Journal of Scientific Computing (2024) 101:49
https://doi.org/10.1007/s10915-024-02689-w

Oracle-Net for Nonlinear Compressed Sensing in Electrical Impedance Tomography Reconstruction Problems

Damiana Lazzaro[1] · **Serena Morigi**[1] · **Luca Ratti**[1]

Received: 9 April 2024 / Revised: 16 August 2024 / Accepted: 24 September 2024 /

Abstract

Sparse recovery principles play an important role in solving many nonlinear ill-posed inverse problems. We investigate a variational framework with learned support estimation for compressed sensing sparse reconstructions, where the available measurements are nonlinear and possibly corrupted by noise. A graph neural network, named Oracle-Net, is proposed to predict the support from the nonlinear measurements and is integrated into a regularized recovery model to enforce sparsity. The derived nonsmooth optimization problem is then efficiently solved through a constrained proximal gradient method. Error bounds on the approximate solution of the proposed Oracle-based optimization are provided in the context of the ill-posed Electrical Impedance Tomography problem (EIT). Numerical solutions of the EIT nonlinear inverse reconstruction problem confirm the potential of the proposed method which improves the reconstruction quality from undersampled measurements, under sparsity assumptions.

Keywords Nonlinear inverse problems · Compressed sensing; Electrical Impedance Tomography · Sparsity-inducing regularization · Nonsmooth numerical optimization · Graph neural networks

1 Introduction

The theory of compressed sensing (CS) is a successful mathematical technology in sparse signal recovery, established several years ago independently by Donoho [21] and Candès [12], successfully carried out by [13, 14], and widespreadly developed in the last two decades. The main idea behind CS is to acquire a small number of linear measurements of a signal that exhibits sparsity or compressibility in a known representation system, and accurately reconstruct the original sparse signal using an appropriate reconstruction algorithm, under the assumption that the sensing matrix used in acquisition is sufficiently incoherent with respect

---

✉ Luca Ratti
luca.ratti5@unibo.it

Damiana Lazzaro
damiana.lazzaro@unibo.it

Serena Morigi
serena.morigi@unibo.it

[1] Department of Mathematics, University of Bologna, Bologna, Italy

to the basis in which the signal is sparse, a property that can be quantified e.g. by verifying the well-known Restricted Isometry Property (RIP). The standard recovery algorithms for finding an approximated CS solution are based either on a greedy approach or on variational models, such as ℓ^1-norm minimization, leading to suitable iterative thresholded gradient descent methods.

A significant portion of research focuses on *linear* CS problems and holds for *well-posed* compressive sensing models. However many real-world applications in physics and biomedical sciences involve inherent nonlinearities. In these cases, the linear model becomes inadequate.

We are interested in a more general setting that extends the concepts of compressive sensing and sparse recovery to inverse and *ill-posed nonlinear* problems.

In this more general setting, we consider the problem of recovering an unknown vector $\sigma^\dagger \in \mathbb{R}^n$ from incomplete and contaminated nonlinear measurements $\Lambda^\delta \in \mathbb{R}^m$, according to the following degradation model

$$\Lambda^\delta = \Phi(\sigma^\dagger) + \eta, \tag{1}$$

where $\eta \in \mathbb{R}^m$ is a vector of unknown perturbations, bounded by a known amount $\|\eta\|_2 \leq \delta$, and $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ represents a nonlinear ill-posed sensing model characterized by an undersampled acquisition in which the number m of available measurements is much smaller ($m \ll n$) than the dimension n of the vector σ.

Moreover we assume that, although σ belongs to the high-dimensional space $\mathbb{R}^n$, it can be represented by a few degrees of freedom. In particular, we depart from the traditional sparsity assumption and seek conductivities σ that exhibit σ_0-sparsity, being σ_0 a reference constant vector. This accounts to say that σ is equal to σ_0 apart from a few coefficients, thus the set of indices of $I_{\mathcal{O}} = \{i : (\sigma_i - \sigma_{0,i}) \neq 0\}$ is assumed to have small cardinality s. In the context of CS, several algorithms have been proposed to enforce (or promote) this property, based on projection and variational techniques. Motivated by the successful application of compressive sensing to linear problems, and aiming to overcome the ill-posed nature that commonly prevents problem (1) from having a unique and stable solution, we propose to recover the vector $\sigma^\dagger$ by solving the following regularized minimization problem:

$$\min_{\sigma \in \mathbb{R}^n} \mathcal{R}(\sigma) \quad \text{subject to} \quad \|\Phi(\sigma) - \Lambda^\delta\|_2 \leq \delta, \tag{2}$$

where $\mathcal{R} \in \Gamma_0(\mathbb{R}^n)$ is a sparsity-promoting penalty and we denote by $\Gamma_0(\mathbb{R}^n)$ the space of proper, lower semi-continuous, and convex functionals on $\mathbb{R}^n$. In particular, we consider functionals $\mathcal{R}$ that pursue sparsity promotion in two different, complementary, ways: by promoting low values of a penalization functional and by projecting on a predetermined set K. The most standard instance in the class of nonsmooth, convex penalties is the ℓ^1-norm functional, which is known to promote sparsity in the solution domain; nevertheless, we may also opt for other choices, such as the Total Variation penalty [51], to enforce other significant prior knowledge on the solution, such as sparsity in the gradient domain. On the other hand, the projection onto the set K can be employed to impose state constraints (such as non-negativity, or box constraint) or to enforce sparsity likewise. The approach we propose in this paper draws connections with the problem known as *support estimation*, namely, with the task of determining the support of the exact solution from the given measurements. We thus introduce the concept of *Oracle*: an operator receiving as an input the measurements Λ^δ and returning a (perfect or approximate) description of its support K, for example through the set of active indices $I_{\mathcal{O}}$. Such an Oracle operator is determined before the resolution of problem (2) to improve the computational efficiency of the minimization method for solving

(2). In this paper, we pursue this task in a supervised statistical learning fashion, by training a Neural Network via a dataset of ground truth solutions (hence, their supports) and associated measurements. We consider, in particular, Graph Neural Networks (GNNs) based on a graph U-Net architecture, which naturally fits the mesh data structure used to represent the spatial domain. GNNs are a class of neural network architectures designed to perform inference on data described by graphs [58]. The benefit of using GNN is related to the specific nonlinear problem considered as a test case of the proposed CS framework, which requires a Finite Element Method (FEM) solution on a mesh domain. Once the Oracle has been learned, we can employ the set K within the definition of $\mathcal{R}$ and tackle the solution of problem (2).

Independently of the choice of $\mathcal{R}$, our strategy then envisages the solution of (2) using a first-order optimization scheme, namely, the Proximal Gradient Method (PGM). Despite the regularizing effect of $\mathcal{R}$, because of the nonlinear nature of Φ, the convergence properties of such a scheme are only guaranteed under additional assumptions.

The main theoretical results of this paper are reported in Proposition 1 and in Theorem 2. The first one shows that minimizer of the proposed regularization functional converges to a solution of the inverse problem (1) as the noise level δ goes to zero, also providing some convergence rates. The second result concerns error bounds on the approximate solution obtained via the proposed Oracle-based PGM. Drawing from a result in [8], we show the convergence of such a scheme to a cluster point whose distance from an exact solution of (1) can be bounded in terms of δ (and, possibly, of the choice of the support K). In order to derive similar error bounds in the context of linear CS, it is usually necessary to introduce theoretical assumptions on the forward operator, such as the RIP, which do not carry over directly to the nonlinear case. In this work, we show that, under the requirements that the Jacobian of the measurement system Φ satisfies a similar RIP condition and that Φ is mildly nonlinear, the accurate recovery of $\sigma^{\dagger}$ is possible using the proposed Oracle-based PGM algorithm.

Given the considerable challenges associated with general nonlinear CS recovering problems, we will concentrate on a specific nonlinear measurement process inherent to Electrical Impedance Tomography. Nevertheless, our Oracle-based framework could be adapted to other nonlinear ill-posed problems. EIT is a promising non-invasive imaging modality mathematically characterized as a highly nonlinear and ill-posed inverse problem [53]. EIT aims to reconstruct the inner conductivity distribution of a medium starting from a set of measured voltages registered by a series of electrodes that are positioned on the surface of the medium. EIT is therefore a noninvasive technique, meaning that it allows analyzing the inner properties of a material or structure without causing damage.

Several CS strategies have been explored in the context of linearized EIT reconstruction problem, see [29, 47], as a response to a growing application need. By leveraging CS, in [52] the authors explore the possibility of achieving accurate breathing monitoring with EIT while reducing the number of measurements needed. The linearized EIT problem is considered from a theoretical perspective in [2] where the authors show that the electrical conductivity may be stably recovered from a number of linearized EIT measurements proportional to the sparsity of the signal with respect to a wavelet basis, up to a log factor.

Real-world EIT data exhibits nonlinearities. To address this challenge, Zong et al. in [59] recently proposed a novel scheme that introduces the concept of compressive learning. Recently, a learned residual approach to EIT has been proposed in [19] for nonlinear undersampled measurements. Moreover, although not specifically labeled as compressed sensing, the authors in [27], [28], [31] treat the reconstruction from EIT nonlinear measurements using sparsity priors. In [16] a mask-guided spatial-temporal graph neural network is proposed to reconstruct multifrequency EIT images in cell culture imaging. The binary masks in [16] are obtained by CT scans and microscopic images using a multimodal imaging setup. In this work, the concept of compressed learning is realized through an Oracle-Net that predicts the optimal prior in the variational model which improves the efficacy of the sparse optimization algorithm. The binary masks are thus automatically determined by the Oracle-Net starting from given measurements. We finally would shed light on the problem of determining how few measurements suffice for an accurate EIT sparsity-regularized reconstruction, a previously unexplored problem in the nonlinear CS context.

The remainder of this paper is structured as follows. In Section 1.1 related works on nonlinear compressed sensing are presented. The link to linear CS is detailed in Section 2. Section 3 presents the EIT forward and inverse models in their continuous settings. In Section 4 we provide general theoretical results on the well-posedness of the regularized problem. In Section 5 we introduce the PGM-based numerical method for EIT reconstruction, and we prove error bounds on the sparsity-aware approximate solution. Section 6 examines some specific regularization models employed and discusses their algorithmic optimization. Section 7 explores the architecture of the proposed Oracle-Net for support estimation. In Section 8 we present the reconstruction results obtained by the Oracle-based PGM algorithm. Finally, Section 9 draws conclusions and discusses future work.

1.1 Related Works

The treatment of linear compressed measurement models for ill-posed problems and recovery accuracy estimates has been first addressed in [34].

A preliminary exploration of nonlinear CS can be found in the work on CS phase retrieval [46] and on 1-bit CS [11]. A first attempt to extend greedy gradient-based strategies to the nonlinear case is in [9]. Subsequently, in [8] Blumensath pioneered the theoretical framework for nonlinear compressed sensing. It has been demonstrated that, under conditions similar to the RIP, the iterative hard thresholding (IHT) algorithm can effectively recover sparse or structured signals from a small number of nonlinear measurements. Along this direction, further researches introduced various approaches for tackling nonlinear CS problems. In [48, 49] Ohlsson et al. proposed algorithms like quadratic basis pursuit and nonlinear basis pursuit, both of which utilize convex relaxations to solve these problems. In addition, research on sparsity-constrained nonlinear optimization, as explored in [5], further enriched the field of nonlinear CS. A generalization of the RIP condition for certain randomized quasi-linear measurements is proposed in [22].

The power of nonlinear CS extends beyond theoretical developments. Practical applications validate its effectiveness. For example, works like [39] demonstrate its application in single-snapshot compressive tomosynthesis, while the proposal in [44] leverages the inherent sparsity of mask patterns to formulate the Optical Proximity Correction problems as inverse nonlinear CS problems, enabling significant efficiency gains.

The aforementioned works seem to provide a CS solution to nonlinear inverse problems neglecting the effect of the ill-posedness of several nonlinear problems. In this work, the combined effect of a sparsifying Oracle-Net and an effective regularization guarantees accurate reconstructions of the original sparse signal using a limited number of nonlinear measurements.

2 Connections to Linear Compressed Sensing

The central focus of this paper is to tackle the variational regularization strategy (2) of the inverse problem (1) that, by a careful choice of the functional $\mathcal{R}$ (see Section 4) can be addressed to as a *sparsity-promoting* regularization. This approach aligns closely with standard compressed sensing, but incorporates distinct characteristics: the aim of this section is to recall the main pillars of linear CS and to show that, although with significant differences, the method we propose and its analysis can be meaningfully understood within the same framework.

i) *Sparsity.* In linear CS, the ground truth signal, or object, to be reconstructed is assumed to be sparse in a suitable basis or dictionary.
   This is also the case in our approach as we assume that the unknown conductivity (or its difference with respect to a known background, or its discretized gradient) has only a few non-vanishing components.

ii) *Incoherence.* The key feature allowing for compressibility in linear CS is that the acquisition operator is sufficiently incoherent with respect to the basis in which the solution is sparse. This concept essentially guarantees that even a few measurements contain information on all the components of the unknown signal. The incoherence property can be expressed through several theoretical properties of the measurement operator Φ, the most relevant of which is the Restricted Isometry Property. This assumption requires that, when restricted to $s-$sparse vectors, the forward operator is invertible, and close to an isometry. Such a condition can be fulfilled (with high probability) in random sensing procedures but is hard to verify in most practical acquisition scenarios (a successful case being, e.g., MRI). In the context of linear problems, [32] explored the connections between RIP and source conditions, which are more common in the theoretical analysis of inverse problems, but are nevertheless rather complicated to verify in applications. In [8], a useful extension of RIP to the nonlinear case is provided, with a few envisaged applications.
   In our analysis, we leverage the RIP on (1) in the version proposed in [8], although with a significant difference. Indeed, by leveraging the theoretical analysis of the local injectivity properties of the EIT problem, we decouple the verification of the RIP from the sparsity level of the unknown conductivity. In particular, we observe that the assumptions of [8] are verified whenever the conductivity belongs to a finite-dimensional subset, e.g., it is discretized through a fixed partition of the physical domain. Although this represents a much more general interpretation of the RIP, allowing us to apply it also in the context of EIT, the decoupling from the sparsity level hinders the possibility of controlling the constants appearing in the RIP explicitly, especially the one in the lower bound. This is reflected in the discussion related to the following points. We nevertheless believe that referring to (21) as a Restricted Isometry Property might strengthen the connection between the CS and the nonlinear EIT literature.

iii) *Sparsity-promoting ℓ^1-norm minimization.* Once the theoretical properties are clearly stated, one of the most commonly used techniques to solve a linear CS problem is to minimize a suitable variational regularization problem containing a ℓ^1-norm penalty functional, leading to the so-called LASSO problem, Total Variation, or many other alternatives. Although the minimizers of such functionals inherently share suitable sparsity properties (see, e.g., [54]), their theoretical analysis is challenging due to the lack of differentiability and strong convexity of the ℓ^1-norm. These difficulties become more severe in the context of (nonlinear) ill-posed problems.

To mitigate these shortcomings, several strategies are possible: in this paper, we introduce in $\mathcal{R}$ an additional ℓ^2-norm regularization term. This strategy, also known as Elastic-Net regularization, is also mentioned in [1], where a suitable extension of the RIP is provided in the case of ill-posed linear problems. Moreover, we enhance the regularization term $\mathcal{R}$ with an additional term imposing the projection onto a specific subset of coefficients. This term does not affect the theoretical property of our strategy, but significantly improves the quality of the reconstructions, especially when designed via data-driven support estimation techniques (Oracle-Net). In classical CS a similar term is not needed, and it is possible to prove (in the noiseless case) the exact recovery of support by simply minimizing the ℓ^1-norm. In our nonlinear, ill-posed scenario, as there are no available exact recovering guarantees for sparse signals, such an Oracle proves to be significantly useful.

iv) *Recovery guarantees.* As already anticipated, one of the most relevant outcomes of the theoretical analysis of linear CS problems, e.g. under the assumption that the RIP is verified, are some exact or asymptotic recovery estimates (in the case of clean or noisy data) of the unknown signal. In such a context, a crucial task is to link the number of measurements required to achieve similar guarantees with the sparsity level of the ground truth signals.
The main obstacle to obtain similar results for our technique is represented by the fact that the RIP condition is, in our case, completely decoupled from the sparsity level. Although this outlines a distance with respect to the goals and achievements of the classical CS literature, we want to emphasize that the introduced setup allows us to pursue a different goal in the theoretical analysis of CS: namely, the deduction of convergence guarantees for the iterative reconstruction method - a topic extensively studied in [10] and [8] for the Iterative Hard Thresholding algorithm applied to fairly general linear and nonlinear problems, and in [30] in the context of inverse scattering. We also point out that, even though the theoretical deduction of recovery estimates in terms of the sparsity level is beyond the purposes of this work, we tried to investigate their validity via a proof-of-concept numerical experiment, as we reported in Section 8.3.

3 EIT Forward and Inverse Models

In this section, we describe the EIT problem in its original continuous setup, focusing on the so-called *complete electrode model* (CEM), both from a forward and an inverse perspective. Our goal is to clearly state the problem of interest in a functional space framework, as well as to motivate the discretization that will be later adopted. Moreover, we collect in Theorem 1 the most relevant theoretical results regarding the nonlinear operator describing the problem, mainly relying on the approach and assumptions adopted in [41].

Consider a conductive body in a 2-dimensional space, described as a bounded, simply connected, Lipschitz domain $\Omega \subset \mathbb{R}^2$ with a piece-wise C^∞ boundary $\partial\Omega$. Its electrical conductivity is modeled as a function $\sigma : \Omega \to \mathbb{R}$, which we assume to be bounded, measurable, and larger than a strictly positive constant c_0:

$$\sigma \in L^\infty(\Omega), \qquad \sigma(x) \geq c_0 > 0 \quad \text{a.e. in } \Omega. \tag{3}$$

The classical formulation of the EIT problem, also known as *continuum model*, seeks to reconstruct the function σ by applying electrical stimuli on the boundary $\partial\Omega$ (i.e., injected currents), and recording the resulting electrical potential, again on the whole boundary $\partial\Omega$.

A more realistic scenario is described via CEM, in which one assumes that the process of probing the electrical properties of Ω is performed by means of some electrodes located on the boundary of the domain. Consider a collection of p electrodes $\{E_j\}_{j=1}^p \subset \partial\Omega$: from a modeling perspective, they are resistive regions (with electrical permittivity z), on which it is possible to apply an external current I_j. We assume that each E_j is an open, non-empty subset of $\partial\Omega$, and that $E_i \cap E_j = \emptyset$ for $i \neq j$. The electrical measurements, both of currents and voltages, associated with the CEM are assumed to be piece-wise constant functions on $\partial\Omega$: we define the associated space as

$$\mathcal{E}_p = \left\{ F \in L^2(\partial\Omega) : \quad F(x) = \sum_{j=1}^{p} F_j \mathbf{1}_{E_j}(x), \quad \int_{\partial\Omega} F(y)dy = \sum_{j=1}^{p} F_j |E_j| = 0 \right\}, \tag{4}$$

where $\mathbf{1}_{E_j}$ denotes the indicator function of $E_j \subset \partial\Omega$. Notice that $F \in \mathcal{E}_p$ can be uniquely determined by means of $p-1$ real parameters, thus $\mathcal{E}_p$ can be equivalently represented as $\mathbb{R}^{p-1}$. We denote by $E = \bigcup_{j=1}^p E_j$ and by $\Gamma = \partial\Omega \setminus E$.

The forward problem of EIT, in the CEM paradigm, consists in determining the couple $(u, U) \in Y = H^1(\Omega) \times \mathcal{E}_p$, representing the inner electrical potential and the boundary voltages at the electrodes, associated with the input boundary currents $I \in \mathcal{E}_p$ through the following differential problem:

$$\begin{cases} -\operatorname{div}(\sigma \nabla u) = 0 & \text{in } \Omega, \\ u + z\sigma \dfrac{\partial u}{\partial n} = U_l & \text{on } \{E_j\}_{j=1}^p, \\ \displaystyle\int_{E_l} \sigma \frac{\partial u}{\partial n} \, \mathrm{d}s = I_l & j = 1, \dots, p, \\ \sigma \dfrac{\partial u}{\partial n} = 0 & \text{on } \Gamma. \end{cases} \tag{5}$$

Here z represents the known contact impedance of the electrodes. The existence and uniqueness of a weak solution of the boundary problem (5) is proved in [53], and its numerical approximation is effectively treated, e.g., via finite element method (FEM).

Denote by $\mathcal{L}(\mathcal{E}_p)$ the space of bounded linear operators from $\mathcal{E}_p$ to $\mathcal{E}_p$ (isomorphic to the matrix space $\mathbb{R}^{(p-1)\times(p-1)}$). For a fixed conductivity σ satisfying (3), the well-posedness result implies that, for every choice of input boundary current $I \in \mathcal{E}_p$, the couple $(u, U) \in Y$ is uniquely determined, and we can denote by Λ_σ^δ the operator that associates a current $I \in \mathcal{E}_p$ with the corresponding boundary voltage $U \in \mathcal{E}_p$. We can thus define the *forward map* of EIT, namely, the operator $F_p : L^\infty(\Omega) \to \mathcal{L}(\mathcal{E}_p)$, which returns, for each conductivity σ, the currents-to-voltages operator Λ_σ^δ:

$$F_p : \sigma \in L^\infty(\Omega) \;\mapsto\; \Lambda_\sigma^\delta = F_p(\sigma) \in \mathcal{L}(\mathcal{E}_p) : \quad \Lambda_\sigma^\delta I = U \;\text{ s.t. } (u, U) \;\text{ solves (5).} \tag{6}$$

Notice that for each σ, the operator Λ_σ^δ can be identified by a square matrix in $\mathbb{R}^{(p-1)\times(p-1)}$. Moreover, since the problem (5) is symmetric in the variables U and I, we can efficiently represent Λ_σ^δ as a symmetric matrix, or as a vector in $\mathbb{R}^{m_p}$, being $m_p = \frac{p(p-1)}{2}$ the number of independent measurements associated with p electrodes.

The *inverse problem* of EIT is to reconstruct σ from the knowledge of Λ_σ^δ. Observe that, since the available measurements are only finite-dimensional, we can only hope to recover conductivities depending on a finite number of degrees of freedom. In particular, we introduce a conformal triangular partition $\mathscr{T}$ over Ω and define the Finite Element space of piecewise affine, continuous, and uniformly bounded functions

$$W_{\mathscr{T}} = \left\{ v \in C(\overline{\Omega}) \quad \text{s.t.} \quad v|_{T_k} \in \mathbb{P}_1(T) \;\; \forall T \in \mathscr{T} \right\}, \tag{7}$$

where $\mathbb{P}_1(T)$ denotes the space of polynomials of degree 1 on the triangle T. This choice fits the assumptions in [41], where the partition can also be non-triangular and non-conformal and the functions in $W_{\mathscr{T}}$ are bounded piece-wise polynomials of arbitrary degree.

A n-dimensional basis can be defined on $W_{\mathscr{T}}$, where n is the number of vertices in the partition $\mathscr{T}$, by means of the piece-wise affine functions $\{\phi_i\}_{i=1}^n$, each of which attains the values 1 at the i-th vertex and 0 at all the other ones. A function in $W_{\mathscr{T}}$ is thus determined by its values at the vertices, $\{v_i\}_{i=1}^n$, which implies that $W_{\mathscr{T}}$ can be identified with $\mathbb{R}^n$, and the forward map F_p restricted on $V_{\mathscr{T}}$ can be interpreted as a nonlinear function from $\mathbb{R}^n$ to $\mathbb{R}^{m_p}$. Its Fréchet derivative with respect to the L^∞ norm, denoted by F_p' can therefore be represented as a $m_p \times n$ matrix, the Jacobian matrix of the described vector field.

In [41], several results regarding the map F_p (and its derivative F_p') can be found. We are nevertheless interested in the properties of a strongly related map, $F_{p,h}$, which arises by a further discretization of the problem, induced by the numerical approximation of the differential problem (5). In particular, as in [41], we introduce a computational mesh $\mathscr{T}_h$ on Ω, in general unrelated to the previously introduced $\mathscr{T}$, and consisting of N_h triangles. We approximate the differential problem (5) through a Finite Element scheme, and introduce the operator $\Lambda^\delta_{\sigma,h} \in \mathcal{L}(\mathcal{E}_p)$, which maps every $I_h \in \mathcal{E}_p$ into the potential $\Lambda^\delta_{\sigma,h} I_h = U_h$ such that (u_h, U_h) is the output of the Finite-Element solver of (5). Then, we can define $F_{p,h}$ as the map from $W_{\mathscr{T}}$ to $\mathcal{L}(\mathcal{E}_p)$ that associates a conductivity σ to $\Lambda^\delta_{\sigma,h}$. We now recap the main theoretical properties satisfied by $F_{p,h}$, which are analyzed in [41]. In particular, we employ further knowledge of the desired solution to limit our search to the space

$$V_{\mathscr{T}} = \left\{ v \in W_{\mathscr{T}} \quad \text{s.t.} \quad c_0 \le v(x) \le c_1 \;\; \forall x \in \Omega \right\}, \tag{8}$$

being $0 < c_0 < c_1$. These uniform bounds of the desired conductivities allow for a simpler expression for the estimates contained in [41].

Theorem 1 (Properties of $F_{p,h}$ and $F'_{p,h}$, from [41])

1. *The operator $F_{p,h}$ is Fréchet differentiable and its derivative $F'_{p,h}(\sigma)$ is Lipschitz continuous for all $\sigma \in V_{\mathscr{T}}$*
2. *Local injectivity: there exist an integer $p_{\mathscr{T}}$ and two positive constants $h_{\mathscr{T}}$, $\Gamma_{\mathscr{T}}$ such that, for $p > p_{\mathscr{T}}$ and $h \le h_{\mathscr{T}}$,*

$$\|F'_{p,h}(\sigma)[\theta]\|_{\mathcal{L}(\mathcal{E}_p)} \ge \Gamma_{\mathscr{T}} \|\theta\|_{L^\infty} \quad \forall \sigma \in V_{\mathscr{T}}, \; \forall \theta \in W_{\mathscr{T}}.$$

3. *Tangential cone condition: for each $\sigma \in \mathrm{int}(V_{\mathscr{T}})$, there exist a radius $r_{\mathscr{T}}$ and a constant $C_{\mathscr{T}}$ such that, for $p > p_{\mathscr{T}}$ and $h \le h_{\mathscr{T}}$, if $\|\tau - \sigma\| \le r_{\mathscr{T}}$,*

$$\|F_{p,h}(\tau) - F_{p,h}(\sigma) - F'_{p,h}(\sigma)[\tau - \sigma]\|_{\mathcal{L}(\mathcal{E}_p)} \le C_{\mathscr{T}} \|F_{p,h}(\tau) - F_{p,h}(\sigma)\|_{\mathcal{L}(\mathcal{E}_p)}$$

Note that the result on the tangential cone condition controls the linearization error by the nonlinear Taylor remainder.

The first statement follows by a combination of Lemma 2.3, Lemma 4.1, and Lemma 4.6 in [41] (see also [40], [57], [36]), whereas the second and the third statements, with minimal modifications, are the objects of Theorem 4.7 and Theorem 4.9 in [41], respectively. In contrast with the formulation of these results presented therein, the values of $p_{\mathscr{T}}$, $h_{\mathscr{T}}$, $\Gamma_{\mathscr{T}}$

depend on the bounds c_0, c_1 on the conductivities, which is considered as a parameter in the definition of $V_{\mathscr{T}}$. Explicit expressions or bounds for $p_{\mathscr{T}}, h_{\mathscr{T}}$ and $\Gamma_{\mathscr{T}}$ are not available, and their definitions are not constructive, as they also involve the Fréchet derivative on the forward map of F_p and of the analogous operator of the continuum model of EIT.

4 A Regularized Constrained Model for the EIT Inverse problem

In this section, we introduce a reconstruction model for the EIT inverse problem in the CEM formulation. Thanks to the discrete nature both of σ and of the measurements $\Lambda^{\delta}_{\sigma,h}$, this problem can be formulated as the resolution of a nonlinear (ill-posed) system of algebraic equations. We introduce a variational regularization strategy, which involves the minimization of a (non-convex, non-smooth) functional, for which we will describe in Section 5 an iterative scheme based on the Proximal-Gradient method [6].

As already discussed in Section 3, we look for solutions of the inverse problem of EIT in a finite-dimensional space of piece-wise constant conductivities, $W_{\mathscr{T}}$, which can be naturally identified with $\mathbb{R}^n$, and more precisely within $W_{\mathscr{T}}$, identified with the hyper-cube $K_{0,1} = [c_0, c_1]^n$. The datum of the inverse problem, i.e., the currents-to-voltage operator $\Lambda^{\delta}_{\sigma,h}$, also belongs to a finite-dimensional space, $\mathcal{L}(\mathcal{E}_{p,h})$, which can be identified with $\mathbb{R}^{m_p}$. Nevertheless, in the context of applications, it is most common to provide alternative, often redundant, representations of the operator $\Lambda^{\delta}_{\sigma,h}$ by considering n_c different current patterns (often associated with the activation of a few, adjacent or opposite, electrodes) and recording the voltage in (other) n_v electrodes. All these modeling choices define the so-called measurement protocol, as well as the total number of measurements $m = n_c n_v$. We assume that $m \geq m_p$, and identify the measurement space with $\mathbb{R}^m$.

For fixed discrete mesh $\mathscr{T}$ and measurement protocol, we denote by $\Phi\colon \mathbb{R}^n \to \mathbb{R}^m$ the nonlinear operator representing $F_{p,h}$ from $W_{\mathscr{T}}$ to the (redundant, m-dimensional) representation of the measurement space $\mathcal{L}(\mathcal{E}_{p,h})$. The inverse problem of EIT is thus equivalent to recovering $\sigma^{\dagger} \in K_{0,1} = [c_0, c_1]^n$ from the noisy measurements $\Lambda^{\delta} \in \mathbb{R}^m$ under the degradation model (1).

In this section, we analyze the variational regularization strategy associated with the following constrained optimization problem:

$$\sigma^{\delta}_{\lambda} \in \underset{\sigma \in K}{\operatorname{argmin}} \left\{ \mathcal{J}^{\delta}_{\lambda}(\sigma) := \frac{1}{2}\|\Phi(\sigma) - \Lambda^{\delta}\|^2 + \frac{\lambda\rho}{2}\|\sigma\|^2 + \lambda R(\sigma) \right\}, \tag{9}$$

where $\rho, \lambda > 0$, $R\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a non-negative, continuous, coercive, and convex functional, and K is a compact, convex subset of $\mathbb{R}^n$ satisfying $K \subset K_{0,1}$.

By introducing the characteristic function χ_K of the set $K \subset \mathbb{R}^n$, namely

$$\chi_K(x) = \begin{cases} 0 & \text{if } x \in K \\ \infty & \text{if } x \notin K \end{cases},$$

allows to incorporate the constraint $\sigma \in K$ into the minimization problem (9), thus obtaining the following equivalent unconstrained optimization problem

$$\sigma^{\delta}_{\lambda} \in \underset{\sigma \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \mathcal{J}^{\delta}_{\lambda}(\sigma) := \frac{1}{2}\|\Phi(\sigma) - \Lambda^{\delta}\|^2 + \frac{\lambda\rho}{2}\|\sigma\|^2 + \lambda R(\sigma) + \chi_K(\sigma) \right\}. \tag{10}$$

The functional $\mathcal{J}^{\delta}_{\lambda}$ presents two regularization terms (namely, the square norm and the convex functional R). It would be possible to consider the two regularization parameters as independent: nevertheless, as this would not provide any significant difference in the theoretical analysis, similarly to the Elastic-Net paradigm [20, 60], we consider their ratio ρ as fixed and interpret the functional as depending on a single parameter, λ.

In the following we discuss some theoretical properties of problem (10) and of its solutions σ_λ^δ. To ease the notation, we denote the overall regularization functional employed in (10) by

$$\mathcal{R}(\sigma) = R(\sigma) + \frac{\rho}{2}\|\sigma\|^2 + \chi_K(\sigma), \tag{11}$$

so that the functional to be minimized can be simply written as

$$\mathcal{J}_\lambda^\delta(\sigma) = \frac{1}{2}\|\Phi(\sigma) - \Lambda^\delta\|^2 + \lambda\mathcal{R}(\sigma).$$

It is easy to observe that such a functional is continuous and coercive and that the minimization is performed on a compact set K: thus, the existence of (at least) a solution of (10) is guaranteed by classic arguments (see e.g. [7]). Nevertheless, we cannot conclude the uniqueness of such solutions due to the nonlinearity of Φ.

The solutions of (10) are also stable with respect to perturbations of the datum Λ^δ: introducing a sequence $\{\Lambda_k\} \subset \mathbb{R}^m$ such that $\Lambda_k \to \Lambda^\delta$, and considering a sequence of minimizers $\{\sigma_k\}$ of the functionals $\mathcal{J}_\lambda^k$, obtained by replacing Λ^δ with Λ_k in $\mathcal{J}_\lambda^\delta$, then the limit of every convergent subsequence of $\{\sigma_k\}$ is a minimizer of $\mathcal{J}_\lambda^\delta$. This can be proved, with slight modifications, as in [23, Theorem 10.2].

We now focus our attention on extending the convergence result [23, Theorem 10.4] to the optimization problem (10). Indeed we prove that, under suitable assumptions on Φ and $\sigma^\dagger$ and for a specific choice of λ, as $\delta \to 0$, the minimizers σ_λ^δ of (10) converge to a solution of the inverse problem $\sigma^\dagger$. We state the result in a general formulation, outlining all the properties required on the operator Φ, which are verified in our setup as discussed in Remark 1. Hereafter, we denote by $J_\Phi(\sigma) \in \mathbb{R}^{m\times n}$ the Jacobian matrix of Φ computed at a point $\sigma \in \mathbb{R}^n$.

Proposition 1 *Let σ_λ^δ be a local minimizers of (10), being $\mathcal{R}$ as in (11) and Φ such that the following assumptions are satisfied:*

1. *Source condition: there exist $p^\dagger \in \partial\mathcal{R}(\sigma^\dagger)$ and $w \in \mathbb{R}^m$ such that*

$$p^\dagger = J_\Phi(\sigma^\dagger)^T w \tag{12}$$

2. *Mild non-linearity of Φ in $\sigma^\dagger$: there exists $\gamma > 0$ such that*

$$\|\Phi(\sigma) - \Phi(\sigma^\dagger) - J_\Phi(\sigma^\dagger)(\sigma - \sigma^\dagger)\|^2 \leq \gamma\|\sigma - \sigma^\dagger\|^2 \qquad \forall\sigma \in K_{0,1}. \tag{13}$$

Assume moreover that $2\gamma\|w\| \leq \rho$. Then, as $\lambda \to 0$, the sequence σ_λ^δ converges to $\sigma^\dagger$ and, for the choice $\lambda \sim \delta$, the following convergence rate holds:

$$\|\sigma_\lambda^\delta - \sigma^\dagger\| = O(\sqrt{\delta}) \tag{14}$$

Proof Since σ_λ^δ is a minimizer of (10), it holds

$$\frac{1}{2}\|\Phi(\sigma_\lambda^\delta) - \Lambda^\delta\|^2 + \lambda\mathcal{R}(\sigma_\lambda^\delta) \leq \frac{1}{2}\|\Phi(\sigma^\dagger) - \Lambda^\delta\|^2 + \lambda\mathcal{R}(\sigma^\dagger);$$

subtracting the term $\lambda\langle p^\dagger, \sigma_\lambda^\delta - \sigma^\dagger\rangle$ on both sides, being $p^\dagger$ as in (12), we get

$$\frac{1}{2}\|\Phi(\sigma_\lambda^\delta)-\Lambda^\delta\|^2+\lambda\big(\mathcal{R}(\sigma_\lambda^\delta)-\mathcal{R}(\sigma^\dagger)-\langle p^\dagger,\sigma_\lambda^\delta-\sigma^\dagger\rangle\big)\le\frac{1}{2}\|\eta\|^2-\lambda\langle p^\dagger,\sigma_\lambda^\delta-\sigma^\dagger\rangle. \tag{15}$$

The second term on the left-hand side of (15) can be interpreted in terms of the Bregman divergence associated with the convex functional $\mathcal{R}$. In particular, it can be denoted as $D_{\mathcal{R}}^{p^\dagger}(\sigma_\lambda^\delta,\sigma^\dagger)$, where for any σ_1,σ_2 the Bregman divergence from σ_2 to σ_2, with respect to $p_2\in\partial\mathcal{R}(\sigma_2)$ is defined as

$$D_{\mathcal{R}}^{p_2}(\sigma_1,\sigma_2)=\mathcal{R}(\sigma_1)-\mathcal{R}(\sigma_2)-\langle p_2,\sigma_1-\sigma_2\rangle.$$

Notice that, for the specific choice of $\mathcal{R}$ outlined in (11), it holds that, for $\sigma_1,\sigma_2\in K$, any $p_2\in\partial\mathcal{R}$ can be written as $\rho\sigma_2+q_2$, with $q_2\in\partial R(\sigma_2)$, and thanks to the convexity of R,

$$\begin{aligned}D_{\mathcal{R}}^{p_2}(\sigma_1,\sigma_2)&=R(\sigma_1)+\frac{\rho}{2}\|\sigma_1\|^2-R(\sigma_2)-\frac{\rho}{2}\|\sigma_2\|^2-\langle\rho\sigma_2+q_2,\sigma_1-\sigma_2\rangle\\&=\frac{\rho}{2}\left(\|\sigma_1\|^2-\|\sigma_2\|^2-2\langle\sigma_2,\sigma_1-\sigma_2\rangle\right)+\big(R(\sigma_1)-R(\sigma_2)-\langle q_2,\sigma_1-\sigma_2\rangle\big)\\&\ge\frac{\rho}{2}\|\sigma_1-\sigma_2\|^2.\end{aligned}$$

Inserting this in (15), and employing the source condition (12), we obtain

$$\frac{1}{2}\|\Phi(\sigma_\lambda^\delta)-\Lambda^\delta\|^2+\frac{\lambda\rho}{2}\|\sigma_\lambda^\delta-\sigma^\dagger\|^2\le\frac{1}{2}\|\eta\|^2-\lambda\langle w,J_\Phi(\sigma^\dagger)(\sigma_\lambda^\delta-\sigma^\dagger)\rangle. \tag{16}$$

Let us now focus on the second term on the right-hand side of (16): adding and subtracting various terms,

$$\begin{aligned}-\langle w,J_\Phi(\sigma^\dagger)(\sigma_\lambda^\delta-\sigma^\dagger)\rangle\;&=\langle w,\big(\Phi(\sigma_\lambda^\delta)-\Phi(\sigma^\dagger)-J_\Phi(\sigma^\dagger)(\sigma_\lambda^\delta-\sigma^\dagger)\big)\\&\qquad+\big(\Lambda^\delta-\Phi(\sigma_\lambda^\delta)\big)+\big(\Phi(\sigma^\dagger)-\Lambda^\delta\big)\rangle;\end{aligned}$$

thus, inserting this in (16),

$$\begin{aligned}&\frac{1}{2}\|\Phi(\sigma_\lambda^\delta)-\Lambda^\delta\|^2+\lambda\langle w,\Phi(\sigma_\lambda^\delta)-\Lambda^\delta\rangle+\frac{\lambda\rho}{2}\|\sigma_\lambda^\delta-\sigma^\dagger\|^2\\&\le\;\lambda\langle w,\Phi(\sigma_\lambda^\delta)-\Phi(\sigma^\dagger)-J_\Phi(\sigma^\dagger)(\sigma_\lambda^\delta-\sigma^\dagger)\rangle\\&\quad+\frac{1}{2}\|\eta\|^2+\lambda\langle w,\eta\rangle,\end{aligned}$$

and as a consequence of (13)

$$\frac{1}{2}\|\Phi(\sigma_\lambda^\delta)-\Lambda^\delta+\lambda w\|^2+\frac{\lambda\rho}{2}\|\sigma_\lambda^\delta-\sigma^\dagger\|^2\le\lambda\gamma\|w\|\|\sigma_\lambda^\delta-\sigma^\dagger\|^2+\|\eta+\lambda w\|^2.$$

Finally, neglecting the first term on the left-hand side, and leveraging the inequality $2\gamma\|w\|\le\rho$,

$$\lambda\|\sigma_\lambda^\delta-\sigma^\dagger\|^2\le\frac{4}{\rho-2\gamma\|w\|}\left(\delta^2+\lambda^2\|w\|^2\right) \tag{17}$$

and for the choice $\lambda\sim\delta$ we get that $\|\sigma_\lambda^\delta-\sigma^\dagger\|^2=O(\delta)$, hence the thesis. □

Remark 1 The functional Φ of EIT satisfies the assumptions of Proposition 1: indeed, the second statement of Theorem 1 implies that for every $\sigma\in K_{0,1}$ (and in particular for $\sigma^\dagger$) the Jacobian matrix $J_\Phi(\sigma)$ is injective, hence its transpose is surjective, and the source condition (12) is verified at any point $p^\dagger$. Condition (13) is moreover verified by Φ in view of the mean value inequality and of the Lipschitz continuity of the Jacobian, guaranteed again by Theorem 1.

5 Proximal-Gradient Method to solve (10)

We now consider a first-order iterative method to solve the minimization problem (10). To simplify its formulation, we introduce the following notation:

$$\begin{aligned} &\mathcal{J}_\lambda^\delta(\sigma) = f(\sigma) + \lambda g(\sigma), \\ &f(\sigma) = \frac{1}{2}\|\Phi(\sigma) - \Lambda^\delta\|^2 + \frac{\lambda\rho}{2}\|\sigma\|^2,\ g(\sigma) = R(\sigma) + \chi_K(\sigma). \end{aligned} \tag{18}$$

In this setup, f is a smooth functional, meaning that it is differentiable with Lipschitz continuous gradient, whereas g is a proper, convex, and continuous function, generally lacking differentiability. Therefore, a natural choice to approximate σ_λ^δ is to rely on a proximal-gradient, or forward-backward, iterative scheme. This accounts for constructing a sequence $\{\sigma^{(n)}\}$ as follows:

$$\sigma^{(n+1)} = \operatorname{prox}_{\mu\lambda g}\big(\sigma^{(n)} - \mu\nabla f(\sigma^{(n)})\big), \tag{19}$$

where $\mu > 0$ is a constant step-size. The convergence properties of such iterates, under suitable limitations on μ and various assumptions of f, g, are widely studied in the literature. For convex g and smooth f, provided that μ is smaller than the inverse of the Lipschitz constant of ∇f, [6, Theorem 1.3] proves that the cluster point of the iterates of (19) satisfies the (necessary) optimality conditions associated with $\mathcal{J}_\lambda^\delta$, also providing a convergence rate of an involved quantity.

In order to prove the convergence of the sequence of the iterates (to a minimizer of $\mathcal{J}_\lambda^\delta$) it is nevertheless necessary to require the convexity of f (see, e.g., [6, Theorem 1.2] or [4, Section 28.5]). Unfortunately, in our case this property is only guaranteed for large values of ρ: in particular, due to the nonlinearity of Φ, the term $\frac{1}{2}\|\Phi(\sigma) - \Lambda^\delta\|^2$ is not convex, even though it can be proved to be weakly convex. For the sake of completeness, we briefly discuss two results related to the convexity of similar functionals in the context of EIT. The first one, [36, Theorem 4.9], shows that, in the continuum model of EIT, the squared-norm mismatch functional is convex when restricted to a suitable subset of $\mathbb{R}^m$, but cannot guarantee that the iterates generated by an iterative scheme analogous to (19) belong to such a set. The second one, [33, Lemma 4.7], proves that, in the continuum model and in the presence of finitely many measurements, the forward operator F_m (substituting Φ in f) is convex with respect to the (semidefinite) Loewner order of semidefinite positive operators. Unfortunately, this does not translate into the convexity of the combination of such a functional with the squared Frobenius norm, which would eventually lead to a formulation analogous to our quadratic mismatch term; as an alternative, [33] relies on an algorithm based on the minimization of a linear functional on a suitable convex set.

An alternative approach to the study of the convergence of first-order methods relies on Kurdyka-Łojasiewicz (KL) conditions, which are less restrictive than convexity: see, for example, [3]. The results in [35, Section 3.2] provide the convergence rates for the function values and the difference of the iterates under weaker assumptions with respect to the ones we consider here (smoothness of f, weak-convexity of g), but the convergence of the iterates (to a stationary point of $\mathcal{J}_\lambda^\delta$, due to the lack of convexity) is ensured only if the KL condition is satisfied, see also [3, Theorem 5.1]. The verification of the KL conditions in the case of EIT is an open problem, and the local injectivity property analyzed in the current dissertation does not prove to be useful: by contrast, the surjectivity of the Jacobian J_Φ would be beneficial (see [42, Theorem 3.2]).

We prove a different result on the sequence $\sigma^{(n)}$, which takes advantage of the local injectivity of Φ. This is an adaptation of Theorem 2 in [8], which is not formulated for a proximal-gradient scheme but for Iterative Hard Thresholding (essentially equivalent to a projected gradient algorithm).

We start by substituting the definition of f in (19), obtaining the following expression of the iterates of PGM:

$$\sigma^{(n+1)} = \operatorname{prox}_{\mu\lambda g}\big(\sigma^{(n)} - \mu\lambda\rho\sigma^{(n)} + \mu J_\Phi(\sigma^{(n)})^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\big), \tag{20}$$

with $\mu > 0$, where $J_\Phi(\sigma)^T \in \mathbb{R}^{n\times m}$ denotes the transpose of the Jacobian matrix of Φ at $\sigma \in \mathbb{R}^n$. The following result characterizes the cluster point $\overline{\sigma}$ of the iterates produced by (20). In particular, it provides a bound for the distance between $\overline{\sigma}$ and any $\sigma \in K$ only in terms of the value of the functional $\mathcal{J}_\lambda$ evaluated in such a σ. This result can be effectively employed to discuss the asymptotic properties of $\overline{\sigma}$ as $\delta \to 0$, as discussed in Corollary 1 below, but may also help in motivating the selection of the space K, as observed in Corollary 2 below and further explored in Section 7. Notice that we do not need to specify the expression of the functional g, which will be the object of Section 6.

Theorem 2 *Let $J_\lambda^\delta(\sigma) = f(\sigma) + \lambda g(\sigma)$, being f as in (18) and g any non-negative, coercive, continuous, and convex functional, whose domain is contained in $K_{0,1}$. Assume that Φ satisfies:*

1. *Restricted Isometry Property: there exists $0 < \alpha \le \beta$ such that*

$$\alpha\|\sigma_1 - \sigma_2\|^2 \le \|J_\Phi(\sigma)(\sigma_1 - \sigma_2)\|^2 \le \beta\|\sigma_1 - \sigma_2\|^2 \quad \forall\sigma \in K_{0,1}, \quad \forall\sigma_1, \sigma_2 \in \mathbb{R}^n \tag{21}$$

2. *Mild non-linearity: there exists $\gamma > 0$ such that*

$$\|\Phi(\sigma_1) - \Phi(\sigma_2) - J_\Phi(\sigma_2)(\sigma_1 - \sigma_2)\|^2 \le \gamma\|\sigma_1 - \sigma_2\|^2 \qquad \forall\sigma_1, \sigma_2 \in K_{0,1}. \tag{22}$$

Require moreover that $\alpha, \beta, \gamma, \lambda, \rho$ and μ satisfy

$$\mu \le \frac{1}{2\beta}, \qquad \mu \le \frac{1}{2\lambda\rho}, \qquad 0 < \alpha + \lambda\rho - 2\gamma. \tag{23}$$

Then, the sequence $\sigma^{(n)}$ defined in (20) converges to a cluster point $\overline{\sigma}$ such that, for any $\sigma \in \operatorname{dom}(g)$

$$\|\sigma - \overline{\sigma}\|^2 \le \frac{4}{\alpha + \lambda\rho - 2\gamma}\mathcal{J}_\lambda(\sigma). \tag{24}$$

Proof We follow the technique adopted in the proof of [8, Theorem 2], with some modifications due to use of the scheme (20) instead of the Iterative Hard Thresholding, studied therein, and a different treatment of the bounds in (21), leading to different conditions that must be satisfied by the parameters. We start by proving the following claim:

$$\|\sigma^{(n+1)} - \sigma\|^2 \le (1 - \mu\lambda\rho - \mu\alpha + \mu\gamma)\|\sigma^{(n)} - \sigma\|^2 + 4\mu\mathcal{J}_\lambda^\delta(\sigma), \tag{25}$$

which holds for any $\sigma \in \operatorname{dom}(g)$. Indeed, let us first consider the definition of $\sigma^{(n+1)}$, which is given by (20) and can be rewritten as follows (to simplify the expressions, we denote by J the matrix $J_\Phi(\sigma^{(n)})$)

$$\sigma^{(n+1)} = \operatorname*{argmin}_{\sigma \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\sigma - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 + \mu\lambda g(\sigma) \right\};$$

therefore, due to the convexity of g, the minimizer $\sigma^{(n+1)}$ satisfies the following optimality conditions:

$$-(\sigma^{(n+1)} - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))) \in \mu\lambda\partial g(\sigma^{(n+1)}),$$

which translates into

$$\mu\lambda(g(\sigma) - g(\sigma^{(n+1)})) \;\geq -\langle \sigma^{(n+1)} - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)})), \sigma - \sigma^{(n+1)} \rangle \quad \forall \sigma \in \operatorname{dom}(g).$$

Using simple algebraic manipulations, we get

$$\begin{aligned} &\frac{1}{2}\|\sigma^{(n+1)} - \sigma\|^2 + \frac{1}{2}\|\sigma^{(n+1)} - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 \\ &\quad - \frac{1}{2}\|\sigma - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 \leq \mu\lambda g(\sigma) - \mu\lambda g(\sigma^{(n+1)}). \end{aligned}$$

Neglecting positive and negative terms on the left and right-hand side, respectively (using the non-negativity of g), we obtain

$$\|\sigma^{(n+1)} - \sigma\|^2 \leq \|\sigma - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 + 2\mu\lambda g(\sigma) \qquad (26)$$

Let us now focus on the first term on the right-hand side of (26):

$$\begin{aligned} &\|\sigma - \sigma^{(n)} + \mu\lambda\rho\sigma^{(n)} - \mu J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 \\ &\quad = \|\sigma - \sigma^{(n)}\|^2 + \mu^2\|\lambda\rho\sigma^{(n)} - J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 \\ &\qquad + 2\mu\lambda\rho\langle \sigma^{(n)}, \sigma - \sigma^{(n)} \rangle - 2\mu\langle J(\sigma - \sigma^{(n)}), \Lambda^\delta - \Phi(\sigma^{(n)}) \rangle \\ &\quad = \|\sigma - \sigma^{(n)}\|^2 + \mu^2 ① + \mu\lambda\rho ② + \mu ③ \end{aligned}$$

We furthermore observe that

$$\begin{aligned} ① &= \|\lambda\rho\sigma^{(n)} - J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 \leq 2\lambda^2\rho^2\|\sigma^{(n)}\|^2 + 2\|J^T(\Lambda^\delta - \Phi(\sigma^{(n)}))\|^2 \\ &\leq 2\lambda^2\rho^2\|\sigma^{(n)}\|^2 + 2\beta\|\Lambda^\delta - \Phi(\sigma^{(n)})\|^2, \end{aligned}$$

where used the upper bound in (21) since $\sigma^{(n)} \in \operatorname{dom}(g) \subset K_{0,1}$; instead,

$$② = 2\langle \sigma^{(n)}, \sigma - \sigma^{(n)} \rangle = \|\sigma\|^2 - \|\sigma^{(n)}\|^2 - \|\sigma - \sigma^{(n)}\|^2$$

and, analogously,

$$\begin{aligned} ③ &= -2\langle J(\sigma - \sigma^{(n)}), \Lambda^\delta - \Phi(\sigma^{(n)}) \rangle \\ &= \|\Lambda^\delta - \Phi(\sigma^{(n)}) - J(\sigma - \sigma^{(n)})\|^2 - \|J(\sigma - \sigma^{(n)})\|^2 - \|\Lambda^\delta - \Phi(\sigma^{(n)})\|^2 \\ &\leq \|\Lambda^\delta - \Phi(\sigma^{(n)}) - J(\sigma - \sigma^{(n)})\|^2 - \alpha\|\sigma - \sigma^{(n)}\|^2 - \|\Lambda^\delta - \Phi(\sigma^{(n)})\|^2, \end{aligned}$$

where we also used the lower bound in (21). The term $\Lambda^\delta - \Phi(\sigma^{(n)}) - J(\sigma - \sigma^{(n)})$ can be bounded as follows:

$$\begin{aligned} \|\Lambda^\delta - \Phi(\sigma^{(n)}) - J(\sigma - \sigma^{(n)})\|^2 &\leq 2\|\Lambda^\delta - \Phi(\sigma)\|^2 + 2\|\Phi(\sigma) - \Phi(\sigma^{(n)}) - J(\sigma - \sigma^{(n)})\|^2 \\ &\leq 2\|\Lambda^\delta - \Phi(\sigma)\|^2 + 2\gamma\|\sigma - \sigma^{(n)}\|^2, \end{aligned}$$

where we have used (22) with $\sigma_1 = \sigma$ and $\sigma_2 = \sigma^{(n)}$. Collecting all the results in (26), we get

$$\begin{aligned}
\|\sigma^{(n+1)} - \sigma\|^2 \leq& \|\sigma - \sigma^{(n)}\|^2 + 2\mu^2\big(\lambda^2\rho^2\|\sigma^{(n)}\|^2 + \beta\|\Lambda^\delta - \Phi(\sigma^{(n)})\|^2\big)\\
&+ \mu\lambda\rho\big(\|\sigma\|^2 - \|\sigma^{(n)}\|^2 - \|\sigma - \sigma^{(n)}\|^2\big)\\
&+ \mu\big(2\|\Lambda^\delta - \Phi(\sigma)\|^2 + 2\gamma\|\sigma - \sigma^{(n)}\|^2 - \alpha\|\sigma - \sigma^{(n)}\|^2 - \|\Lambda^\delta - \Phi(\sigma^{(n)})\|^2\big)\\
&+ 2\mu\lambda g(\sigma)\\
=&(1 - \mu\lambda\rho - \mu\alpha + 2\mu\gamma)\|\sigma^{(n)} - \sigma\|^2 + (2\mu^2\lambda^2\rho^2 - \mu\lambda\rho)\|\sigma^{(n)}\|^2\\
&+ (2\mu^2\beta - \mu)\|\Lambda^\delta - \Phi(\sigma^{(n)})\|^2 + \mu\lambda\rho\|\sigma\|^2 + 2\mu\|\Phi(\sigma) - \Lambda^\delta\|^2 + 2\mu\lambda g(\sigma).
\end{aligned}$$

Imposing that $2\mu^2\lambda^2\rho^2 - \mu\lambda\rho \leq 0$ (i.e., $\mu \leq \frac{1}{2\lambda\rho}$) and $2\mu^2\beta - \mu \leq 0$ (i.e., $\mu \leq \frac{1}{2\beta}$), we finally retrieve (25), where we used again the non-negativity of g. Let us now define $q = 1 - \mu\lambda\rho - \mu\alpha + 2\mu\gamma$: then, applying (25) recursively, we get

$$\|\sigma^{(n+1)} - \sigma\|^2 \leq q^n\|\sigma^{(0)} - \sigma\|^2 + 4\mu\mathcal{J}_\lambda^\delta(\sigma)\sum_{i=0}^{n} q^i.$$

This shows that, if we impose that $q < 1$ (which motivates the last bound in (23)), the sequence $\{\sigma^{(n)}\}$ is bounded and convergent to a cluster point $\overline{\sigma}$ such that

$$\|\overline{\sigma} - \sigma\|^2 \leq \frac{4\mu}{1-q}\mathcal{J}_\lambda^\delta(\sigma),$$

and substituting the expression of q we recover (24). □

Let us briefly discuss the assumptions introduced in Theorem 2.

1. The hypotheses on the functional f are verified when the operator Φ is the one of EIT, for a sufficiently large number of electrodes p and small discretization size h. Indeed, the local injectivity of $F'_{p,h}$ and the Lipschitz continuity and differentiability of $F_{p,h}$ on $V_{\mathscr{T}}$ in Theorem 1 ensure the lower and upper bounds in (21), whereas (22) is guaranteed by the Lipschitz continuity of $F'_{p,h}$.
2. The requirement $0 < \alpha + \lambda\rho - 2\gamma$ in (23) imposes an important connection between the ill-conditioning of J_Φ (expressed by α since, from (21), the smallest singular value of $J_\Phi(\sigma)$ is larger than α^2) and the non-linearity of Φ (encoded in γ). It is in general very difficult to check if the stronger condition $2\gamma < \alpha$ is verified in the case of EIT; nevertheless, we can leverage the presence of the regularization parameters $\lambda\rho$. Despite this being restrictive in an asymptotic scenario as $\lambda \to 0$, it is possible to verify that the requirements in (23) are less restrictive than the ones entailed by the original computations contained in [8], which would imply a condition like $\beta + \lambda\rho < \frac{3}{2}(\alpha + \lambda\rho) - 4\gamma$.

As a corollary of Proposition 1 and of Theorem 2, we deduce the following convergence rate for the error between the cluster point $\overline{\sigma}$ of (20) and the solution of the inverse problem.

Corollary 1 *Let $\overline{\sigma}$ be the cluster point of the iterates (20) and $\sigma^\dagger$ the solution of the inverse problem (1). Suppose the assumptions of Proposition 1 and of Theorem 2 are verified, and that moreover the last expression in (23) is replaced by $2\gamma < \alpha$. Then, for sufficiently small δ, under the choice $\lambda = C\delta$, there exists a constant c (depending on $\alpha, \gamma, \rho, \sigma^\dagger$ and C) such that*

$$\|\overline{\sigma} - \sigma^\dagger\| \leq c\sqrt{\delta} \tag{27}$$

Proof Substituting σ_λ^δ in (24) and leveraging the explicit expression of the convergence rate (17),

$$\begin{aligned}\|\overline{\sigma}-\sigma^\dagger\|^2 &\le 2\|\overline{\sigma}-\sigma_\lambda^\delta\|^2+2\|\sigma_\lambda^\delta-\sigma^\dagger\|^2\\ &\le \frac{8}{\alpha+\lambda\rho-2\gamma}\mathcal{J}_\lambda^\delta(\sigma_\lambda^\delta)+\frac{8}{\lambda(\rho-2\gamma\|w\|)}(\delta^2+\lambda^2\|w\|^2).\end{aligned}$$

Now, since by definition $\mathcal{J}_\lambda^\delta(\sigma_\lambda^\delta)\le\mathcal{J}_\lambda^\delta(\sigma^\dagger)$,

$$\|\overline{\sigma}-\sigma^\dagger\|^2\le\frac{8}{\alpha+\lambda\rho-2\gamma}\left(\frac{\delta^2}{2}+\frac{\lambda\rho}{2}\|\sigma^\dagger\|^2+\lambda g(\sigma^\dagger)\right)+\frac{8}{\lambda(\rho-2\gamma\|w\|)}(\delta^2+\lambda^2\|w\|^2).$$

Exploiting now the fact that $\alpha-2\gamma>0$ (and $\lambda\rho>0$), under the choice $\lambda=C\delta$, we recover the desired estimate for $\delta<1$:

$$\|\overline{\sigma}-\sigma^\dagger\|^2\le\left(\frac{4+4C\rho+8g(\sigma^\dagger)}{\alpha-2\gamma}+\frac{8(1+C^2\|w\|^2)}{C(\rho-2\gamma\|w\|)}\right)\delta. \tag{28}$$

□

The next corollary provides an important insight into the choice of the regularization functional g. Let us focus on the model expressed in (18), assuming that g is the sum of a non-negative, convex, and continuous functional R (whose domain is, without loss of generality, $\mathbb{R}^n$), and the characteristic function χ_K of the compact, convex set $K\subset K_{0,1}$. In particular, the following result discussed the importance of an educated choice of the set K, determining the domain of g.

Corollary 2 *Suppose the assumptions of Theorem 2 are verified, and that the last expression in (23) is replaced by* $2\gamma<\alpha$. *Let* σ_K *be the orthogonal projection of the exact solution* $\sigma^\dagger$ *onto* K, *i.e.,*

$$\sigma_K=\mathrm{proj}_K(\sigma^\dagger)=\operatorname*{argmin}_{\sigma\in K}\left\{\|\sigma^\dagger-\sigma\|\right\}. \tag{29}$$

Then, for sufficiently small δ, *under the choice* $\lambda=C\delta$, *there exist two constants* c_1,c_2 *(depending on* α,γ,ρ,g *and* C*) such that*

$$\|\overline{\sigma}-\sigma^\dagger\|\le c_1\sqrt{\delta}+c_2\|\sigma_K-\sigma^\dagger\|. \tag{30}$$

Proof We employ the inequality (24) replacing σ by σ_K. Then, it holds

$$\begin{aligned}\|\overline{\sigma}-\sigma^\dagger\|^2 &\le 2\|\overline{\sigma}-\sigma_K\|^2+2\|\sigma_K-\sigma^\dagger\|^2\\ &\le\frac{8}{\alpha+\lambda\rho-2\gamma}\mathcal{J}_\lambda^\delta(\sigma_K)+2\|\sigma_K-\sigma^\dagger\|^2\\ &\le\frac{8}{\alpha+\lambda\rho-2\gamma}\left(\delta^2+\|\Phi(\sigma_K)-\Phi(\sigma^\dagger)\|^2+\frac{\lambda\rho}{2}\|\sigma_K\|^2+\lambda g(\sigma_K)\right)\\ &\quad+2\|\sigma_K-\sigma^\dagger\|^2\end{aligned}$$

We now recall that the map Φ is Lipschitz-continuous with a constant smaller than β, due to (21) and to the mean-value inequality. Moreover, since K is contained within the compact set $K_{0,1}$, the quantities $\|\sigma_K\|^2$ and $g(\sigma_K)=R(\sigma_K)$ can be bounded by the constants Γ_1 and Γ_2 respectively, both independent of K. Thus, for $\lambda=C\delta$ and $\delta<1$, also employing that $\alpha-2\gamma>0$, we get

$$\|\overline{\sigma} - \sigma^\dagger\|^2 \leq \frac{8 + 4C\rho\Gamma_1 + \Gamma_2}{\alpha - 2\gamma}\,\delta + \left(\frac{8\beta}{\alpha - 2\gamma} + 2\right)\|\sigma_K - \sigma^\dagger\|^2. \tag{31}$$

□

When comparing (27) and (30), one observes that, although the latter bound does not guarantee the asymptotic accuracy of $\overline{\sigma}$ as $\delta \to 0$, it clearly outlines its dependence on the choice of K. In particular, for a fixed δ, (30) shows that the error might be reduced if K is chosen such that $\|\sigma_K - \sigma^\dagger\|$ is extremely small, or even 0.

If the solution $\sigma^\dagger$ is known to be $s-$sparse (i.e., to have at most s non-vanishing components), one possible way to leverage this information (see e.g. [8]) is by considering K as the union of all the $s-$dimensional coordinate hyperplanes in $\mathbb{R}^n$. The projection onto such a space can be easily computed (by selecting the s largest components of a vector, as in the Iterative Hard Thresholding algorithm in [8]), but an overestimation of the sparsity level s may still lead to an inefficient reconstruction.

In our case, we wish to select K in a more insightful way. In particular, suppose that an *Oracle* function is available, taking as an input the measurements Λ^δ and returning the exact support of $\sigma^\dagger$. Employing such an output as the set K in algorithm (20) would allow canceling the last term on the right-hand side of (30), thus obtaining better estimates. The quest for such an Oracle functional can also be interpreted as the estimation of the *support*, or the *sparsity pattern* of the solution, a problem that has been extensively studied in the context of compressed sensing (see, e.g., [25, 45, 56]). Our approach, discussed in Section 7, relies instead on statistical learning techniques, and in particular based on Graph Neural Networks, to provide a data-driven approximation of the optimal Oracle functional.

6 Some explicit regularization models

We now want to study more closely some choices of R and K in (10), which will also reflect in more explicit expressions for the proximal operator of g appearing in (20) and defined in (18).

In particular, we consider the following possible expressions for R:

(1) Sparsity-promoting regularization through ℓ^1 norm:

$$R(\sigma) = \|\sigma - \sigma_0\|_1, \qquad \sigma_0 \in [c_0, c_1] \subset \mathbb{R}^n. \tag{32}$$

This reflects the assumption that σ differs from a known reference conductivity σ_0 only in a few components. The proximal map in this case reads as follows:

$$\sigma^* = \mathrm{prox}_{\lambda\mu R}(\sigma) = \sigma_0 + S_{\lambda\mu}(\sigma - \sigma_0),$$

where S_t is the element-wise soft-thresholding function, namely $S_t\colon \mathbb{R}^n \to \mathbb{R}^n$ such that, for $i = 1, \ldots, n$,

$$[S_t(v)]_i = \mathrm{sign}(v_i)\max(0, |v_i| - t).$$

(2) Anisotropic Total Variation (TV) on meshes (see [17]):

$$R(\sigma) = \mathrm{TV}(\sigma) = \sum_{i=1}^{n}\sum_{k\in\mathcal{N}_i} w_{ik}|\sigma_i - \sigma_k|, \tag{33}$$

where, for each i, $\mathcal{N}_i$ denotes the set of indices $k \in \{1, \ldots, \deg(i)\}$, with $\deg(i)$ denoting the valence of vertex ith. The positive weight w_{ik} is instead defined as the inverse of the

(Euclidean) distance between the i-th vertex and its k-th adjacent vertex. The proximal map associated with this choice

$$\sigma^* = \mathrm{prox}_{\lambda\mu R}(\sigma) = \arg\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|x-\sigma\|_2^2 + \lambda\mu TV(x) \right\} \tag{34}$$

can be approximately computed according to a procedure proposed in [43] and generalized in [17] for polygonal meshes. Let σ^* be the fixed point of the following equations, stated on its components

$$\sigma_i^* = \operatorname*{argmin}_{x\in\mathbb{R}} \left\{ \frac{1}{2}(x-\sigma_i)^2 + \lambda\mu \sum_{k\in\mathcal{N}_i} w_{ik}|x-\sigma_k^*| \right\}. \tag{35}$$

Following [43, Theorem 3.2 and Remark 3.1], the solution of problem (35) can be obtained as the limit of the sub-iterations

$$\sigma_i^{(l+1)} = \operatorname*{argmin}_{x\in\mathbb{R}} \left\{ \frac{1}{2}(x-\sigma_i)^2 + \lambda\mu \sum_{k\in\mathcal{N}_i} w_{ik}|x-\sigma_k^{(l)}| \right\}, \tag{36}$$

and for each i and l the unique solution of (36) can be efficiently computed by the median formula proposed in [43]

$$\sigma_i^{(l+1)} = \mathrm{median}\left\{\sigma_1^{(l)}, \ldots, \sigma_n^{(l)}, \sigma_i^{(l)} + \lambda\mu W_0, \sigma_i^{(l)} + \lambda\mu W_1 \ldots, \sigma_i^{(l)} + \lambda\mu W_n\right\}, \tag{37}$$

where the values σ_k of the vertices $k \in \mathcal{N}_i$ are sorted in increasing order, and

$$W_j = -\sum_{k=1}^{j} w_k + \sum_{k=j+1}^{n} w_k, \qquad j = 0, .., n. \tag{38}$$

A discussion on the convergence to the global anisotropic TV problem (34) by iterating the local optimization problem (35), is provided in [43].

Regarding K we consider the following cases:

(i) Box constraint in $\mathbb{R}^n$: given $0 < c_0 < c_1$,

$$K = [c_0, c_1]^n, \qquad \chi_K(\sigma) = \prod_{i=1}^{n} \chi_{[c_0,c_1](\sigma_i)}. \tag{39}$$

This restriction is mandatory in order to ensure that the conductivity σ is non-vanishing and bounded away from 0, hence the EIT problem is well-defined.

(ii) Oracle-based projection:

$$K = [c_0, c_1]^n \cap \Pi_{\mathcal{O}}, \tag{40}$$

being $\Pi_{\mathcal{O}}$ a coordinate hyperplane conducted through a reference point σ_0, i.e., of the form:

$$\Pi_{\mathcal{O}} = \left\{x \in \mathbb{R}^n : \quad x_i - \sigma_{0,i} = 0 \;\; \forall i \in \{1, \ldots, n\} \setminus I_{\mathcal{O}}\right\}, \tag{41}$$

where $I_{\mathcal{O}}$ denotes the set of active coordinates. The knowledge of $\Pi_{\mathcal{O}}$ (i.e., of the expected support of $\sigma - \sigma_0$) can be seen as the outcome of a separate support estimation problem, and acts as an Oracle for the PGM scheme (20). The projection onto $\Pi_{\mathcal{O}}$ can be performed as follows:

$$\mathrm{proj}_{\Pi_{\mathcal{O}}}(\sigma) = \sigma_0 + M_{\mathcal{O}} \odot (\sigma - \sigma_0),$$

where $\odot$ denotes the element-wise product of vectors and $M_{\mathcal{O}} \in \mathbb{R}^n$ is the *mask* associated with the Oracle $\mathcal{O}$, namely, a vector such that $[M_{\mathcal{O}}]_i = 1$ if $i \in I_{\mathcal{O}}$ and 0 otherwise.

The following result significantly simplifies the computation of $\mathrm{prox}_{\lambda\mu g}$ in (20) for our proposed choices of R and K.

Proposition 2 (Proximal map of g) *Let g be a proper, convex and continuous function defined as*

$$g(\sigma) = R(\sigma) + \chi_K(\sigma), \tag{42}$$

where R is a non-negative, proper, continuous, convex functional, and $K \subset K_{0,1} = [c_0, c_1]^n$ is a convex and closed set in $\mathbb{R}^n$. If R is chosen as in (32) or (33) and K as in (39) or (40), then the proximal of g is summative, $\mathrm{prox}_g = \mathrm{prox}_{\chi_K} \circ \mathrm{prox}_R$, *and, in particular, it satisfies*

$$\mathrm{prox}_{\lambda\mu g} = \mathrm{proj}_K \circ \mathrm{prox}_{\lambda\mu R}\,. \tag{43}$$

Proof Both the choices of K lead to characteristic functions $\chi_K(\sigma)$ which can be written as the product of n functions, each in one component σ_i. This is also the case for R in (32). Hence, in these cases, (43) follows by [15, Proposition 2.2]. Instead, when R is chosen as in (33), both R and K fit the hypotheses of [50, Proposition II.2], which entails (43). □

The previous result can be extended also to other separable functionals R, as the one in (32), which can be decomposed with respect to the components of σ, i.e., that can be written as $R(\sigma) = \sum_{i=1}^{n} \psi_i(\sigma_i)$, being $\psi_i : \mathbb{R} \to \mathbb{R}$ non-negative, continuous and convex. This class includes, for example, the choices $R(\sigma) = \|\sigma\|_p^p$ for $p \geq 1$. Moreover, the theoretical treatment of case (33) can be extended to functionals that have different definitions of the weights w_{ik}, also replacing the absolute value $|\sigma_i - \sigma_k|$ by the difference $\sigma_i - \sigma_k$. This includes, for example, wavelet operators defined on meshes. Moreover, it can be extended to the case $R(\sigma) = \mathrm{TV}(\sigma - \sigma_0)$, for $\sigma_0 \in [c_0, c_1]^n$.

To summarize, our sparsity-inducing proposal to solve the inverse EIT reconstruction problem is based on the solution of the optimization problem (10) with $g(\sigma)$ defined as in (42), which reads as

$$\sigma_\lambda^\delta \in \underset{\sigma \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \mathcal{J}_\lambda^\delta(\sigma) := \frac{1}{2}\|\Phi(\sigma) - \Lambda^\delta\|^2 + \frac{\lambda\rho}{2}\|\sigma\|^2 + \lambda R(\sigma) + \chi_K(\sigma) \right\}. \tag{44}$$

To this aim we propose the Proximal-Gradient method defined in (20) which is characterized by four different algorithms according to the choice of R and K in (44), which consequently affects the solution of the $\mathrm{prox}_{\lambda\mu g}$ problem with g defined in (42). In particular, by considering the four different combinations of R and K, we achieve the following algorithms:

PGM-ℓ^1	: R is chosen as in (32) and K as in (39), promoting sparsity of the difference from the reference σ_0 and imposing the state constraints on σ;
PGM-TV	: R is chosen as in (33) and K as in (39), promoting sparsity of gradient of σ and imposing the state constraints on σ;
PGM-ℓ^1-$M_{\mathcal{O}}$	: R is chosen as in (32) and K as in (40), employing a pre-trained Oracle to select the support of $\sigma - \sigma_0$;

PGM-TV-$M_{\mathcal{O}}$: R is chosen as in (33), and K as in (40), employing a pre-trained Oracle to select the support of $\sigma - \sigma_0$.

The last two algorithms require the preliminary estimate of the support by means of the pre-trained Oracle, procedure that will be described in the next section.

7 Support Estimation via Oracle-Net

We devised an Oracle-based strategy to predict the expected support $I_{\mathcal{O}}$ of a conductivity distribution σ starting from a set of measurements Λ^{δ}. The support $I_{\mathcal{O}}$ allows to classify σ on the vertices of the mesh domain $\mathscr{T}_h$ as belonging to an inclusion, when $\sigma_i \neq \sigma_{0,i}$, or belonging to the background σ_0. The Oracle $\mathcal{O}$ produces a binary mask $M_{\mathcal{O}}$ on the vertices of the mesh $\mathscr{T}_h$, with $[M_{\mathcal{O}}]_i = 1$ for vertex v_i if $i \in I_{\mathcal{O}}$, and 0 otherwise.

We employed a Graph-U-Net architecture, termed Oracle-Net, to model $\mathcal{O}$. This architecture, inspired by U-Net, is tailored for graph/mesh data as described in [18, 24]. The Graph-U-Net is an autoencoder architecture based on convolutional graph operators and gPool and gUnpool operators. The pool (gPool) operator samples some nodes to form a coarser mesh while the unpool (gUnpool) operator performs the inverse process, by increasing the number of nodes exploiting the list of node locations selected in the corresponding gPool, see [26] for details.

The architecture of the Oracle-Net neural network is depicted in Fig. 1. Each layer T_ℓ is characterized by the composition of a GCN-based graph convolution [38], a ReLU activation function s, and a gPool/gUnpool operator, here denoted by a generic p. T_ℓ applies to the input feature array $X \in \mathbb{R}^{n \times n_c}$, namely

$$T_\ell : X \mapsto s(GCN(p(X); \Theta_\ell)), \tag{45}$$

where $\Theta_\ell \in \mathbb{R}^{n_c \times n_f}$ denotes the trainable weight matrix of layer ℓ, and n_c, n_f are, respectively, the number of input, output features of the layer T_ℓ. The Oracle-Net with $2P$ layers performs the following composite function

$$\mathcal{O}(X) := \underbrace{T_{2P}^{gU} \circ \cdots \circ T_{P+1}^{gU}}_{decoder} \circ \underbrace{T_{P}^{gP} \circ \cdots \circ T_{1}^{gP}}_{encoder}(X), \tag{46}$$

where T_ℓ^{gP}, for $\ell = 1, \ldots, P$ applies a gPool operation to the feature vector, while T_ℓ^{gU}, for $\ell = P+1, \ldots, 2P$ applies a gUnpool operation. The last decoder layer uses a sigmoid activation function s in (45) which returns the probability that v belongs to the support. A binary mask $M_{\mathcal{O}}$ in $\{0, 1\}^n$ is generated by applying a threshold of value σ_{th} during the inference phase. We finally remark that the technicality of skip-connections has been inherited in Oracle-Net from Graph-U-Net, although not evident from the abstract model reported in (46).

Oracle-Net processes a mesh $\mathscr{T}_h$ (identical for the discretization of the variable σ and for the numerical solution of the differential problem, hence $\mathscr{T} = \mathscr{T}_h$) and its adjacency matrix as input. Additionally, for each vertex $v \in \mathscr{T}_h$, it receives a weight vector $w \in \mathbb{R}^m$ that modulates the impact of potential measurements on v. According to the selected measurement protocol the weight distribution w changes, as illustrated in Fig.2(a) for adjacent injection, adjacent measurement, $(I_{[E_k,E_{k+1}]}, V_{[E_j,E_{j+1}]})$, and in Fig.2(b) for opposite injection, adjacent measurement, $(I_{[E_k,E_{k+p/2}]}, V_{[E_j,E_{j+1}]})$. Specifically, each vector component w_i corresponding to an injection-measurement pair quantifies the measurement's influence

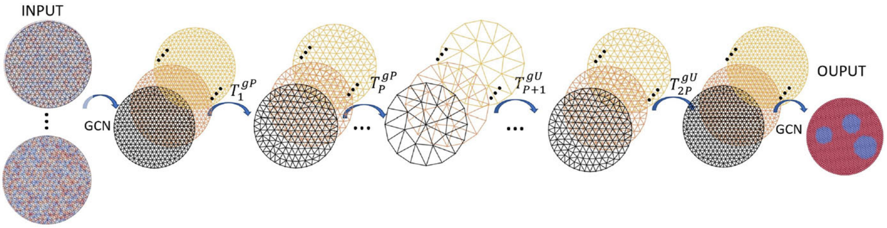
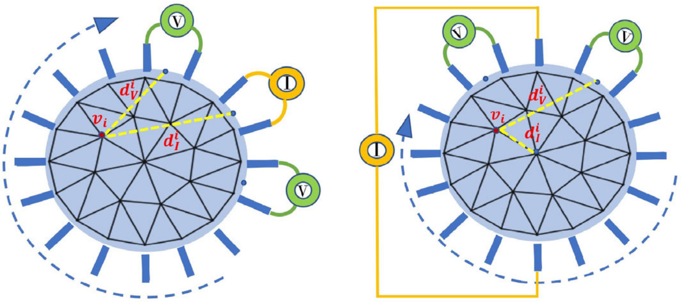


Fig. 1 Oracle-Net architecture



(a) (b)

Fig. 2 Oracle-Net, weight vector setup: (a) adjacent injection-adjacent measurements; (b) opposite injection-adjacent measurements

on mesh vertex v_i and is calculated as follows:

$$w_i = \frac{1}{d_V^i + d_I^i} V_{[E_j, E_{j+1}]}, \tag{47}$$

where d_V^i is the distance from the vertex v_i to the midpoint between the pairs of electrodes involved in the measurements $[E_j, E_{j+1}]$, and d_I^i is the distance between the vertex v_i and the midpoint between the pairs of electrodes used for the injection, $[E_k, E_{k+1}]$ or $[E_k, E_{k+p/2}]$.

Oracle-Net is a supervised GNN which in the training phase takes in input the exact support mask $M_{\mathcal{O}}$ for each training sample j for $j = 1, \ldots, N$, and computes the optimal parameters Θ by minimizing the joint loss function:

$$\mathcal{L}(\Theta) := \sum_{j=1}^{N} \Big(\sum_i BCE(p_i(\Theta), [M_{\mathcal{O}}]_i^j) + \sum_i MSE(p_i(\Theta), [M_{\mathcal{O}}]_i^j) \Big) \tag{48}$$

where $BCE(q_1, q_2)$ is binary cross-entropy between the target GT (q_2) and the input probability (q_1) in one single class, and $MSE(q_1, q_2)$ applying an L_2 norm to minimize the difference between the ground truth q_2 and the probability q_1 obtained by the graph network at vertex v_i.

8 Numerical Results

This section presents numerical results that illustrate the performance of the proposed Oracle-Net-based Proximal Gradient Method for the nonlinear inverse EIT reconstruction problem. Section 8.1 presents a validation of Oracle-Net for support estimation. Section 8.2 demonstrates reconstruction performance on a 2D dataset, considering both noise-free and noisy conditions. Finally, Section 8.3 explores the relationship between the number of measurements and Oracle mask sparsity, an area well-understood in linear compressed sensing but still open for nonlinear problems like inverse EIT.

The data used for the experimental session consists of a 2D synthetic EIT conductivities generated on a mesh $\mathscr{T}_h$ composed of $n = 1602$ vertices and $N_h = 3073$ triangles. We will refer to them as Ground Truth (GT). All examples simulate a circular tank slice of unitary radius. In the circular boundary ring, $p = 32$ equally spaced electrodes are located. The conductivity of the background material is set to be $\sigma_0 = 1.0\ \Omega m^{-1}$. Each sample consists of a random number from 1 to 4 of inclusions inside the circular tank, localized randomly and characterized by random radius in the range $[0.15, 0.25]$ and magnitude in the range $[0.2, 2]$. Each inclusion consists of a homogeneous material with the same conductivity intensity. The region covered by the inclusions is significantly less than the total tank area, this leads to the sparsity both in the solution vector $\sigma - \sigma_0$, and in its gradient.

The acquisition of m measurements is simulated through adjacent injection - adjacent measurement protocol for the results reported in Section 8.1, and opposite injection - adjacent measurement protocol for the experiments illustrated in Sections 8.2 and 8.3. In all the examples the setup is considered blind, that is no a priori information about the sizes or locations of the inclusions is considered. In the forward calculations of $\Phi(\sigma)$ in (10) we applied the KTCFwd forward solver, a two-dimensional version of the FEM described in [55], kindly provided by the authors (website https://github.com/CUQI-DTU/KTC2023-CUQI4), which is based on a FEM implementation of the CEM model on triangle elements. The electric potential is discretized using second-order polynomial basis functions, while the conductivity is discretized on the nodes using linear basis functions on triangle elements.

8.1 Oracle-Net Validation

For the training of the Oracle-Net, we used an ad hoc designed dataset which consists of 5000 instances Λ^δ, each of dimension $m = 992$. A portion of the dataset which is the 70% of the total number is used for training, the 15% is used as a validation set and the remaining 15% is employed as the test set for performance assessment. Training of Oracle-Net has been performed with ADAM optimizer, [37], using a learning rate equal to 2.5×10^{-3} through 2034 epochs with a mini-batch size of 10 instances.

As a figure of merit for assessing the performance of the Oracle-Net, we used the False Negative (FN) value, which represents the percentage on the test sample of misclassified vertices that corresponds to the vertices in the domain $\mathscr{T}_h$ belonging to inclusions which are misclassified as background. For the success of the reconstruction algorithm, this represents the major drawback produced by the Oracle-Net since the misclassified vertices will not be carried out by the reconstruction algorithm.

In Fig.3 we report the ideal Oracle mask (directly obtained by the ground truth conductivity distribution) superimposed to the mask $M_{\mathcal{O}}$ obtained by the Oracle-Net for six different samples. For increasing threshold values $\sigma_{th} = \{0.4, 0.8, 0.9\}$ the support estimated by the Oracle-Net enlarges, while the FN measure, reported in the bottom of each mask, decreases. In

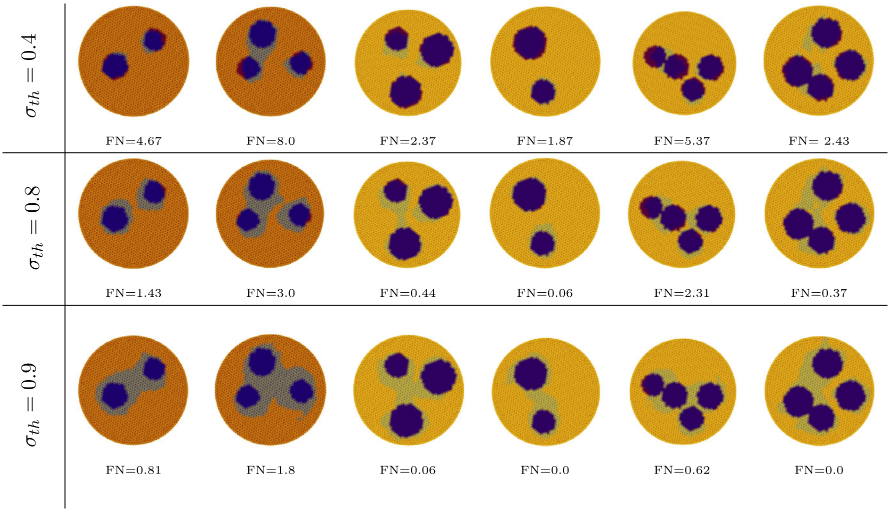


Fig. 3 Ideal Oracle (solid) superimposed to the learned Oracle mask $M_{\mathcal{O}}$ (in transparency). Top to bottom: three different thresholds σ_{th} (rows) applied to six different samples (columns)

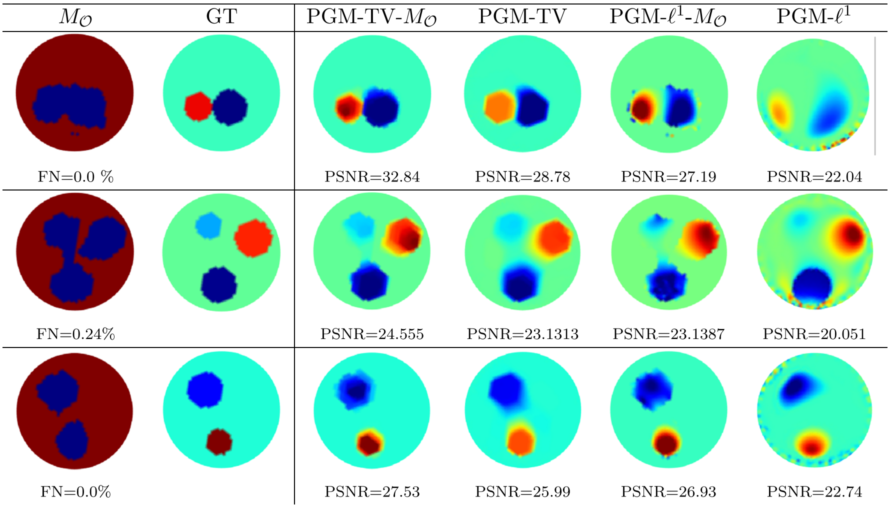


Fig. 4 Noise-free datasets, reconstructions with the different algorithms applying opposite-adjacent protocol and $\sigma_{th} = 0.8$ for the threshold of the Oracle-Net result

Fig. 3, in the first two columns the samples are obtained by an adjacent-adjacent protocol (see Fig.2(a)), while for the results in the other columns an opposite-adjacent protocol has been applied (see Fig.2(b)). The protocol indeed has no impact on the quality of the resulting masks. We finally remark that Oracle-Net generates the mask $M_{\mathcal{O}}$ in real-time during inference.

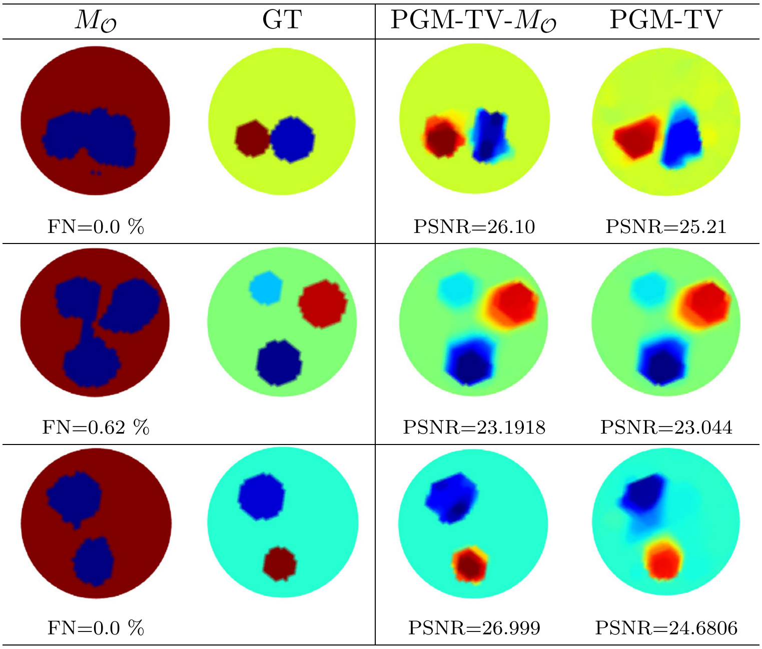


Fig. 5 Noisy dataset, reconstructions with the different algorithms applying opposite-adjacent protocol and $\sigma_{th} = 0.8$ for the threshold of the Oracle-Net result

8.2 EIT Reconstruction Results

The inference phase of the Oracle-Net allows us to determine the estimated support mask $M_{\mathcal{O}}$ for a given set of new measurements (unseen from the Oracle). Then the conductivity σ is computed by applying the PGM algorithm.

In order to investigate the influence on the performance of the Oracle-Net in the EIT reconstruction problem (10), we compared the results of PGM-TV-$M_{\mathcal{O}}$ and PGM-ℓ^1-$M_{\mathcal{O}}$, which exploit the Oracle-Net estimated support, with the standard PGM algorithms PGM-TV and PGM-ℓ^1, analogously regularized. In all the reported experiments, we selected the optimal λ value through trials and errors, while the ρ value was fixed to be 10^{-12}. All reconstruction algorithms terminate when either the maximum number of iterations (1000) is reached or the relative change in the iterative solution values falls below a tolerance of 10^{-6}.

First, we consider the case where the measurement error η in (1) vanishes.

In Fig.4 we show some sample conductivity reconstructions and we report below the corresponding Peak-SNR (PSNR), used to measure the quality of the reconstruction. For each sample, the first column represents the estimated support $M_{\mathcal{O}}$ pre-computed by the Oracle-Net together with the FN value associated, the second column illustrates the target GT reconstruction, and from the third to the sixth columns the computed reconstructions achieved by the different algorithms are shown, with corresponding PSNR. From a visual inspection of the illustrated results and from the PSNR values reported, we can observe the benefit of incorporating the Oracle-Net estimated support into the PGM algorithmic framework. PGM-TV and PGM-ℓ^1 achieve lower performance than their Oracle-Net-based counterparts. Moreover, the use of the TV regularizer seems to be beneficial with or without Oracle exploitation. We

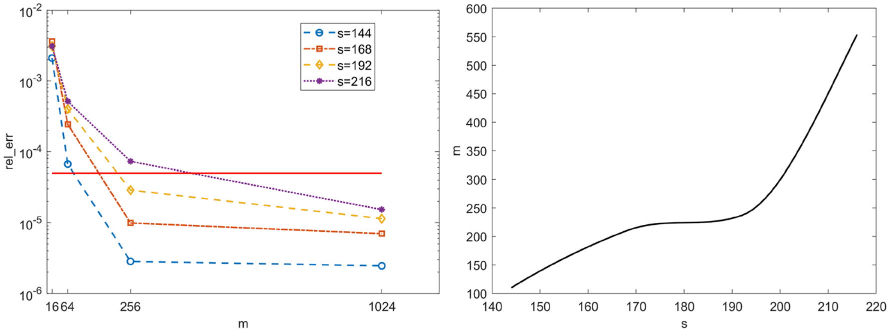


Fig. 6 Left - Plot of the relative errors in the recovery of four samples σ with increasing gradient s-sparsity obtained by PGM-TV-$M_{\mathcal{O}}$ with an increasing number of measurements m. Right - Plot of the empirical relation between the number of measurements m and the sparsity factor s

remark that the benefits of using the estimated support $M_{\mathcal{O}}$ produced by Oracle-Net are not limited to the qualitative aspect, but also concern computational efficiency. Reconstruction methods utilizing the $M_{\mathcal{O}}$ mask converge at twice the speed of their counterparts without it, as demonstrated by average performance across multiple experiments.

We finally evaluate the robustness against measurements corrupted by additive noise. The measured voltage is computed by the forward model KTCFwd and recorded as a vector $V_m = \Phi(\sigma) \in \mathbb{R}^m$. According to the degradation model (1), the noiseless measures V_m are corrupted by additive white Gaussian noise to simulate experimentally measured voltages, with noise level $\bar{\delta}$:

$$\eta = \bar{\delta}\ \|V_m\|_2\ \bar{n}, \quad \bar{n} \in \mathcal{N}(0, 1). \tag{49}$$

The Signal to Noise Ratio (SNR) in dB is calculated as $SNR = 10\log_{10}(\frac{\|V_m\|_2^2}{\|\eta\|_2^2})$.

In Fig.5 some examples of conductivity reconstructions are illustrated. The collected voltage has been corrupted by adding a realization of random noise with Gaussian distribution with noise level $\bar{\delta} = 2.5 \times 10^{-3}$ (added noise is such that the intrinsic SNR is 40dB). We observe that, even though the Oracle-Net was trained on noiseless data, the mask $M_{\mathcal{O}}$ obtained in inference with noise measurements - see Fig.5 first column - is sufficiently accurate, and PGM algorithm performs very well also with noise measurements. A natural consequence of having noise-corrupted measurements is the degradation in the conductivity reconstruction. A quantitative comparison of the third columns in Fig.4 and Fig.5 reveals a degradation in reconstruction performance as measured by PSNR values. Similarly, the fourth columns of these figures, which do not incorporate the Oracle Mask, exhibit analogous behavior.

8.3 Experimental Compressed Sensing Issues

The quality of the reconstructed conductivity σ is strictly related to the quantity m and the quality of the acquired measurements V_M. In this simplified context we are aware of neglecting many other factors that could affect the reconstruction accuracy, i.e., the mismodeling of the domain, or the misplacement of electrodes. However, it is clear that obtaining high-quality EIT reconstructions with a reduced number of electrodes (and thus of measurements) would be of great help to reduce the costs and increase the reliability of EIT systems in practical applications. It would be even more useful to know the sufficient number of measurements needed to recover an optimal σ, under sparsity conditions on the unknown conductivity.

A well-known result in well-posed linear CS asserts that when $\sigma \in \mathbb{R}^n$ is s-sparse, the recovery via ℓ^1-minimization is provably exact, using at least a number of measurements m roughly of order $O(s \log n)$. To the best of our knowledge, a similar result has not been achieved yet in the case of nonlinear measurements. To conclude this numerical session we would like to investigate, at least from an experimental point of view, the relation between the s-sparsity factor, which characterizes the conductivity σ to be reconstructed, and the number of nonlinear measurements m, needed to obtain an optimal recovery.

At this aim we applied PGM-TV-$M_{\mathcal{O}}$ algorithm to the reconstruction of four conductivity samples with different levels of sparsity of the gradient, using an increased number m of measurements, randomly chosen among the total available measures. In the idealized set-up, each sample is noise-free and contains a single circular inclusion, with a different radius; for increasing radius values the sample represents a decreasing s-sparsity value, characterized by $s := ||\nabla\sigma||_0 = \{144, 168, 192, 216\}$. In Fig. 6, left panel, we report the average of the relative ℓ^2 reconstruction errors $rel\_err := \|\sigma - \sigma^*\|_2/\|\sigma\|_2$ obtained for each of the four sample, running 100 times the EIT reconstruction algorithm. The number of measurements m used for solving the inverse EIT reconstruction problem varies in the range $m = \{16, 64, 256, 1024\}$.

The horizontal solid red line indicates the error threshold below which the reconstruction can be considered optimal (with relative error $rel\_err = 5 \times 10^{-5}$, which corresponds to PSNR=90). As expected, when the gradient-sparsity decreases, which corresponds to increasing values of s, the number of measurements necessary to obtain an optimal reconstruction increases. Finally, it should be emphasized that, for samples with a more severe s-sparsity, beyond a certain threshold, the improvement in reconstruction, in the face of an increase in the number of measurements (from $m = 256$ to $m = 1024$), is no longer significant.

The plot in Fig.6, right panel, represents the empirical relation between the number of measurements m and the sparsity factor s, obtained by interpolating the points in the plot - Fig. 6(left panel) - which intersect the optimal recovery line (in solid red). The employed CS framework enables efficient reconstruction of conductivity distributions from a significantly reduced number of measurements compared to the data dimensionality. Consequently, the plot in Fig.6, right panel, aids in determining the minimum measurement count required for optimal conductivity recovery σ, assuming sparsity conditions on the unknown conductivity.

9 Conclusions

This paper demonstrated a proof of concept study in using CS techniques in the numerical solution of nonlinear ill-posed inverse problems. We proposed a sparsity-aware PGM for the solution of a variational formulation of the EIT inverse problem. The sparsity inducing role is taken by a new concept of “Oracle” which infers the optimal support for a given set of nonlinear measurements. By exploiting the sparsity or compressibility of the signal distribution, CS reduces the amount of data needed for accurate reconstruction. The Oracle is designed by an autoencoder GNN that automatically predicts a binary mask which localizes the inclusions thus reducing memory requirements and processing time while maintaining recovery accuracy. The accurate recovery is demonstrated, using the proposed sparsity-aware PGM algorithm, under the requirements that the Jacobian of the measurement system Φ satisfies a RIP-like condition and that Φ is mildly nonlinear. Moreover, we shed light on the problem of determining how few measurements suffice for an accurate EIT sparsity-regularized reconstruction, a well-known result in well-posed linear CS. An interesting future direction will address this issue. Finally, we will consider other interesting nonlinear CS contexts where the proposed Oracle-based strategy can be successfully applied.

Acknowledgements This work was supported in part by the National Group for Scientific Computation (GNCS-INDAM), Research Projects 2024, and in part by MIUR RFO projects. The research of LR has been funded by PNRR - M4C2 - Investimento 1.3. Partenariato Esteso PE00000013 - “FAIR - Future Artificial Intelligence Research” - Spoke 8 “Pervasive AI”, which is funded by the European Commission under the NextGeneration EU programme. LR also acknowledges the support of “Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni” of the “Istituto Nazionale di Alta Matematica” through project GNAMPA-INdAM 2023, code CUP_E53C22001930001. The work of SM and DL was supported by PRIN2022_MORIGI, titled "Inverse Problems in the Imaging Sciences (IPIS)" 2022 ANC8HL - CUP J53D23003670006, and PRIN2022_PNRR_CRESCENTINI CUP J53D23014080001.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Data Availability Data will be made available on request.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Alberti, G. S., Felisi, A., Santacesaria, M., Trapasso, S. I.: Compressed sensing for inverse problems and the sample complexity of the sparse radon transform. arXiv preprint arXiv:2302.03577, (2023)
2. Alberti, G.S., Santacesaria, M.: Infinite dimensional compressed sensing from anisotropic measurements and applications to inverse problems in pde. Applied and Computational Harmonic Analysis **50**, 105–146 (2021)
3. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. Mathematical Programming **137**(1–2), 91–129 (2013)
4. Bauschke, H.H., Combettes, P.L., Bauschke, H.H., Combettes, P.L.: Correction to: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer (2017)
5. Beck, A., Eldar, Y.C.: Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. SIAM Journal on Optimization **23**(3), 1480–1509 (2013)
6. Beck, A., Teboulle, M.: Gradient-based algorithms with applications to signal-recovery problems, page 42-88. Cambridge University Press, (2009)
7. Benning, M., Burger, M.: Modern regularization methods for inverse problems. Acta numerica **27**, 1–111 (2018)
8. Blumensath, T.: Compressed sensing with nonlinear observations and related nonlinear optimization problems. IEEE Transactions on Information Theory **59**(6), 3466–3474 (2013)
9. Blumensath, T., Davies, M. E.: Gradient pursuit for non-linear sparse signal modelling. In 2008 16th European Signal Processing Conference, pages 1–5, (2008)
10. Blumensath, T., Davies, M.E.: Iterative hard thresholding for compressed sensing. Applied and computational harmonic analysis **27**(3), 265–274 (2009)

11. Boufounos, P. T., Baraniuk, R. G.: 1-bit compressive sensing. In 2008 42nd Annual Conference on Information Sciences and Systems, pages 16–21, (2008)
12. Candes, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory **52**(2), 489–509 (2006)
13. Candes, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Processing Magazine **25**(2), 21–30 (2008)
14. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics **59**(8), 1207–1223 (2006)
15. Chaux, C., Pesquet, J.-C., Pustelnik, N.: Nested iterative algorithms for convex constrained image recovery problems. SIAM Journal on Imaging Sciences **2**(2), 730–762 (2009)
16. Chen, Z., Liu, Z., Ai, L., Zhang, S., Yang, Y.: Mask-guided spatial-temporal graph neural network for multifrequency electrical impedance tomography. IEEE Transactions on Instrumentation and Measurement **71**, 1–10 (2022)
17. Colibazzi, F., Lazzaro, D., Morigi, S., Samoré, A.: Learning nonlinear electrical impedance tomography. Journal of Scientific Computing **90**(1), 58 (2022)
18. Colibazzi, F., Lazzaro, D., Morigi, S., Samoré, A.: Deep-plug-and-play proximal gauss-newton method with applications to nonlinear, ill-posed inverse problems. Inverse Problems and Imaging **17**(6), 1226–1248 (2023)
19. Colibazzi, F., Lazzaro, D., Morigi, S., Samorè, A.: Limited electrodes models in electrical impedance tomography reconstruction. In: Calatroni, L., Donatelli, M., Morigi, S., Prato, M., Santacesaria, M.: editors, Scale Space and Variational Methods in Computer Vision, pages 68–80, Cham, (2023). Springer International Publishing
20. De Mol, C., De Vito, E., Rosasco, L.: Elastic-net regularization in learning theory. Journal of Complexity **25**(2), 201–230 (2009)
21. Donoho, D.: Compressed sensing. IEEE Transactions on Information Theory **52**(4), 1289–1306 (2006)
22. Ehler, M., Fornasier, M., Sigl, J.: Quasi-linear compressed sensing. Multiscale Modeling & Simulation **12**(2), 725–754 (2014)
23. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media (1996)
24. Fey, M., Lenssen, J. E.: Fast Graph Representation Learning with PyTorch Geometric. arXiv e-prints, page arXiv:1903.02428, Mar. (2019)
25. Fletcher, A.K., Rangan, S., Goyal, V.K.: Necessary and sufficient conditions for sparsity pattern recovery. IEEE Transactions on Information Theory **55**(12), 5758–5772 (2009)
26. Gao, H., Ji, S.: Graph u-nets. In Proceedings of the 36th International Conference on Machine Learning, (2019)
27. Garde, H., Knudsen, K.: 3d reconstruction for partial data electrical impedance tomography using a sparsity prior. In Proceedings of the 10th AIMS Conference on Dynamical Systems, Differential Equations and Applications (2014), pages 495–504. American Institute of Mathematical Sciences (AIMS), 2015. 10th AIMS Conference on Dynamical Systems, Differential Equations and Applications (2014) ; Conference date: 07-07-2014 Through 11-07-2014
28. Garde, H., Knudsen, K.: Sparsity prior for electrical impedance tomography with partial data. Inverse Problems in Science and Engineering **24**(3), 524–541 (2016)
29. Gehre, M., Kluth, T., Lipponen, A., Jin, B., Seppänen, A., Kaipio, J.P., Maass, P.: Sparsity reconstruction in electrical impedance tomography: An experimental evaluation. Journal of Computational and Applied Mathematics **236**(8), 2126–2136 (2012). (**Inverse Problems: Computation and Applications**)
30. Gilbert, A.C., Levinson, H.W., Schotland, J.C.: Nonlinear iterative hard thresholding for inverse scattering. SIAM Journal on Imaging Sciences **13**(1), 108–140 (2020)
31. González, G., Kolehmainen, V., Seppänen, A.: Isotropic and anisotropic total variation regularization in electrical impedance tomography. Computers & Mathematics with Applications **74**(3), 564–576 (2017)
32. Grasmair, M., Scherzer, O., Haltmeier, M.: Necessary and sufficient conditions for linear convergence of l1-regularization. Communications on Pure and Applied Mathematics **64**(2), 161–182 (2011)
33. Harrach, B.: The calderón problem with finitely many unknowns is equivalent to convex semidefinite optimization. SIAM Journal on Mathematical Analysis **55**(5), 5666–5684 (2023)
34. Herrholz, E., Teschke, G.: Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. Inverse Problems **26**(12), 125012 (2010)
35. Hurault, S., Chambolle, A., Leclaire, A., Papadakis, N.: A relaxed proximal gradient descent algorithm for convergent plug-and-play with proximal denoiser. In International Conference on Scale Space and Variational Methods in Computer Vision, pages 379–392. Springer, (2023)
36. Kindermann, S.: On the tangential cone condition for electrical impedance tomography. ETNA - Electronic Transactions on Numerical Analysis, (2021)

37. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. In Proceedings of 3rd International Conference on Learning Representations, (2015)
38. Kipf, T. N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. In International Conference on Learning Representations, (2017)
39. Klodt, M., Hauser, R.: Nonlinear compressed sensing for multi-emitter x-ray imaging. In: Pelillo, M., Hancock, E.: editors, Energy Minimization Methods in Computer Vision and Pattern Recognition, pages 189–204, Cham, (2018). Springer International Publishing
40. Lechleiter, A., Rieder, A.: Newton regularizations for impedance tomography: a numerical study. Inverse Problems **22**(6), 1967–1987 (2006)
41. Lechleiter, A., Rieder, A.: Newton regularizations for impedance tomography: convergence by local injectivity. Inverse problems **24**(6), 065009 (2008)
42. Li, G., Pong, T.K.: Calculus of the exponent of kurdyka-łojasiewicz inequality and its applications to linear convergence of first-order methods. Foundations of computational mathematics **18**(5), 1199–1232 (2018)
43. Li, Y., Osher, S.: A new median formula with applications to PDE based denoising. Communications in Mathematical Sciences **7**(3), 741–753 (2009)
44. Ma, X., Wang, Z., Li, Y., Arce, G.R., Dong, L., Garcia-Frias, J.: Fast optical proximity correction method based on nonlinear compressive sensing. Opt. Express **26**(11), 14479–14498 (2018)
45. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. (2006)
46. Moravec, M. L., Romberg, J. K., Baraniuk, R. G.: Compressive phase retrieval. In D. Van De Ville, V. K. Goyal, and M. Papadakis, editors, Wavelets XII, volume 6701 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, page 670120, Sept. (2007)
47. Nasehi Tehrani, J., Jin, C., McEwan, A., van Schaik, A.: A comparison between compressed sensing algorithms in electrical impedance tomography. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pages 3109–3112, (2010)
48. Ohlsson, H., Allen, Y., Roy, D., Verhaegen, M., Shankar, S.: Quadratic basis pursuit. Kernels, and Support Vector Machines, London. Chapman and Hall/CRC, In Regularization, Optimization (2014)
49. Ohlsson, H., Yang, A. Y., Dong, R., Sastry, S. S.: Nonlinear basis pursuit. In 47th Asilomar Conference on Signals, Systems and Computers, pages 115–119. IEEE, (2013)
50. Pustelnik, N., Condat, L.: Proximity operator of a sum of functions; application to depth map estimation. IEEE Signal Processing Letters **24**(12), 1827–1831 (2017)
51. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**, 259–268 (1992)
52. Shiraz, A., Khodadad, D., Nordebo, S., Yerworth, R., Frerichs, I., van Kaam, A., Kallio, M., Papadouri, T., Bayford, R., Demosthenous, A.: Compressive sensing in electrical impedance tomography for breathing monitoring. Physiological Measurement **40**(3), 034010 (2019)
53. Somersalo, E., Cheney, M., Isaacson, D.: Existence and uniqueness for electrode models for electric current computed tomography. SIAM J. Appl. Math. **52**(4), 1023–1040 (1992)
54. Unser, M., Fageot, J., Gupta, H.: Representer theorems for sparsity-promoting l1 regularization. IEEE Transactions on Information Theory **62**(9), 5167–5180 (2016)
55. Vauhkonen, P., Vauhkonen, M., Savolainen, T., Kaipio, J.: Three-dimensional electrical impedance tomography based on the complete electrode model. IEEE Transactions on Biomedical Engineering **46**(9), 1150–1160 (1999)
56. Wainwright, M. J.: Sharp thresholds for high-dimensional and noisy recovery of sparsity. arXiv preprint math/0605740, (2006)
57. Wang, J.: Non-convex lp regularization for sparse reconstruction of electrical impedance tomography. Inverse Problems in Science and Engineering **29**(7), 1032–1053 (2021)
58. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M.: Graph neural networks: A review of methods and applications. AI Open **1**, 57–81 (2020)
59. Zong, Z., Wang, Y., He, S., Zhu, Y.-J., Wei, Z.: A compressive learning-based scheme for nonlinear reconstructions in electrical impedance tomography. IEEE Transactions on Instrumentation and Measurement **73**, 1–13 (2024)
60. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology **67**(2), 301–320 (2005)