average RE		
Predictor	iter 1	iter 50	95% CI	Predictor	iter 1	iter 50	95% CI
Affine	9.1663	8.9985	8.9651 - 9.0319	Affine	11.1383	10.8471	10.6356 - 11.0586
Greedy	10.3048	10.1686	10.1416 - 10.1956	Greedy	11.5483	11.2712	11.0304 - 11.512
No predict	14.7439	14.7116	14.6822 - 14.741	No predict	15.1044	14.8975	14.5571 - 15.2379
Primal only	10.2875	10.1508	10.124 - 10.1776	Primal only	11.7061	11.4329	11.1949 - 11.6709



(a) Relative objective value

(b) Relative error to ground-truth

Figure 1: Iteration-wise relative objective values and iterate errors in *Baseline* experiment.

Main problem parameters We set $\Sigma^{-1/2} = 200\,\mathrm{Id}$, $\alpha = 0.5$, $x_m = 10^{-5}$, and $x_M = 10^5$. We do not use the knowledge of the precise noise statistics or the conductivity range.

Step length parameters We take constant $\tau_k \equiv \tau := 0.85\|\Sigma^{-1/2}\nabla I(\check{x})(\Sigma^{-1/2}\nabla I(\check{x}))^*\|^{-1}$ and $\sigma_k \equiv \sigma := 1$.

Incompressible flow parameters We set $\beta_1 = 10^{-3}$ and $\beta_2 = 10^{-5}$.

Affine predictor parameters To promote gradient sparsity in calm areas, we take $c := 10\max_\xi\{0, 1 - 10^{-12}|h^k|^{-1}\}^2$ for h^k the estimated displacement.

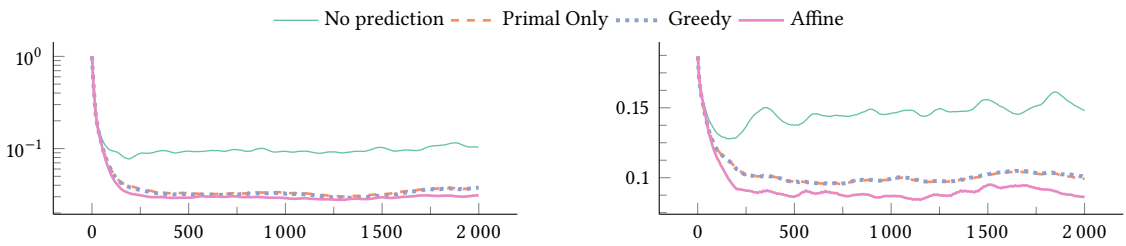
The step length parameter choices are discussed in detail in Appendix B.2.

4.3 RESULTS

The *Baseline* experiment features an inclusion moving at constant speed. Figure 1 shows the relative objective values and iterate errors

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

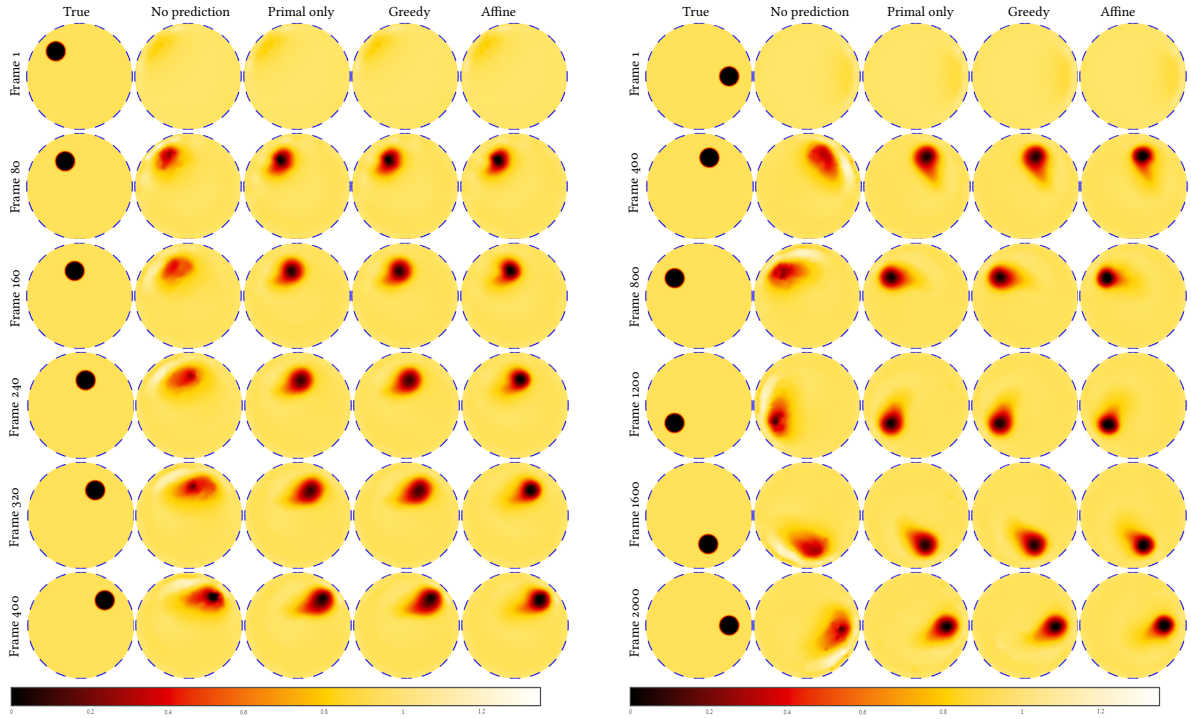
where x^k_{true} is the ground-truth for iteration k with the tested predictor configurations. *No Prediction* performs the worst and *Affine* predictor the best. *Primal Only* and *Greedy* perform similarly, outperforming *No Prediction* but falling behind *Affine*. This is also confirmed by Table 1.



(a) Relative objective value

(b) Relative error to ground-truth

Figure 2: Iteration-wise objective value and relative error in *Circular Motion* experiment.



(a) *Baseline* experiment

(b) *Circular Motion* experiment

Figure 3: A selection of reconstructed frames in *Baseline* and *Circular Motion* experiments.

Reconstructions in Figure 3a confirm what we observed in Figure 1 and Table 1. The incompressible Affine predictor produces slightly sharper reconstructions than other Flow predictors, with No prediction yielding the blurriest results.

In the *Circular Motion* experiment, an inclusion moves in a circular trajectory. By Figures 2 and 3b and Table 1, *Affine* again provides the sharpest reconstructions and lowest objective values, while *No Prediction* performs the worst.

Figure 4 compares the true target, a static solution using RIPGN, and online reconstruction with *Affine* at frame 500. The static reconstruction is sharper but significantly slower, taking 71.9 seconds, whereas online reconstruction with Algorithm 1 captures essential features in just 12.1 milliseconds.

The third experiment involves a halting inclusion. By Figures 5 and 7a and Table 1, the reconstruction quality, the relative objective values and iterate errors align with previous experiments. Notably, as objects slow down, especially with *No Prediction*, reconstructions improve. This aligns with expectations, as, aside from the noise, the data is static, where identity prediction is optimal.

The final experiment features two inclusions vanishing at different frames (500 and 1000) and

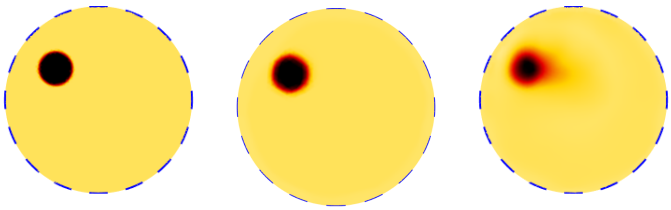
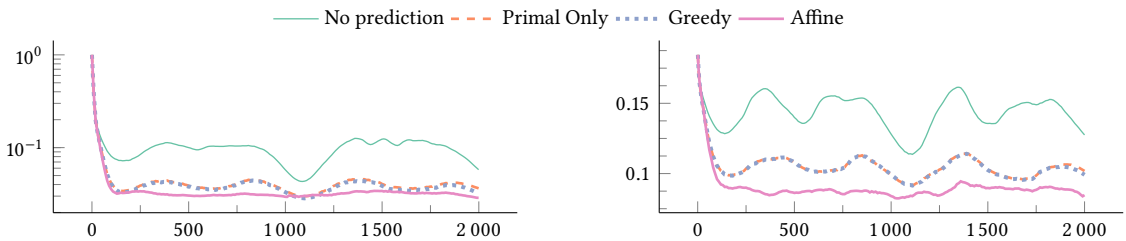


Figure 4: *Circular Motion* experiment. Comparison of the reconstruction quality. Left: true target. Middle: static reconstruction with RIPGN. Right: online reconstruction with Algorithm 1 and Affine prediction.



(a) Relative objective value

(b) Relative error to ground-truth

Figure 5: Iteration-wise objective value and relative error in *Halting Motion* experiment.

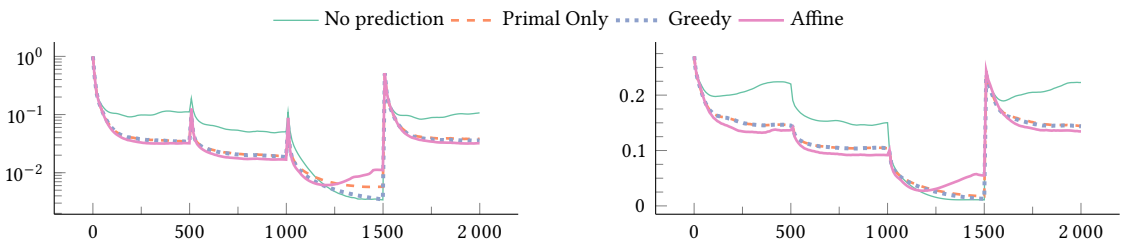
reappearing at frame 1500. Table 1 and Figure 6 show similar results to the previous cases, although by average relative error, the differences between *Affine* and *No Prediction* and *Primal Only* are smaller. However, Figure 6 reveals abrupt spikes in the objective value and in the relative error when objects disappear, followed by subsequent decreases as reconstructions exhibit fewer edges, resulting in lower total variation penalties. *No Prediction* dominates when both inclusions disappear since, aside from the noise, the data is completely static. By Figure 7b, the reconstructions accurately capture the process of inclusions disappearing and reappearing.

5 CONCLUSIONS

Online optimisation offers a real-time option for solving sequential optimisation problems. While its application in dynamic inverse problems remains relatively rare, we introduced a predictive online primal-dual proximal splitting method tailored for objective functions with non-smooth and non-convex components.

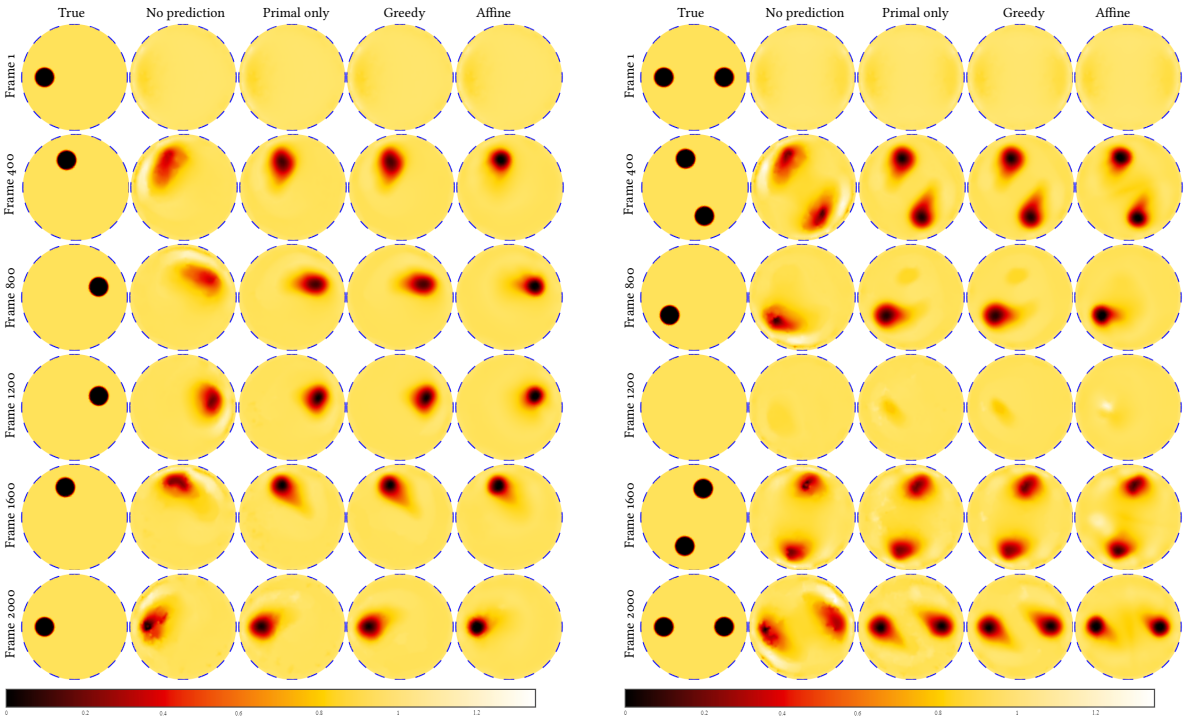
We established a regret bound for this method and we comprehensively evaluated the method in dynamic EIT. Through numerical evaluations using incompressible constant flow-based predictors, we have demonstrated a substantial enhancement in reconstruction quality when compared to uninformed predictions, even in cases where the assumptions of constancy and incompressibility were violated. Remarkably, this improvement was achieved while maintaining minimal computational times, averaging just around 12 milliseconds.

It is worth noting that, in our experiments, the online reconstruction quality, albeit slightly inferior, remained competitive with significantly slower and computationally more costly static reconstructions. We anticipate further enhancements in the computational speed of our algorithm through optimised numerical implementations, promising even more efficient and effective solutions in the future.



(a) Relative objective value (b) Relative error to ground-truth

Figure 6: Iteration-wise objective value and relative error for *Disappearing Inclusions*.



(a) *Halting Motion* experiment (b) *Disappearing Inclusions* experiment

Figure 7: A selection of reconstructed frames in *Halting Motion* and *Disappearing Inclusions* experiments.

REFERENCES

[1] D. J. Albers, P. A. Blancquart, M. E. Levine, E. E. Seylabi, and A. Stuart, Ensemble Kalman methods with constraints, *Inverse Problems* 35 (2019), 095007, doi:10.1088/1361-6420/ab1c09.

[2] 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.

[3] N. Bastianello, A. Simonetto, and R. Carli, Primal and dual prediction-correction methods for time-varying convex optimization (2020), arXiv:2004.11709.

[4] S. S. Beauchemin and J. L. Barron, The computation of optical flow, *ACM computing surveys (CSUR)* 27 (1995), 433–466.

[5] E. V. Belmega, P. Mertikopoulos, R. Negrel, and L. Sanguinetti, Online convex optimization and no-regret learning: Algorithms, guarantees and applications, 2018, arXiv:804.04529.

[6] M. Benning, L. Gladden, D. Holland, C. B. Schönlieb, and T. Valkonen, Phase reconstruction from velocity-encoded MRI measurements – A survey of sparsity-promoting variational approaches, *Journal of Magnetic Resonance* 238 (2014), 26–43, doi:10.1016/j.jmr.2013.10.003.

[7] A. Bernstein, E. Dall'Anese, and A. Simonetto, Online primal-dual methods with measurement feedback for time-varying convex optimization, *IEEE Transactions on Signal Processing* 67 (2019), 1978–1991.

[8] T. J. Chang and S. Shahrampour, On Online Optimization: Dynamic Regret Analysis of Strongly Convex and Smooth Problems, *Proceedings of the AAAI Conference on Artificial Intelligence* 35 (2021), 6966–6973, doi:10.1609/aaai.v35i8.16858.

[9] X. Chen, B. Tang, J. Fan, and X. Guo, Online gradient descent algorithms for functional data learning, *Journal of Complexity* 70 (2022), 101635.

[10] K. S. Cheng, D. Isaacson, J. Newell, and D. G. Gisser, Electrode models for electric current computed tomography, *IEEE Transactions on Biomedical Engineering* 36 (1989), 918–924.

[11] C. Clason, S. Mazurenko, and T. Valkonen, Acceleration and global convergence of a first-order primal-dual method for nonconvex problems, *SIAM Journal on Optimization* 29 (2019), 933–963, doi:10.1137/18m1170194, arXiv:1802.03347.

[12] C. Clason and T. Valkonen, Introduction to Nonsmooth Analysis and Optimization, 2024, arXiv:2001.00216. Submitted.

[13] J. Dardé, N. Hyvönen, T. Kuutela, and T. Valkonen, Contact Adapting Electrode Model for Electrical Impedance Tomography, *SIAM Journal on Applied Mathematics* 82 (2022), 427–449.

[14] N. Dizon, J. Jauhiainen, and T. Valkonen, Prediction techniques for dynamic imaging with online primal-dual methods, *Journal of Mathematical Imaging and Vision* (2024), doi:10.1007/s10851-024-01214-w, arXiv:2405.02497.

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

[16] M. Gerber and K. Heine, Online inference with multi-modal likelihood functions, *The Annals of Statistics* 49 (2021), doi:10.1214/21-aos2076.

[17] 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.

[18] E. Hazan et al., Introduction to online convex optimization, *Foundations and Trends in Optimization* 2 (2016), 157–325.

[19] 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.

[20] B. K. Horn and B. G. Schunck, Determining optical flow, *Artificial intelligence* 17 (1981), 185–203.

[21] 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.

[22] J. Jauhiainen, N. Dizon, and T. Valkonen, Codes for "Online optimization for dynamic electrical impedance tomography", 2025, doi:10.5281/zenodo.15028468. Software.

[23] 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.

[24] 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* (2020).

[25] 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.

[26] J. Jauhiainen, A. Seppanen, and T. Valkonen, Mumford-Shah regularization in electrical impedance tomography with complete electrode model, *Inverse Problems* (2022).

[27] 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.

[28] B. Jin and P. Maass, *An analysis of electrical impedance tomography with applications to Tikhonov regularization*, volume 18, 2012, doi:10.1051/cocv/2011193.

[29] 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.

[30] S. Lan, S. Li, and M. Pasha, Bayesian spatiotemporal modeling for inverse problems, *Statistics and Computing* 33 (2023), doi:10.1007/s11222-023-10253-z.

[31] A. Lechleiter, Dynamic inverse problems: modelling—regularization—numerics, *Inverse Problems* 34 (2018), 4pp.

[32] 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.

[33] A. I. Nachman, Global uniqueness for a two-dimensional inverse boundary value problem, *Annals of Mathematics* (1996), 71–96.

[34] M. Nonhoff and M. A. Müller, Online gradient descent for linear dynamical systems, *IFAC-PapersOnLine* 53 (2020), 945–952.

[35] F. Orabona, A Modern Introduction to Online Learning, 2020, arXiv:1912.13213.

[36] S. Siltanen, J. Mueller, and D. Isaacson, An implementation of the reconstruction algorithm of A Nachman for the 2D inverse conductivity problem, *Inverse Problems* 16 (2000), 681.

[37] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, Time-varying convex optimization: Time-structured algorithms and applications, *Proceedings of the IEEE* 108 (2020), 2032–2048.

[38] E. Somersalo, M. Cheney, and D. Isaacson, Existence and uniqueness for electrode models for electric current computed tomography, *SIAM Journal on Applied Mathematics* 52 (1992), 1023–1040.

[39] Y. Tang, E. Dall'Anese, A. Bernstein, and S. Low, Running primal-dual gradient method for time-varying nonconvex problems, *SIAM Journal on Control and Optimization* 60 (2022), 1970–1990.

[40] T. Valkonen, A primal-dual hybrid gradient method for non-linear operators with applications to MRI, *Inverse Problems* 30 (2014), 055012, doi:10.1088/0266-5611/30/5/055012, arXiv:1309.5032.

[41] T. Valkonen, Testing and non-linear preconditioning of the proximal point method, *Applied Mathematics and Optimization* 82 (2020), doi:10.1007/s00245-018-9541-6, arXiv:1703.05705.

[42] T. Valkonen, Predictive online optimisation with applications to optical flow, *Journal of Mathematical Imaging and Vision* 63 (2021), 329–355, doi:10.1007/s10851-020-01000-4, arXiv:2002.03053.

[43] 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), 1–10.

[44] 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.

[45] L. Zhang, H. Liu, and X. Xiao, Regrets of proximal method of multipliers for online non-convex optimization with long term constraints, *Journal of Global Optimization* 85 (2022), 61–80, doi:10.1007/s10898-022-01196-2.

[46] Y. Zhang, E. Dall'Anese, and M. Hong, Online proximal-admm for time-varying constrained convex optimization, *IEEE Transactions on Signal and Information Processing over Networks* 7 (2021), 144–155.

[47] Y. Zhang, R. J. Ravier, V. Tarokh, and M. M. Zavlanos, Distributed Online Convex Optimization with Improved Dynamic Regret, 2019, arXiv:1911.05127.

[48] M. Zinkevich, Online convex programming and generalized infinitesimal gradient ascent, in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, AAAI, 2003, 928–936.

APPENDIX A THE DATA FITTING TERM AND ITS PROPERTIES

In this appendix, we prove Corollary A.5 on smoothness inequalities for

$$E_k(x) := \frac{1}{2}\|S_k(x) - b_k\|_{Z_k}^2,$$

required in the proof of Theorem 3.6. We assume:

Assumption A.1. Let $S_k : X_k \mapsto Z_k$ be twice Fréchet differentiable between the Hilbert spaces X_k and Z_k, and suppose $S_k'(x) \in \mathbb{L}(X_k; Z_k)$ and $S_k''(x) \in \mathbb{L}(X_k \times X_k; Z_k)$ satisfy

$$\begin{aligned} \|S_k'(x)(y-x)\|_{Z_k} &\le S'_{\max}\|y-x\|_{X_k} \quad \text{and} \\ \|S_k''(x)(y-x, z-x)\|_{Z_k} &\le S''_{\max}\|y-x\|_{X_k}\|z-x\|_{X_k} \quad \text{for all} \quad x, y, z \in B(\bar{x}, \delta), \end{aligned}$$

For brevity, in the proofs, we drop the time index k as it bears no impact on our analysis. We also abbreviate $S_z'(h) := S'(z)(h)$ and $R(x) = S_k(x) - b_k$. Then $E(x) = \frac{1}{2}\|R(x)\|^2$, keeping in mind that $S_k' = R'$ and $S_k'' = R''$. We start with an auxiliary lemma.

Lemma A.2. *Suppose that Assumption A.1 holds for some $\bar{x} \in X_k$ and $\delta > 0$, and that $x, z \in B(\bar{x}, \delta)$. Then, $c = z + t(x - z)$ and some $t \in [0, 1]$,*

$$\langle \nabla E_k(z), z - x \rangle_{X_k} = E_k(z) - E_k(x) + \frac{1}{2}\|S_k'(z)(x - z)\|_{Z_k}^2 - \frac{1}{8}\|S_k''(c)(x - z, x - z)\|_{Z_k}^2 + \frac{1}{2}\langle S_k''(c)(x - z, x - z), S_k(x) - b_k \rangle_{Z_k}.$$

Proof. Since S_k and therefore R is twice differentiable, Taylor expansion gives for $c := x + t(z - x)$ and some $t \in [0, 1]$ that

$$R(x) = R(z) + R_z'(x - z) + \frac{1}{2}R_c''(x - z, x - z). \tag{A.1}$$

Expanding $E(x) - E(z)$ and using (A.1) and $\langle \nabla E(z), x - z \rangle = \langle R_z'(x - z), R(z) \rangle$ then yields

$$\begin{aligned} E(x) - E(z) &= \frac{1}{2}\left(\|R(x)\|^2 - \|R(z)\|^2\right) = \frac{1}{2}\langle R(x) - R(z), R(x) + R(z) \rangle \\ &= \frac{1}{2}\langle R(z) + R_z'(x - z) + \frac{1}{2}R_c''(x - z, x - z) - R(z), R(x) + R(z) \rangle = \frac{1}{2}\langle \nabla E(z), x - z \rangle + A, \end{aligned} \tag{A.2}$$

where an application of (A.1) establishes

$$\begin{aligned} A &:= \frac{1}{2}\left(\left\langle R_z'(x - z) + \frac{1}{2}R_c''(x - z, x - z), R(x) \right\rangle + \frac{1}{2}\left\langle R_c''(x - z, x - z), R(z) \right\rangle\right) \\ &= \frac{1}{2}\left\langle R_z'(x - z) + \frac{1}{2}R_c''(x - z, x - z), R(z) + R_z'(x - z) + \frac{1}{2}R_c''(x - z, x - z) \right\rangle \\ &\quad + \frac{1}{4}\langle R_c''(x - z, x - z), R(z) \rangle = \frac{1}{2}\langle \nabla E(z), x - z \rangle + \frac{1}{2}\|R_z'(x - z)\|^2 + B. \end{aligned} \tag{A.3}$$

Here again, an application of (A.1), gives

$$\begin{aligned} B &:= \frac{1}{8}\|R_c''(x - z, x - z)\|^2 + \frac{1}{2}\langle R_c''(x - z, x - z), R(z) + R_z'(x - z) \rangle \\ &= -\frac{1}{8}\|R_c''(x - z, x - z)\|^2 + \frac{1}{2}\langle R_c''(x - z, x - z), R(x) \rangle. \end{aligned} \tag{A.4}$$

By combining (A.2)–(A.4), we obtain the claim. □

With the help of Lemma A.2, we next derive a lower bound for $\langle \nabla E_k(z), x - \bar{x} \rangle_{X_k}$.

Lemma A.3. *Suppose that Assumption A.1 holds for some $\bar{x} \in X_k$ and $\delta > 0$. Then*

$$\langle \nabla E_k(z), x - \bar{x} \rangle_{X_k} \geq E_k(x) - E_k(\bar{x}) + \frac{1 - \beta}{2}\|S_k'(z)(x - \bar{x})\|_{Z_k}^2 - \frac{1}{2\beta}\|S_k'(z)(x - z)\|_{Z_k}^2 - \frac{A}{2}\|x - \bar{x}\|_{X_k}^2 - \frac{B}{2}\|x - z\|_{X_k}^2$$

for all $x, z \in B(\bar{x}, \delta)$, where for any $\beta > 0$,

$$\begin{aligned} A &:= S_{\max}''\left(2\|S_k(\bar{x}) - b_k\|_{Z_k} + \delta^2 S_{\max}''/2\right) \quad \textit{and} \\ B &:= S_{\max}''\left(S_{\max}'\delta + 3\|S_k(\bar{x}) - b_k\|_{Z_k} + \delta^2 S_{\max}''/2\right). \end{aligned} \tag{A.5}$$

Proof. Since $x, z, \bar{x} \in B(\bar{x}, \delta)$, applying Lemma A.2 twice establishes for some $t_1, t_2 \in [0, 1]$ and $b := z + t_1(x - z)$ and $c := z + t_2(\bar{x} - z)$ the identity

$$\begin{aligned} \langle \nabla E(z), x - \bar{x} \rangle &= \langle \nabla E(z), x - z \rangle + \langle \nabla E(z), z - \bar{x} \rangle \\ &= E(x) - E(\bar{x}) + \frac{1}{2}\|R_z'(\bar{x} - z)\|^2 - \frac{1}{8}\|R_c''(\bar{x} - z, \bar{x} - z)\|^2 + \frac{1}{2}\langle R_c''(\bar{x} - z, \bar{x} - z), R(\bar{x}) \rangle \\ &\quad - \frac{1}{2}\|R_z'(x - z)\|^2 + \frac{1}{8}\|R_b''(x - z, x - z)\|^2 - \frac{1}{2}\langle R_b''(x - z, x - z), R(x) \rangle. \end{aligned} \tag{A.6}$$

Next, we estimate the $\bar{x} - z$ terms with similar $\bar{x} - x$ and $x - z$ terms. First, let us inspect the term $\|R_z'(\bar{x} - z)\|^2$. By Young's inequality and the linearity of R',

$$\frac{1}{2}\|R_z'(\bar{x} - z)\|^2 = \frac{1}{2}\|R_z'(\bar{x} - x) - R_z'(z - x)\|^2 \geq \frac{1-\beta}{2}\|R_z'(\bar{x} - x)\|^2 + \frac{1-\beta^{-1}}{2}\|R_z'(z - x)\|^2. \tag{A.7}$$

Next, we inspect the term $\|R_c''(\bar{x}-z, \bar{x}-z)\|^2$. R is bilinear and bounded by $R''_{\max} = S''_{\max}$ by Assumption A.1, thus Young's inequality, and $z \in B(\bar{x}, \delta)$ gives

$$\begin{aligned} \|R_c''(\bar{x} - z, \bar{x} - z)\|^2 &\leq 2\|R_c''(\bar{x} - x, \bar{x} - z)\|^2 + 2\|R_c''(x - z, \bar{x} - z)\|^2 \\ &\leq 2(R''_{\max})^2\|\bar{x} - x\|^2\delta^2 + 2(R''_{\max})^2\|x - z\|^2\delta^2. \end{aligned} \tag{A.8}$$

Finally, let us inspect the term $\langle R_c''(\bar{x} - z, \bar{x} - z), R(\bar{x})\rangle - \langle R_b''(x - z, x - z), R(x)\rangle$. Notice that by mean value theorem $R(x) = R(\bar{x}) + R_a'(x - \bar{x})$ for $a = \bar{x} + t_3(x - \bar{x})$ with some $t_3 \in [0, 1]$. Using this, bilinearity and symmetricity of R'', boundedness of R' and R'', Cauchy-Schwartz and Young's inequalities, and $x \in B(\bar{x}, \delta)$ we obtain

$$\begin{aligned} &\langle R_c''(\bar{x} - z, \bar{x} - z), R(\bar{x})\rangle - \langle R_b''(x - z, x - z), R(x)\rangle \\ &= \langle R_c''(\bar{x} - x, \bar{x} - x), R(\bar{x})\rangle - \langle R_b''(x - z, x - z), R_a'(x - \bar{x})\rangle \\ &\quad + \langle R_c''(x - z, x - z) - R_b''(x - z, x - z), R(\bar{x})\rangle + 2\langle R_c''(\bar{x} - x, x - z), R(\bar{x})\rangle \\ &\geq -S''_{\max}\|R(\bar{x})\|\|\bar{x} - x\|^2 - S''_{\max}S'_{\max}\delta\|x - z\|^2 - 2S''_{\max}\|R(\bar{x})\|\|x - z\|^2 \\ &\quad - 2S''_{\max}\|R(\bar{x})\|\|\bar{x} - x\|\|x - z\| \\ &\geq -S''_{\max}\left(2\|R(\bar{x})\|\|\bar{x} - x\|^2 + \left(3\|R(\bar{x})\| + S'_{\max}\delta\right)\|x - z\|^2\right). \end{aligned} \tag{A.9}$$

Finally, applying the above estimates (A.7)–(A.9) to (A.6) shows that

$$\begin{aligned} \langle \nabla E(z), x - \bar{x}\rangle \geq E(x) - E(\bar{x}) + \frac{1-\beta}{2}\|R_z'(\bar{x} - x)\|^2 - \frac{1}{2\beta}\|R_z'(z - x)\|^2 - S''_{\max}\|R(\bar{x})\|\|\bar{x} - x\|^2 \\ - \frac{S''_{\max}(S'_{\max}\delta + 3\|R(\bar{x})\|)}{2}\|x - z\|^2 - \frac{\delta^2(R''_{\max})^2}{4}\left(\|\bar{x} - x\|^2 + \|x - z\|^2\right). \end{aligned} \qquad \square$$

The next corollary transfers $S_k'(z)$ to $S_k'(\bar{x})$ in one of the terms.

Corollary A.4. *Suppose Assumption A.1 holds for some $\bar{x} \in X_k$ and $\delta > 0$. Then*

$$\begin{aligned} \langle \nabla E_k(z), x - \bar{x}\rangle_{X_k} \geq E_k(x) - E_k(\bar{x}) + \frac{(1-\beta)^2}{2}\|S_k'(\bar{x})(x - \bar{x})\|_{Z_k}^2 - \frac{1}{2\beta}\|S_k'(z)(x - z)\|_{Z_k}^2 \\ - \frac{C}{2}\|x - \bar{x}\|_{X_k}^2 - \frac{B}{2}\|x - z\|_{X_k}^2, \end{aligned}$$

for all $x, z \in B(\bar{x}, \delta)$, where, for any $1 > \beta > 0$, B is given in (A.5), and

$$C := A + S''_{\max}\beta^{-1}(\beta - 1)^2\delta^2 = S''_{\max}\left(2\|S_k(\bar{x}) - b_k\|_{Z_k} + \delta^2 S''_{\max}/2 + \beta^{-1}(\beta - 1)^2\delta^2\right). \tag{A.10}$$

Proof. With fixed h, $R_z'(h)$ is continuously differentiable by Assumption A.1. Thus, by the main value theorem, for $a = \bar{x} + t_1(x - \bar{x})$ with some $t_1 \in [0, 1]$, we have

$$\begin{aligned} R_z'(\bar{x} - x) &= R'(z)(\bar{x} - x) = R'(\bar{x})(\bar{x} - x) + R''(a)(\bar{x} - x, z - \bar{x}) \\ &= R_{\bar{x}}'(\bar{x} - x) + R_a''(\bar{x} - x, z - \bar{x}). \end{aligned} \tag{A.11}$$

Since $\beta > 0$, (A.11), $z \in B(\bar{x}, \delta)$, Young's inequality and the boundedness of R'' (due to boundedness of S'' in Assumption A.1) show that

$$\begin{aligned} \|R_z(\bar{x} - x)\|^2 &= \|R_{\bar{x}}'(\bar{x} - x)\|^2 + \|R_a''(\bar{x} - x, z - \bar{x})\|^2 + 2\langle R_{\bar{x}}'(z - \bar{x}), R_a''(\bar{x} - x, z - \bar{x})\rangle \\ &\geq (1-\beta)\|R_{\bar{x}}'(\bar{x} - x)\|^2 - (\beta^{-1} - 1)\|R_a''(\bar{x} - x, z - \bar{x})\|^2 \\ &\geq (1-\beta)\|R_{\bar{x}}'(\bar{x} - x)\|^2 - (\beta^{-1} - 1)S''_{\max}\delta^2\|\bar{x} - x\|^2. \end{aligned}$$

Since $(1 - \beta)(\beta^{-1} - 1) = (\beta - 1)^2/\beta$, using this and Lemma A.3 yields the claim. $\square$

Finally, we state conditions that guarantee the growth estimates of Assumptions 2.3 and 2.5 for E. Essentially, $S'(\bar{x})^* S'(\bar{x})$ has to be sufficiently elliptic.

Corollary A.5. *Suppose that Assumption A.1 holds for some $\bar{x} \in X_k$ and $\delta > 0$. Let $x, z \in B(\bar{x}, \delta)$, and suppose some $1 > \beta > 0$ and some $\varepsilon \in \mathbb{R}$ that*

$$\|S_k'(\bar{x})(x - \bar{x})\|_{Z_k}^2 \geq \frac{C + 2\varepsilon}{(1-\beta)^2} \|x - \bar{x}\|_{X_k}^2 \tag{A.12}$$

for C defined in (A.10). Then, for B defined in (A.5) and $D := \frac{1}{2}B + \frac{1}{2\beta}(S'_{\max})^2$,

$$\langle \nabla E_k(z), x - \bar{x} \rangle_{X_k} \geq E_k(x) - E_k(\bar{x}) + \varepsilon \|x - \bar{x}\|_{X_k}^2 - D\|x - z\|_{X_k}^2 \tag{A.13}$$

Additionally, if for some $\tilde{\varepsilon} \in \mathbb{R}$ also

$$\|S_k'(\bar{x})(x - \bar{x})\|_{Z_k}^2 \geq 2\left(\tilde{\varepsilon} + S''_{\max}\left(\left(\frac{\delta}{8} + S'_{\max}\right)\delta + \|S_k(\bar{x}) - b_k\|_{Z_k}\right)\right)\|x - \bar{x}\|_{X_k}^2, \tag{A.14}$$

then

$$\langle \nabla E(z) - \nabla E(\bar{x}), x - \bar{x} \rangle_{X_k} \geq (\varepsilon + \tilde{\varepsilon})\|x - \bar{x}\|_{X_k}^2 - D\|x - z\|_{X_k}^2. \tag{A.15}$$

Proof. The boundedness of R', Corollary A.4, and (A.12) yield

$$\begin{aligned} E(\bar{x}) - E(x) + \langle \nabla E(z), x - \bar{x} \rangle &\geq \frac{(1-\beta)^2}{2}\|R_{\bar{x}}(x - \bar{x})\|^2 - \frac{1}{2\beta}\|R'(z)(x - z)\|^2 \\ &\quad - \frac{C}{2}\|x - \bar{x}\|^2 - \frac{B}{2}\|x - z\|^2 \\ &\geq \varepsilon\|x - \bar{x}\|^2 - \frac{1}{2\beta}(S'_{\max})^2\|x - z\|^2 - \left(D - \frac{1}{2\beta}(S'_{\max})^2\right)\|x - z\|^2 \\ &= \varepsilon\|x - \bar{x}\|^2 - D\|x - z\|^2. \end{aligned}$$

This gives (A.13). Subtracting $\langle \nabla E(\bar{x}), x - \bar{x} \rangle$ from both sides, we further obtain

$$\langle \nabla E(z) - \nabla E(\bar{x}), x - \bar{x} \rangle \geq E(x) - E(\bar{x}) + \varepsilon\|\bar{x} - x\|^2 - D\|x - z\|^2 - \langle \nabla E(\bar{x}), x - \bar{x} \rangle. \tag{A.16}$$

Recall that by the mean value theorem $R(x) = R(\bar{x}) + R_a'(x - \bar{x})$. Thus

$$\|R(x)\| = \|R(x) - R(\bar{x}) + R(\bar{x})\| \leq \|R_a'(x - \bar{x})\| + \|R(\bar{x})\| \leq S'_{\max}\|x - \bar{x}\| + \|R(\bar{x})\|.$$

This, Lemma A.2 with $z = \bar{x}$, $x \in B(\bar{x}, \delta)$, and (A.14) give

$$\begin{aligned} E(x) - E(\bar{x}) - \langle \nabla E(\bar{x}), x - \bar{x} \rangle &= \frac{1}{2}\|R_{\bar{x}}'(x - \bar{x})\|^2 - \frac{1}{8}\|R_a''(x - \bar{x}, x - \bar{x})\|^2 + \langle R_a''(x - \bar{x}, x - \bar{x}), R(x) \rangle \\ &\geq \frac{1}{2}\|R_{\bar{x}}'(x - \bar{x})\|^2 - \frac{S''_{\max}}{8}\|x - \bar{x}\|^4 - S''_{\max}\|R(x)\|\|x - \bar{x}\|^2 \\ &\geq \frac{1}{2}\|R_{\bar{x}}'(x - \bar{x})\|^2 - S''_{\max}\left(\frac{\delta^2}{8} + S'_{\max}\delta + \|R(\bar{x})\|\right)\|x - \bar{x}\|^2 \geq \tilde{\varepsilon}\|x - \bar{x}\|^2. \end{aligned}$$

Plugging this into (A.16) yields (A.15) □

APPENDIX B IMPLEMENTATION DETAILS

We now elaborate on the background gradient approximation scheme and the step length parameters choices used in the numerical experiments.

APPENDIX B.1 BACKGROUND SOLUTION OF PDES

Formally, the gradient of the approximation scheme of Section 4.2.1 takes

$$\widetilde{\nabla E}(x^k) := \nabla S_j(x^j)^* \left(S_j(x^j) - b_k\right),$$

for some past time index $0 \le j \le k$. As we keep the excitation potentials fixed at every frame, $U^k = U^j$ for any $j, k \ge 0$, hence $I^k(x) = I^j(x)$ and $\nabla \tilde{S}_k(x) = \Sigma^{-1/2}\nabla \tilde{I}^k(x) = \Sigma^{-1/2}\nabla I^j(x) = \nabla \tilde{S}_j(x)$ for all x, meaning that $S_k = S_j$.

Lemma B.1. *Suppose $0 < x_m < x_M < \infty$, $\mathcal{X}_k$ is finite dimensional, and that $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain. Further, define E_k by (3.20), and suppose that $S_k = S_j$ and $Y_k = Y_j$ for any $k, j \in \mathbb{N}$, and that E_k, x^k, and $\bar{x}^k$ satisfy the assumptions of Theorem 3.6. Then $E_k(x)$ and $\widetilde{\nabla E}(x^k)$ satisfy Assumption 2.5 with error terms*

$$\bar{e}_k = e_k = \hat{e}_k = (S'_{\max})^2(S_{\max} + \|b_k\|_{Y_k})\delta\|x^j - x^k\|_{X_k}.$$

Proof. Since S_k is differentiable by Lemma 3.3, so is E_k. Let $\check{x} = x^j$ for some $j \le k$. Since $S_k = S_j$ and $Y_k = Y_k$,

$$\begin{aligned}\langle \nabla E_j(x^j), x^k - \bar{x}^k\rangle_{X_j} &= \langle S'_j(x^j)(x^k - \bar{x}^k), S_j(x^j) - b_k\rangle_{Y_j} \\ &= \langle S'_k(x^j)(x^k - \bar{x}^k), S_k(x^j) - b_k\rangle_{Y_k} = \langle \nabla E_k(x^j), x^k - \bar{x}^k\rangle_{X_k}.\end{aligned}$$

Thus for a $z \in \bar{B}(x^k, \|x^j - x^k\|)$, by the mean value theorem,

$$\langle \nabla E_k(x^j), x^k - \bar{x}^k\rangle_{X_k} = \langle \nabla E_k(\check{x}^k), x^k - \bar{x}^k\rangle_{X_k} + \langle \nabla E'_k(z)(x^j - x^k), x^k - \bar{x}^k\rangle_{X_k}.$$

Since $x^k \in B(\bar{x}^k, \delta)$,

$$\begin{aligned}\langle \nabla E'_k(z)(x^j - x^k), x^k - \bar{x}^k\rangle_{X_k} &= \langle S''_k(z)(x^k - \bar{x}^k)(x^j - x^k), S_k(z) - b_k\rangle_{Y_k} \\ &\quad + \langle S'_k(z)(x^k - \bar{x}^k), S'_k(z)(x^j - x^k)\rangle_{Y_k} \\ &\le S''_{\max}(S_{\max} + \|b_k\|_{Y_k})\|x^k - \bar{x}^k\|_{X_k}\|x^j - x^k\|_{X_k} + (S'_{\max})^2\|x^k - \bar{x}^k\|_{X_k}\|x^j - x^k\|_{X_k} \\ &\le S''_{\max}(S_{\max} + \|b_k\|_{Y_k})\delta\|x^j - x^k\|_{X_k} + (S'_{\max})^2\delta\|x^j - x^k\|_{X_k} =: e_k\end{aligned}$$

Now given that S_k satisfies the assumptions of Theorem 3.6, we have

$$\begin{aligned}E_k(x^k) - E_k(\bar{x}^k) + \theta\|x^k - \bar{x}^k\|_X^2 - D\|x^k - \check{x}^k\|_X^2 &\le \langle \nabla E_k(\check{x}^k), x^k - \bar{x}^k\rangle_{X_k} \\ = \langle \nabla E_k(\check{x}^k), x^k - \bar{x}^k\rangle_{X_k} + \langle \nabla E'_k(z)(x^j - x^k), x^k - \bar{x}^k\rangle_{X_k} &\le \langle \nabla E_k(\check{x}^k), x^k - \bar{x}^k\rangle_{X_k} + e_k.\end{aligned}$$

Thus Assumption 2.5 holds as claimed. □

The error term e_k (and likewise $\bar{e}_k$ and $\hat{e}_k$) depends on the measurements b_k and on the linearisation lag through the term $x^k - \bar{x}^k$. The exact values of $S'_{\max}$ and $S''_{\max}$ are given in Section 3.3 and it is easily verified from (3.7) and (3.20) that $S_{\max} = N_2\|\Sigma^{-1/2}\|_2 \max_{j,k}(C_2\|U^{j,k}\|_2 + \|\mathcal{I}^{j,k}\|_2)$.

APPENDIX B.2 STEP LENGTH PARAMETER CHOICE

The reasoning behind the choice of τ and σ is as follows: We set $\kappa_k \equiv \kappa := 0.15$ and assume that $\frac{1}{2}\lambda_{E,k}$, $\bar{\lambda}_{E,k}$, and $\hat{\lambda}_{E,k}$ are all bounded by $L_{\nabla E}$, where $L_{\nabla E}$ satisfies $L_{\nabla E} \le \frac{1}{2}\|\Sigma^{-1/2}\nabla I(\check{x})(\Sigma^{-1/2}\nabla I(\check{x}))^*\|$. This norm is computationally efficient, with a manageable size of 240×240. Under these assumptions, the first term in (2.7b) becomes

$$0.85 > \max\left\{\lambda_{E,k}, -2(\gamma_k + \bar{\gamma}_{E,k}), \frac{2}{1-\kappa_k}\bar{\lambda}_{E,k}, \frac{2}{1-\kappa_k}\hat{\lambda}_{E,k}\right\}$$

From numerical experiments, we determined that $\|K_k\| \approx 16 \cdot 10^{-6}$ and that, in all cases, $2L_{\nabla E} \leq 10^4$. Consequently, the choice of τ and σ ensures that $0.85 \cdot 10^4 \cdot 16 \cdot 10^{-6} < 0.15$, satisfying the condition in (2.7b).

The approximation of $\lambda_{E,k}$, $\bar{\lambda}_{E,k}$, and $\hat{\lambda}_{E,k}$ is heuristic, derived from the convex static case where all these terms would equal $L_{\nabla E}$. Numerically, we observed that the algorithm typically diverged when $\kappa < 0.15$, indicating the reasonability of the approximation.