HEURISTIC DISCREPANCY PRINCIPLE FOR VARIATIONAL REGULARIZATION OF INVERSE PROBLEMS

HUAN LIU, ROMMEL REAL, XILIANG LU, XIANZHENG JIA, AND QINIAN JIN

Abstract. We consider the variational regularization for inverse problems in a general form. Based on the discrepancy principle, we propose a heuristic parameter choice rule for choosing the regularization parameter which does not require the information on the noise level and is therefore purely data driven. Under variational source conditions, we obtain a posteriori error estimates. According to the Bakushinskii veto, convergence in the worst case scenario can not be expected in general. However, by imposing certain conditions on the random noise, we establish a convergence result for the heuristic rule. Applications of the results are addressed and numerical simulations are reported.

1. **Introduction**

Inverse problems have received tremendous attention due to their wide applications. In this paper we consider a general framework of variational regularization for solving inverse problems. Let $(\mathcal{Q}, \tau_{\mathcal{Q}})$ and $(\mathcal{U}, \tau_{\mathcal{U}})$ be two topological spaces with the topology $\tau_{\mathcal{Q}}$ and $\tau_{\mathcal{U}}$ respectively. We consider inverse problems that can be formulated as the form

$$\mathcal{S}(q, u^\dagger) = 0, \tag{1.1}$$

where $u^\dagger \in \mathcal{U}$ is the exact data, $q \in \mathcal{Q}$ is the parameter to be determined, and $\mathcal{S} : \mathcal{Q} \times \mathcal{U} \to [0, \infty]$ is a proper data misfit functional. Here the properness means that $\mathcal{S}$ takes a finite value at some point in $\mathcal{Q} \times \mathcal{U}$. The formulation (1.1) provides a general framework to cover a broad range of inverse problems. For instance, the inverse problems of the form

$$F(q) = u^\dagger \tag{1.2}$$

has been studied extensively, where $F : \mathcal{Q} \to \mathcal{U}$ is an operator from a topological space $(\mathcal{Q}, \tau_{\mathcal{Q}})$ to a Banach space $\mathcal{U}$. This type of inverse problems can be formulated in the form (1.1) by taking

$$\mathcal{S}(q, u) = \|F(q) - u\|^r, \qquad \forall (q, u) \in \mathcal{Q} \times \mathcal{U},$$

where $0 < r < \infty$ is a number and $\|\cdot\|$ denotes the norm on $\mathcal{U}$. More examples of inverse problems that can be formulated in the form (1.1) will be provided later.

The equation (1.1) may have many solutions if there exists one. In order to pick the one with the desired feature, we choose a proper penalty functional $\mathcal{R} : \mathcal{Q} \to (-\infty, \infty]$ to determine a solution $q^\dagger$ such that

$$\mathcal{R}(q^\dagger) = \min \left\{ \mathcal{R}(q) : q \in \mathcal{Q} \text{ and } \mathcal{S}(q, u^\dagger) = 0 \right\}, \tag{1.3}$$

which, if exists, is called an $\mathcal{R}$-minimizing solution of (1.1). Determining an $\mathcal{R}$-minimizing solution of (1.1) is ill-posed in general. In order to determine an $\mathcal{R}$-minimizing solution stably, regularization techniques should be employed.

Let $\tilde{u}$ be a noisy data. The variational regularization for solving (1.1) is to use the solution of the minimization problem

$$\tilde{q}_\alpha \in \arg\min_{q\in\mathcal{Q}} \{\mathcal{S}(q,\tilde{u}) + \alpha\mathcal{R}(q)\} \tag{1.4}$$

to approximate an $\mathcal{R}$-minimizing solution $q^\dagger$, where $\alpha > 0$ is the so-called regularization parameter. The approximation accuracy of $\tilde{q}_\alpha$ to $q^\dagger$ depends heavily on the choice of the regularisation parameter $\alpha > 0$. In case a good estimate on $\mathcal{S}(q^\dagger, \tilde{u})$ is known, one may choose α by *a priori* or *a posteriori* manner. However, the knowledge on $\mathcal{S}(q^\dagger, \tilde{u})$ is usually obtained by estimation and a good estimate may be difficult to obtain. In case $\mathcal{S}(q^\dagger, \tilde{u})$ is overestimated or underestimated, it could result in a poor approximate solution. Therefore, it is necessary to develop parameter choice rules which depends only on the data $\tilde{u}$.

For variational regularization for linear ill-posed inverse problems in Hilbert spaces, Hanke and Raus proposed in [13] a heuristic rule for selecting the regularization parameter α, which can be viewed as a heuristic variant of the discrepancy principle. This heuristic parameter choice rule was extended in [24] for convex variational regularization for solving linear ill-posed problems $Fq = u^\dagger$, where $F : \mathcal{Q} \to \mathcal{U}$ is a bounded linear operator from a Banach space $\mathcal{Q}$ to a Hilbert space $\mathcal{U}$, and a convergence result was established under a source condition on the sought solution and a randomness condition on the noisy data. The heuristic rule of Hanke and Raus was further generalized in [25] for the variational regularization methods for solving linear as well as nonlinear ill-posed inverse problems of the form (1.2) in Banach spaces and some convergence results were proved under a randomness condition on the noisy data without relying on any source conditions on the sought solution. Since $\mathcal{U}$ is allowed to be a general Banach space, the results in [25] can be used to deal with the situation that the data is corrupted by various types of noise. However, the results in [25] requires the forward operator F to satisfy certain conditions, such as the tangential cone condition, when F is a nonlinear operator, which restricts the range of applications. In this paper we will extend the heuristic discrepancy principle of Hanke and Raus to the variational regularization (1.4) under most general conditions and provide applications which can not be covered by the model in [25].

This paper is organized as follows. In Section 2 we first give the precise description of the heuristic discrepancy principle, we then derive an *a posteriori* error estimates when the sought solution satisfies a variational source condition and establish a convergence result under a randomness condition on the noisy data without using any source condition on the sought solution. We also give a discussion on the variational source condition for inverse problems of the form (1.1) by borrowing an idea from the recent work [9]. In Section 3 we will provide examples of inverse problems which can be cast into the form (1.1); in particular we give a detailed discussion on a convex regularization framework for a class of elliptic parameter estimation problems and on the Kohn-Vogenius formulation for the electrical impedance tomography. In Section 4 we provide numerical simulations to test the theoretical results.

2. **Theory**

Throughout this paper we will assume that (1.1) has a solution in $\mathrm{dom}(\mathcal{R})$, the domain of $\mathcal{R}$. In this section we will consider the variational regularization (1.4) for solving (1.1). We will formulate the heuristic discrepancy principle and give its detailed analysis. Throughout the paper we will use $\xrightarrow{\tau_\mathcal{Q}}$ and $\xrightarrow{\tau_\mathcal{U}}$ to denote the convergence with respect to the topologies $\tau_\mathcal{Q}$ and $\tau_\mathcal{U}$ respectively. Our analysis is based on the following conditions on the penalty functional $\mathcal{R}:\mathcal{Q}\to[0,\infty]$ and the data misfit term $\mathcal{S}:\mathcal{Q}\times\mathcal{U}\to[0,\infty]$.

Assumption 2.1.
(i) *$\mathcal{R}:\mathcal{Q}\to[0,\infty]$ is proper, nonnegative and $\tau_\mathcal{Q}$-lower semi-continuous.*
(ii) *For every $\rho>0$, the sublevel set*

$$\mathcal{M}_\rho := \{q\in\mathcal{Q} : \mathcal{R}(q)\le\rho\}$$

is $\tau_\mathcal{Q}$-sequentially compact. That is, every sequence $\{q_n\}$ in $\mathcal{M}_\rho$ has a $\tau_\mathcal{Q}$-convergent subsequence in $\mathcal{Q}$.
(iii) *The mapping $(q,u)\to\mathcal{S}(q,u)$ is $(\tau_\mathcal{Q}\times\tau_\mathcal{U})$-lower semi-continuous over $\mathcal{Q}\times\mathcal{U}$, i.e. $\{(q_n,u_n)\}\subset\mathcal{Q}\times\mathcal{U}$ with $q_n\xrightarrow{\tau_\mathcal{Q}} q\in\mathcal{Q}$ and $u_n\xrightarrow{\tau_\mathcal{U}} u\in\mathcal{U}$ implies that*

$$\mathcal{S}(q,u)\le\liminf_{n\to\infty}\mathcal{S}(q_n,u_n).$$

(iv) *For each $q\in\mathcal{Q}$ the function $u\to\mathcal{S}(q,u)$ is $\tau_\mathcal{U}$-continuous.*

Under Assumption 2.1, by a standard argument one can show that (1.1) has an $\mathcal{R}$-minimizing solution $q^\dagger$ and the minimization problem (1.4) has a solution $\tilde{q}_\alpha$ for each $\alpha>0$. In case a good estimate on $\mathcal{S}(q^\dagger,\tilde{u})$ is available, one may choose α by *a priori* or *a posteriori* manner to investigate the approximation property of $\tilde{q}_\alpha$. Since the knowledge on $\mathcal{S}(q^\dagger,\tilde{u})$ is usually hard to obtain, we consider to choose the regularization parameter α by a heuristic rule which depends only on $\tilde{u}$. Motivated by the work in [13], [7, Chapter 4] and [25], we propose the following heuristic discrepancy principle to choose the regularization parameter.

Rule 2.1. *Let $A\ge 0$, $\alpha_0>0$ and $0<\gamma<1$ be given numbers, and let*

$$\Delta_\gamma=\left\{\alpha_0\gamma^j : j=0,1,\cdots\right\}.$$

Define $\alpha_* := \alpha_*(\tilde{u})\in\Delta_\gamma$ such that

$$\alpha_*\in\arg\min_{\alpha\in\Delta_\gamma}\left\{\Theta(\alpha,\tilde{u}) := \left(\frac{1}{\alpha}+A\right)\mathcal{S}(\tilde{q}_\alpha,\tilde{u})\right\}$$

and use $\tilde{q}_{\alpha_*}$ as an approximate solution.

When using Rule 2.1, we should avoid to choose $\alpha_0>0$ too small since otherwise it may produce a small regularization parameter α_* and hence result in a bad approximate solution that is oscillatory. On the other hand, by the definition of $\tilde{q}_\alpha$ we have

$$\mathcal{S}(\tilde{q}_\alpha,\tilde{u})+\alpha\mathcal{R}(\tilde{q}_\alpha)\le\mathcal{S}(q_m,\tilde{u})+\alpha\mathcal{R}(q_m),$$

where $q_m\in\mathcal{Q}$ denotes an element such that $\mathcal{R}(q_m)=\min_{q\in\mathcal{Q}}\mathcal{R}(q)$ whose existence is guaranteed by Assumption 2.1. Thus $\mathcal{S}(\tilde{q}_\alpha,\tilde{u})\le\mathcal{S}(q_m,\tilde{u})$ and so

$$\frac{1}{\alpha}\mathcal{S}(\tilde{q}_\alpha,\tilde{u})\to 0 \quad \text{as } \alpha\to\infty.$$

Consequently, if α_0 is too large and $A = 0$, Rule 2.1 may output a large parameter α_* which may result in an approximate solution with less accuracy. Therefore, when using Rule 2.1, we suggest to choose $A > 0$ to be suitably large and $\alpha_0 > 0$ to be suitably small. The choice of $\alpha_0 > 0$ can be based on a rough guess of the optimal regularization parameter.

2.1. **A posteriori error estimates.** We first derive *a posteriori* error estimates on $\tilde{q}_{\alpha_*}$ with α_* chosen by Rule 2.1. To achieve this, we need conditions on the sought $\mathcal{R}$-minimizing solution $q^\dagger$. We will assume the following variational source condition on $q^\dagger$.

Assumption 2.2. *There is an error function $\mathcal{E} : \mathcal{Q} \times \mathcal{Q} \to [0, \infty)$ with $\mathcal{E}(q, q) = 0$ whenever $q \in \mathcal{Q}$ such that*

$$\mathcal{E}(q, q^\dagger) \le \mathcal{R}(q) - \mathcal{R}(q^\dagger) + \varphi(\mathcal{S}(q, u^\dagger)) \tag{2.1}$$

for all $q \in \mathcal{M}_\rho := \{q \in \mathcal{Q} : \mathcal{R}(q) \le \rho\}$ with $\rho > \mathcal{R}(q^\dagger)$, where $\varphi : [0, \infty) \to [0, \infty)$ is a concave index function. Here φ is called an index function if it is continuous and strictly increasing with $\varphi(0) = 0$.

We will give a detailed discussion on Assumption 2.2 in Proposition 2.4. In order to carry out the derivation of the error estimate, we need the following "quasi-triangle inequality" on the data misfit term $\mathcal{S}$.

Assumption 2.3. *There is a constant $C_S \ge 1$ such that for any $u^\dagger \in \mathcal{U}$ and $q^\dagger \in \mathcal{Q}$ satisfying $\mathcal{S}(q^\dagger, u^\dagger) = 0$ and any $u \in \mathcal{U}$ and $q \in \mathcal{Q}$ there holds*

$$\mathcal{S}(q, u^\dagger) \le C_{\mathcal{S}} \left(\mathcal{S}(q^\dagger, u) + \mathcal{S}(q, u) \right).$$

Now we are ready to prove the following *a posteriori* error estimate.

Theorem 2.2. *Let $\mathcal{R}$ and $\mathcal{S}$ satisfy Assumption 2.1 and Assumption 2.3. Assume that $q^\dagger$ is an $\mathcal{R}$-minimizing solution of (1.1) satisfying Assumption 2.2 with $\Phi(t) := t/\varphi(t)$ being strictly increasing. Let α_* be determined by Rule 2.1. If $\tilde{q}_{\alpha_*} \in \mathcal{M}_\rho$ and $\delta_* := \mathcal{S}(\tilde{q}_{\alpha_*}, \tilde{u}) \ne 0$, then*

$$\mathcal{E}(\tilde{q}_{\alpha_*}, q^\dagger) \le C \left(1 + \frac{\delta}{\delta_*} \right) (\delta + \varphi(\delta + \delta_*)). \tag{2.2}$$

where $\delta := \mathcal{S}(q^\dagger, \tilde{u})$ denotes the noise level and C is a constant depending only on α_0, γ, A and $C_{\mathcal{S}}$.

Proof. We first show that if $\tilde{q}_\alpha \in \mathcal{M}_\rho$, then

$$\mathcal{E}(\tilde{q}_\alpha, q^\dagger) \le \frac{\delta}{\alpha} + C_{\mathcal{S}} \varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})), \tag{2.3}$$

$$\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le 2\delta + \Phi^{-1}(3 C_{\mathcal{S}} \alpha). \tag{2.4}$$

To see this, by using the minimality of $\tilde{q}_\alpha$ we have

$$\alpha \left(\mathcal{R}(\tilde{q}_\alpha) - \mathcal{R}(q^\dagger) \right) + \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \mathcal{S}(q^\dagger, \tilde{u}) = \delta.$$

From Assumption 2.2 it then follows that

$$\alpha \mathcal{E}(\tilde{q}_\alpha, q^\dagger) + \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \delta + \alpha \varphi(\mathcal{S}(\tilde{q}_\alpha, u^\dagger)). \tag{2.5}$$

Note that φ is a concave index function, we have $\varphi(bt) \le b\varphi(t)$ for any $t \ge 0$ and $b \ge 1$. By virtue of Assumption 2.3 we then have

$$\begin{aligned}\varphi(\mathcal{S}(\tilde{q}_\alpha, u^\dagger)) &\le \varphi(C_{\mathcal{S}}(\mathcal{S}(q^\dagger, \tilde{u}) + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})) \\ &\le C_{\mathcal{S}}\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})).\end{aligned}$$

Combining this with (2.5) yields

$$\alpha\mathcal{E}(\tilde{q}_\alpha, q^\dagger) + \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \delta + C_{\mathcal{S}}\alpha\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})). \tag{2.6}$$

Since $\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \ge 0$, by dropping this term on the left hand side of (2.6) we can obtain (2.3). Next, by dropping the nonnegative term $\alpha\mathcal{E}(\tilde{q}_\alpha, q^\dagger)$ on the left hand side of (2.6) we have

$$\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \delta + C_{\mathcal{S}}\alpha\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})). \tag{2.7}$$

If $C_{\mathcal{S}}\alpha\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})) \le \delta$, then

$$\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le 2\delta. \tag{2.8}$$

If $C_{\mathcal{S}}\alpha\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})) > \delta$, we have from (2.7) that

$$\begin{aligned}\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) &\le 2\delta + C_{\mathcal{S}}\alpha\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})) \\ &\le 3C_{\mathcal{S}}\alpha\varphi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})).\end{aligned}$$

Consequently $\Phi(\delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u})) \le 3C_{\mathcal{S}}\alpha$ and hence

$$\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \delta + \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \Phi^{-1}(3C_{\mathcal{S}}\alpha)$$

which together with (2.8) shows (2.4).

Now we are ready to derive the a posteriori error estimate (2.2). Since $\tilde{q}_{\alpha_*}$ is assumed to be in $\mathcal{M}_\rho$, we may use (2.3) to derive that

$$\begin{aligned}\mathcal{E}(\tilde{q}_{\alpha_*}, q^\dagger) &\le \frac{\delta}{\alpha_*} + C_{\mathcal{S}}\varphi(\delta + \mathcal{S}(\tilde{q}_{\alpha_*}, \tilde{u})) \\ &\le \frac{\delta}{\delta_*}\Theta(\alpha_*, \tilde{u}) + C_{\mathcal{S}}\varphi(\delta + \delta_*).\end{aligned} \tag{2.9}$$

We need to estimate $\Theta(\alpha_*, \tilde{u})$. If $\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le 3\delta$ for all $\alpha \in \Delta_\gamma$, we have

$$\begin{aligned}\Theta(\alpha_*, \tilde{u}) \le \Theta(\alpha_0, \tilde{u}) &\le \left(\frac{1}{\alpha_0} + A\right)\mathcal{S}(\tilde{q}_{\alpha_0}, \tilde{u}) \\ &\le 3\left(\frac{1}{\alpha_0} + A\right)\delta.\end{aligned} \tag{2.10}$$

Next we will assume that $\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) > 3\delta$ for some $\alpha \in \Delta_\gamma$. By the minimality of $\tilde{q}_\alpha$ and the nonnegativity of $\mathcal{R}$ we have

$$\begin{aligned}\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) &\le \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) + \alpha\mathcal{R}(\tilde{q}_\alpha) \\ &\le \alpha\mathcal{R}(q^\dagger) + \mathcal{S}(q^\dagger, \tilde{u}) \\ &= \alpha\mathcal{R}(q^\dagger) + \delta.\end{aligned}$$

This implies that $\lim_{\alpha\searrow 0}\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \le \delta$. Therefore there must exist a largest number $\hat{\alpha}$ in Δ_γ such that

$$\mathcal{S}(\tilde{q}_{\gamma\hat{\alpha}}, \tilde{u}) \le 3\delta < \mathcal{S}(\tilde{q}_{\hat{\alpha}}, \tilde{u}).$$

We need a lower bound for $\hat{\alpha}$. By the minimality of $\tilde{q}_{\hat{\alpha}}$ we have

$$\begin{aligned} 3\delta + \hat{\alpha}\mathcal{R}(\tilde{q}_{\hat{\alpha}}) &\le \mathcal{S}(\tilde{q}_{\hat{\alpha}}, \tilde{u}) + \hat{\alpha}\mathcal{R}(\tilde{q}_{\hat{\alpha}}) \\ &\le \mathcal{S}(q^\dagger, \tilde{u}) + \hat{\alpha}\mathcal{R}(q^\dagger) \\ &= \delta + \hat{\alpha}\mathcal{R}(q^\dagger). \end{aligned}$$

This implies that $\mathcal{R}(\tilde{q}_{\hat{\alpha}}) \le \mathcal{R}(q^\dagger) < \rho$ and thus $\tilde{q}_{\hat{\alpha}} \in \mathcal{M}_\rho$. Consequently we may use (2.4) to obtain

$$3\delta \le \mathcal{S}(\tilde{q}_{\hat{\alpha}}, \tilde{u}) \le 2\delta + \Phi^{-1}(3C_{\mathcal{S}}\hat{\alpha}).$$

Thus $\delta \le \Phi^{-1}(3C_{\mathcal{S}}\hat{\alpha})$. This together with the monotonicity of Φ gives

$$\hat{\alpha} \ge \frac{1}{3C_{\mathcal{S}}}\Phi(\delta).$$

Now we may use the minimality of $\Theta(\alpha_*, \tilde{u})$ to derive that

$$\begin{aligned} \Theta(\alpha_*, \tilde{u}) \le \Theta(\gamma\hat{\alpha}, \tilde{u}) &= \left(\frac{1}{\gamma\hat{\alpha}} + A\right)\mathcal{S}(\tilde{q}_{\gamma\hat{\alpha}}, \tilde{u}) \\ &\le 3\left(\frac{1}{\gamma\hat{\alpha}} + A\right)\delta \le 3A\delta + \frac{9C_{\mathcal{S}}}{\gamma}\frac{\delta}{\Phi(\delta)} \\ &= 3A\delta + \frac{9C_{\mathcal{S}}}{\gamma}\varphi(\delta). \end{aligned} \tag{2.11}$$

Combining (2.10) and (2.11) with (2.9) we obtain the desired estimate (2.2). □

The a posteriori estimate in Theorem 2.2 involves the quantity δ_*. If δ_* is about the order of δ, it gives convergence rate $O(\varphi(\delta))$. If δ_* is much larger than δ, only weaker convergence rates are available. If δ_* is significantly smaller than δ, the factor δ/δ_* blows up and the approximation may diverge. Therefore, the quantity δ_* provides an a posteriori check of Rule 2.1, its value should always be monitored and the computed approximation should be discarded if δ_* is presumably too small.

In Theorem 2.2 we have assumed that Rule 2.1 defines a positive number α_* with $\tilde{q}_{\alpha_*} \in \mathcal{M}_\rho$. This requirement can be guaranteed if the noisy data $\tilde{u}$ satisfies the following condition.

Assumption 2.4. *There is a constant $\kappa > 0$ such that*

$$\mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \ge \kappa\mathcal{S}(q^\dagger, \tilde{u}) \tag{2.12}$$

for any minimizer $\tilde{q}_\alpha$ of (1.4) with any $\alpha \in \Delta_\gamma$.

For an interpretation of Assumption 2.4 one may refer to [25] when the inverse problem takes the form (1.2). Rough speaking, Assumption 2.4 stipulates a randomness condition on the noise that corrupts the data. It would be interesting to explore a statistical interpretation on Assumption 2.4 which however remains open.

Lemma 2.3. *Let $\mathcal{R}$ and $\mathcal{S}$ satisfy Assumption 2.1 and Assumption 2.3. Let the noisy data $\tilde{u}$ satisfy Assumption 2.4. If*

$$\mathcal{S}(q^\dagger, \tilde{u}) \le \alpha_0\mathcal{R}(q^\dagger) \quad \text{and} \quad \left(1 + \frac{2(1 + A\alpha_0)}{\kappa}\right)\mathcal{R}(q^\dagger) \le \rho,$$

then Rule 2.1 determines a positive parameter $\alpha_* \in \Delta_\gamma$ with $\tilde{q}_{\alpha_*} \in \mathcal{M}_\rho$.

Proof. Let $\delta := \mathcal{S}(q^\dagger, \tilde{u})$. From Assumption 2.4 it follows that

$$\Theta(\alpha, \tilde{u}) = \left(\frac{1}{\alpha} + A\right) \mathcal{S}(\tilde{q}_\alpha, \tilde{u}) \geq \left(\frac{1}{\alpha} + A\right) \kappa \mathcal{S}(q^\dagger, \tilde{u}) = \left(\frac{1}{\alpha} + A\right) \kappa\delta$$

for all $\alpha > 0$. This implies that $\Theta(\alpha, \tilde{u}) \to \infty$ as $\alpha \to 0$. Therefore there must exist an $\alpha_* \in \Delta_\gamma$ that gives the minimum of $\Theta(\alpha, \tilde{u})$ over Δ_γ. By using Assumption 2.4 and the minimizing property of $\tilde{q}_{\alpha_0}$ we have

$$\begin{aligned}\left(\frac{1}{\alpha_*} + A\right) \kappa\delta \leq \Theta(\alpha_*, \tilde{u}) \leq \Theta(\alpha_0, \tilde{u}) &= \left(\frac{1}{\alpha_0} + A\right) \mathcal{S}(\tilde{q}_{\alpha_0}, \tilde{u}) \\ &\leq \left(\frac{1}{\alpha_0} + A\right) \left(\delta + \alpha_0 \mathcal{R}(q^\dagger)\right) \\ &= (1 + A\alpha_0) \left(\frac{\delta}{\alpha_0} + \mathcal{R}(q^\dagger)\right).\end{aligned}$$

In view of the condition $\delta = \mathcal{S}(q^\dagger, \tilde{u}) \leq \alpha_0 \mathcal{R}(q^\dagger)$ we can obtain

$$\frac{\kappa\delta}{\alpha_*} \leq \left(\frac{1}{\alpha_*} + A\right) \kappa\delta \leq 2(1 + A\alpha_0)\mathcal{R}(q^\dagger)$$

which shows that

$$\alpha_* \geq \frac{\kappa\delta}{2(1 + A\alpha_0)\mathcal{R}(q^\dagger)}.$$

Consequently, it follows from the minimizing property of $\tilde{q}_{\alpha_*}$ that

$$\mathcal{R}(\tilde{q}_{\alpha_*}) \leq \frac{\delta}{\alpha_*} + \mathcal{R}(q^\dagger) \leq \left(\frac{2(1 + A\alpha_0)}{\kappa} + 1\right) \mathcal{R}(q^\dagger) \leq \rho.$$

Thus $\tilde{q}_{\alpha_*} \in \mathcal{M}_\rho$ and the proof is complete. $\square$

We remark that the *a posteriori* error estimate in Theorem 2.2 is based on the variational source condition on $q^\dagger$ specified in Assumption 2.2. Variational source conditions were first introduced in [21], as a generalization of the standard source conditions in Hilbert spaces, to derive convergence rates of Tikhonov regularization in Banach spaces. This kind of source conditions was further generalized and refined later, see [2, 12, 22, 33] for more information. Recently it has been shown in [9] that, for inverse problems of the form (1.2) in Banach spaces, under certain natural conditions, a variational source condition is always satisfied with a suitable concave index function φ. The error function $\mathcal{E}$ in Assumption 2.2 is used to measure the speed of convergence. In the original version of variational source conditions, Bregman distance induced by $\mathcal{R}$ is used as an error function. Use of a general error function has the advantage of covering a wider range of applications, see [9, 11]. In the following result we extend the work in [9] for inverse problems of the form (1.1).

Proposition 2.4. *Let Assumption 2.1 hold and let $\rho > \mathcal{R}(q^\dagger)$. Furthermore, assume that there is an error function $\mathcal{E} : \mathcal{Q} \times \mathcal{Q} \to [0, \infty)$ with $\mathcal{E}(q, q) = 0$ whenever $q \in \mathcal{Q}$ such that the following conditions hold:*

(i) *$\mathcal{E}(q^*, q^\dagger) \leq \mathcal{R}(q^*) - \mathcal{R}(q^\dagger)$ for all $q^* \in \mathcal{M}_\rho$ with $\mathcal{S}(q^*, u^\dagger) = 0$;*
(ii) *The mapping $q \to -\mathcal{E}(q, q^\dagger) + \mathcal{R}(q)$ is $\tau_\mathcal{Q}$ lower semi-continuous on $\mathcal{M}_\rho$;*
(iii) *There exist $\beta > 1$ and $c \in \mathbb{R}$ such that $\beta\mathcal{E}(q, q^\dagger) - \mathcal{R}(q) \leq c$ for all $q \in \mathcal{M}_\rho$.*

Then there exists a concave strictly increasing index function $\varphi : [0,\infty) \to [0,\infty)$ *such that*

$$\mathcal{E}(q, q^\dagger) \le \mathcal{R}(q) - \mathcal{R}(q^\dagger) + \varphi(\mathcal{S}(q, u^\dagger)), \quad \forall q \in \mathcal{M}_\rho.$$

Proof. The proof essentially follows [9] with refinements. For $r \ge 0$ we define

$$D(r) := \sup_{q\in\mathcal{M}_\rho} \left\{\mathcal{E}(q, q^\dagger) - \mathcal{R}(q) + \mathcal{R}(q^\dagger) - r\mathcal{S}(q, u^\dagger)\right\}.$$

It is clear that $D(r)$ is monotonically decreasing with respect to $r \in [0,\infty)$. Furthermore, by (iii) we have $D(r) \le D(0) \le c + \mathcal{R}(q^\dagger) < \infty$ for all $r \in [0,\infty)$. We claim that

$$\lim_{r\to\infty} D(r) \le 0.$$

If there is an $\hat{r} \ge 0$ such that $D(\hat{r}) \le 0$, then this claim holds obviously by the monotonicity of $D(r)$. So we may assume that $D(r) > 0$ for all $r \ge 0$. Let $\varepsilon > 0$ be an arbitrarily small number. For each $r \ge 1$ we may choose $q_r \in \mathcal{M}_\rho$ such that

$$\mathcal{E}(q_r, q^\dagger) - \mathcal{R}(q_r) + \mathcal{R}(q^\dagger) - r\mathcal{S}(q_r, u^\dagger) \ge D(r) - \varepsilon. \tag{2.13}$$

According to (iii) we have

$$\begin{aligned}\mathcal{E}(q_r, q^\dagger) - \mathcal{R}(q_r) &= \frac{1}{\beta}\left(\beta\mathcal{E}(q_r, q^\dagger) - \mathcal{R}(q_r)\right) - \left(1 - \frac{1}{\beta}\right)\mathcal{R}(q_r)\\ &\le \frac{c}{\beta} - \left(1 - \frac{1}{\beta}\right)\mathcal{R}(q_r).\end{aligned}$$

Combining the above two inequalities gives

$$\begin{aligned}\left(1 - \frac{1}{\beta}\right)\mathcal{R}(q_r) + r\mathcal{S}(q_r, u^\dagger) &\le \frac{c}{\beta} + \mathcal{R}(q^\dagger) + \varepsilon - D(r)\\ &\le \frac{c}{\beta} + \mathcal{R}(q^\dagger) + \varepsilon. \end{aligned}\tag{2.14}$$

This implies that $\{\mathcal{R}(q_r)\}$ is bounded. Thus, by Assumption 2.1 (ii), there exist $\hat{q} \in \mathcal{Q}$ and a sequence $\{r_n\}$ with $r_n \to \infty$ such that $q_{r_n} \xrightarrow{\tau_\mathcal{Q}} \hat{q}$ as $n \to \infty$. According to Assumption 2.1 (i) we have $\hat{q} \in \mathcal{M}_\rho$. From (2.14) we have $\mathcal{S}(q_{r_n}, u^\dagger) \to 0$ as $n \to \infty$. Therefore, in view of Assumption 2.1 (iii), $\mathcal{S}(\hat{q}, u^\dagger) = 0$. Thus it follows from (2.13), the nonnegativity of $\mathcal{S}$, (ii) and (i) that

$$\begin{aligned}\lim_{r\to\infty} D(r) &= \lim_{n\to\infty} D(r_n) = \limsup_{n\to\infty} D(r_n)\\ &\le \limsup_{n\to\infty}\left\{\mathcal{E}(q_{r_n}, q^\dagger) - \mathcal{R}(q_{r_n}) + \mathcal{R}(q^\dagger) - r_n\mathcal{S}(q_{r_n}, u^\dagger) + \varepsilon\right\}\\ &\le \limsup_{n\to\infty}\left\{\mathcal{E}(q_{r_n}, q^\dagger) - \mathcal{R}(q_{r_n})\right\} + \mathcal{R}(q^\dagger) + \varepsilon\\ &\le \mathcal{E}(\hat{q}, q^\dagger) - \mathcal{R}(\hat{q}) + \mathcal{R}(q^\dagger) + \varepsilon\\ &\le \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we must have $\lim_{r\to\infty} D(r) = 0$.

It is now ready to show the validity of the variational source condition. If $D(r) \le 0$ for some $r \ge 0$, we may take $\varphi(t) = rt$. Therefore we may assume $D(r) > 0$ for all $r \ge 0$. By the definition of $D(r)$ we have

$$\begin{aligned}\mathcal{E}(q, q^\dagger) - \mathcal{R}(q) + \mathcal{R}(q^\dagger) &= \mathcal{E}(q, q^\dagger) - R(q) + \mathcal{R}(q^\dagger) - r\mathcal{S}(q, u^\dagger) + r\mathcal{S}(q, u^\dagger)\\ &\le D(r) + r\mathcal{S}(q, u^\dagger)\end{aligned}$$

for all $r \geq 0$ and $q \in \mathcal{M}_\rho$. Let

$$\varphi(t) := \inf_{r \geq 0} \{D(r) + rt\}.$$

Then

$$\begin{aligned}\mathcal{E}(q, q^\dagger) - \mathcal{R}(q) + \mathcal{R}(q^\dagger) &\leq \inf_{r\geq 0} \left\{D(r) + r\mathcal{S}(q, u^\dagger)\right\} \\ &= \varphi(\mathcal{S}(q, u^\dagger)).\end{aligned}$$

By definition φ is a nonnegative, concave, upper semi-continuous, increasing function on $[0, \infty)$. Because $D(r) \to 0$ as $r \to \infty$, we have $\varphi(0) = 0$. Moreover, it follows from (iii) that

$$\varphi(t) \leq D(0) \leq \sup_{q \in \mathcal{M}_\rho} \left\{\beta\mathcal{E}(q, q^\dagger) - \mathcal{R}(q) + \mathcal{R}(q^\dagger)\right\} \leq c + \mathcal{R}(q^\dagger) < \infty$$

for all $t \geq 0$. Thus φ is continuous on $[0, \infty)$. Moreover, by replacing $\varphi(t)$ by $\varphi(t) + \sqrt{t}$ if necessary, we may gurantee the validity of the variational source condition with a concave strictly increasing index function. $\square$

2.2. **Convergence.** In Theorem 2.2, we have derived *a posteriori* error estimates on $\mathcal{E}(\tilde{q}_{\alpha_*}, q^\dagger)$ for a given a noisy data $\tilde{u}$ under variational source conditions on $q^\dagger$ for the regularization parameter α_* chosen by Rule 2.1. It is natural to ask, for a family of noisy data $\{u^\delta\}$ satisfying $u^\delta \xrightarrow{\tau_\mathcal{U}} u^\dagger$ as $\delta \to 0$, if q_α^δ is defined as

$$q_\alpha^\delta \in \arg\min_{q \in \mathcal{Q}} \left\{\mathcal{S}(q, u^\delta) + \alpha\mathcal{R}(q)\right\}, \tag{2.15}$$

and $\alpha_* := \alpha_*(u^\delta)$ is chosen by Rule 2.1 with $\Theta(\alpha, \tilde{u})$ replaced by $\Theta(\alpha, u^\delta)$, i.e.

$$\alpha_* \in \arg\min_{\alpha \in \Delta_\gamma} \left\{\Theta(\alpha, u^\delta) := \left(\frac{1}{\alpha} + A\right)\mathcal{S}(q_\alpha^\delta, u^\delta)\right\}, \tag{2.16}$$

is it possible to guarantee the convergence of $q_{\alpha_*}^\delta$ to $q^\dagger$ as $\delta \to 0$ without using any source conditions? Bakushinski's veto ([1]) states that heuristic rules can not lead to convergence in the sense of worst case scenario for any regularisation method. In order to obtain a convergence result, certain conditions on the noisy data should be imposed. We will use the following condition which means that $\{u^\delta\}$ satisfies Assumption 2.4 in a uniform sense.

Assumption 2.5. *$\{u^\delta\}$ is a family of noisy data with $u^\delta \xrightarrow{\tau_\mathcal{U}} u^\dagger$ as $\delta \to 0$, and there is a constant $0 < \kappa < 1$ such that*

$$\mathcal{S}(q_\alpha^\delta, u^\delta) \geq \kappa\mathcal{S}(q^\dagger, u^\delta)$$

for every u^δ and every solution q_α^δ of (2.15) with all $\alpha \in \Delta_\gamma$.

Now we are ready to prove the following convergence result on Rule 2.1.

Theorem 2.5. *Let $\mathcal{R}$ and $\mathcal{S}$ satisfy Assumption 2.1, let $\{u^\delta\}$ be a family of noisy data satisfying Assumption 2.5, and let $\alpha_* := \alpha_*(u^\delta) \in \Delta_\gamma$ be chosen by (2.16). Then*

$$\mathcal{S}(q_{\alpha_*}^\delta, u^\delta) \to 0 \quad \textit{and} \quad \mathcal{R}(q_{\alpha_*}^\delta) \to \mathcal{R}^\dagger \quad \textit{as } \delta \to 0,$$

where

$$\mathcal{R}^\dagger := \min\{\mathcal{R}(q) : q \in \mathcal{Q} \textit{ and } \mathcal{S}(q, u^\dagger) = 0\}.$$

Moreover, any sequence from $\{q_{\alpha_*}^\delta\}$ *contains a* $\tau_{\mathcal{Q}}$*-convergent subsequence whose limit is an* $\mathcal{R}$*-minimizing solution of (1.1). If, in addition, (1.1) has a unique* $\mathcal{R}$*-minimizing solution* $q^\dagger$*, then* $q_{\alpha_*}^\delta \xrightarrow{\tau_{\mathcal{Q}}} q^\dagger$ *as* $\delta \to 0$*.*

Proof. We first show that

$$\Theta(\alpha_*(u^\delta), u^\delta) \to 0 \quad \text{ as } \delta \to 0. \tag{2.17}$$

To this end, let $q^\dagger \in \mathcal{Q}$ be an $\mathcal{R}$-minimizing solution of (1.1). We take a constant $\tau > 1$ and define $\hat{\alpha} := \hat{\alpha}(u^\delta) \in \Delta_\gamma$ to be the largest number such that

$$\hat{\alpha}^2 + \mathcal{S}(q_{\hat{\alpha}}^\delta, u^\delta) \le \tau \mathcal{S}(q^\dagger, u^\delta). \tag{2.18}$$

This $\hat{\alpha}$ is well-defined. In fact, for any $\alpha > 0$ we have from the minimality of q_α^δ that

$$\mathcal{S}(q_\alpha^\delta, u^\delta) + \alpha \mathcal{R}(q_\alpha^\delta) \le \mathcal{S}(q^\dagger, u^\delta) + \alpha \mathcal{R}(q^\dagger)$$

which implies that

$$\limsup_{\alpha \to 0} \mathcal{S}(q_\alpha^\delta, u^\delta) \le \mathcal{S}(q^\dagger, u^\delta). \tag{2.19}$$

Therefore, for $\tau > 1$, there must exist a finite number $\hat{\alpha} \in \Delta_q$ satisfying (2.18). Since $u^\delta \xrightarrow{\tau_{\mathcal{U}}} u^\dagger$, by Assumption 2.1 (iv) we have $\mathcal{S}(q^\dagger, u^\delta) \to \mathcal{S}(q^\dagger, u^\dagger) = 0$ as $\delta \to 0$. Thus, by the definition of $\hat{\alpha}(u^\delta)$ we can see that $\hat{\alpha} := \hat{\alpha}(u^\delta) \to 0$ as $\delta \to 0$. We claim that

$$\frac{\mathcal{S}(q^\dagger, u^\delta)}{\hat{\alpha}} \to 0 \quad \text{ as } \delta \to 0. \tag{2.20}$$

To see this, we use the minimizing property of $q_{\hat{\alpha}/\gamma}^\delta$ to derive that

$$\mathcal{S}(q_{\hat{\alpha}/\gamma}^\delta, u^\delta) + \frac{\hat{\alpha}}{\gamma} \mathcal{R}(q_{\hat{\alpha}/\gamma}^\delta) \le \mathcal{S}(q^\dagger, u^\delta) + \frac{\hat{\alpha}}{\gamma} \mathcal{R}(q^\dagger). \tag{2.21}$$

Since $\hat{\alpha} < \hat{\alpha}/\gamma$, we have

$$\left(\frac{\hat{\alpha}}{\gamma}\right)^2 + \mathcal{S}(q_{\hat{\alpha}/\gamma}^\delta, u^\delta) > \tau \mathcal{S}(q^\dagger, u^\delta)$$

which together with (2.21) implies that

$$\begin{aligned} \tau \mathcal{S}(q^\dagger, u^\delta) + \frac{\hat{\alpha}}{\gamma} \mathcal{R}(q_{\hat{\alpha}/\gamma}^\delta) &\le \mathcal{S}(q_{\hat{\alpha}/\gamma}^\delta, u^\delta) + \frac{\hat{\alpha}}{\gamma} \mathcal{R}(q_{\hat{\alpha}/\gamma}^\delta) + \left(\frac{\hat{\alpha}}{\gamma}\right)^2 \\ &\le \mathcal{S}(q^\dagger, u^\delta) + \frac{\hat{\alpha}}{\gamma} \mathcal{R}(q^\dagger) + \left(\frac{\hat{\alpha}}{\gamma}\right)^2. \end{aligned} \tag{2.22}$$

Since $\tau > 1$, we obtain

$$\mathcal{R}(q_{\hat{\alpha}/\gamma}^\delta) \le \frac{\hat{\alpha}}{\gamma} + \mathcal{R}(q^\dagger).$$

In view of $\hat{\alpha} \to 0$ as $\delta \to 0$, we can derive that

$$\limsup_{\delta \to 0} \mathcal{R}(q_{\hat{\alpha}/\gamma}^\delta) \le \mathcal{R}(q^\dagger). \tag{2.23}$$

This shows that $\{q_{\hat{\alpha}/\gamma}^\delta\} \subseteq \mathcal{M}_\rho$ with $\rho = \mathcal{R}(q^\dagger) + 1$ for all small $\delta > 0$. By Assumption 2.1 (ii), $\{q_{\hat{\alpha}/\gamma}^\delta\}$ has a $\tau_{\mathcal{Q}}$-convergent subsequence, and, for simplicity of exposition, we denote this subsequence again by $\{q^\delta_{\hat\alpha/\gamma}\}$. Let $\hat q \in \mathcal{Q}$ be the $\tau_{\mathcal{Q}}$-limit of this subsequence. By (2.23) and the $\tau_{\mathcal{Q}}$-lower semi-continuity of $\mathcal{R}$, we have

$$\mathcal{R}(\hat q) \le \liminf_{\delta\to 0} \mathcal{R}(q^\delta_{\hat\alpha/\gamma}) \le \limsup_{\delta\to 0} \mathcal{R}(q^\delta_{\hat\alpha/\gamma}) \le \mathcal{R}(q^\dagger). \tag{2.24}$$

Since $u^\delta \xrightarrow{\tau_{\mathcal{U}}} u^\dagger$ and $q^\delta_{\hat\alpha/\gamma} \xrightarrow{\tau_{\mathcal{Q}}} \hat q$ as $\delta \to 0$, we may use Assumption 2.1 (iii) and (2.21) to obtain

$$\mathcal{S}(\hat q, u^\dagger) \le \liminf_{\delta\to 0} \mathcal{S}(q^\delta_{\hat\alpha/\gamma}, u^\delta) \le \liminf_{\delta\to 0} \left\{ \mathcal{S}(q^\dagger, u^\delta) + \frac{\hat\alpha}{\gamma}\mathcal{R}(q^\dagger) \right\} = 0.$$

Therefore, it follows from the nonnegativity of $\mathcal{S}$ that $\mathcal{S}(\hat q, u^\dagger) = 0$. Consequently, $\hat q$ is a solution of $S(q, u^\dagger) = 0$ with $\mathcal{R}(\hat q) \le \mathcal{R}(q^\dagger)$. By the $\mathcal{R}$-minimality of $q^\dagger$, we must have $\mathcal{R}(\hat q) = \mathcal{R}(q^\dagger)$ which together with (2.24) shows that

$$\lim_{\delta\to 0} \mathcal{R}(q^\delta_{\hat\alpha/\gamma}) = \mathcal{R}(\hat q) = \mathcal{R}(q^\dagger). \tag{2.25}$$

In view of (2.22), we have

$$(\tau - 1)\frac{\mathcal{S}(q^\dagger, u^\delta)}{\hat\alpha} \le \frac{\hat\alpha}{\gamma^2} + \frac{1}{\gamma}\left(\mathcal{R}(q^\dagger) - \mathcal{R}(q^\delta_{\hat\alpha/\gamma})\right).$$

Since $\tau > 1$, we may use $\hat\alpha \to 0$ and (2.25) to conclude that $\mathcal{S}(q^\dagger, u^\delta)/\hat\alpha \to 0$ as $\delta \to 0$ which shows (2.20).

Now we are ready to show (2.17). In fact, by the minimality of $\alpha_*(u^\delta)$, the choice of $\hat\alpha$, and (2.20) we have

$$\begin{aligned} 0 \le \Theta(\alpha_*, u^\delta) \le \Theta(\hat\alpha, u^\delta) &= \left(\frac{1}{\hat\alpha} + A\right)\mathcal{S}(q^\delta_{\hat\alpha}, u^\delta) \\ &\le \tau\left(\frac{1}{\hat\alpha} + A\right)\mathcal{S}(q^\dagger, u^\delta) \to 0 \end{aligned}$$

as $\delta \to 0$ which shows (2.17). Note that

$$\Theta(\alpha_*, u^\delta) = \left(\frac{1}{\alpha_*} + A\right)\mathcal{S}(q^\delta_{\alpha_*}, u^\delta) \ge \frac{\mathcal{S}(q^\delta_{\alpha_*}, u^\delta)}{\alpha_*},$$

we may use (2.17), $\alpha_* \le \alpha_0$ and $\mathcal{S}(q^\delta_{\alpha_*}, u^\delta) \ge \kappa\mathcal{S}(q^\dagger, u^\dagger)$ from Assumption 2.5 to derive that

$$\mathcal{S}(q^\delta_{\alpha_*}, u^\delta) \to 0 \quad \text{and} \quad \frac{\mathcal{S}(q^\dagger, u^\delta)}{\alpha_*} \to 0 \quad \text{as } \delta \to 0. \tag{2.26}$$

Next we show the convergence result. By using the minimality of $q^\delta_{\alpha_*}$, the nonnegativity of $\mathcal{S}$ and (2.26) we have

$$\limsup_{\delta\to 0} \mathcal{R}(q^\delta_{\alpha_*}) \le \limsup_{\delta\to 0} \left\{ \frac{\mathcal{S}(q^\dagger, u^\delta)}{\alpha_*} + \mathcal{R}(q^\dagger) \right\} = \mathcal{R}(q^\dagger).$$

Thus $q^\delta_{\alpha_*} \in \mathcal{M}_\rho$ with $\rho = \mathcal{R}(q^\dagger) + 1$ for all small $\delta > 0$. Therefore, we can use a similar argument for dealing with the convergence of $\{q^\delta_{\hat\alpha/\gamma}\}$ to conclude that, by choosing a subsequence if necessary, there exists an $\mathcal{R}$-minimizing solution q_* of (1.1) such that $q^\delta_{\alpha_*} \xrightarrow{\tau_{\mathcal{Q}}} q_*$ and $\mathcal{R}(q^\delta_{\alpha_*}) \to \mathcal{R}(q_*) = \mathcal{R}(q^\dagger)$ as $\delta \to 0$.

If $q^\dagger$ is the unique $\mathcal{R}$-minimizing solution (1.1), then the above argument shows that every sequence from $\{q^\delta_{\alpha_*}\}$ has a subsequence converging to $q^\dagger$. Using a subsequence-subsequence argument, the whole sequence $\{q^\delta_{\alpha_*}\}_{\delta>0}$ must converge to $q^\dagger$. The proof is complete. $\square$

3. Applications

In this section we will address the choices of the regularization functional $\mathcal{R}$ and the data misfit term $\mathcal{S}$ with applications of the theory developed in section 2.

The regularization functional $\mathcal{R}$ can be chosen in various manners depending on a priori information on the sought parameters. If the parameter to be reconstructed is a function defined on a bounded domain Ω in $\mathbb{R}^n$, one may take ([32])

$$\mathcal{R}(q) = a\|q\|_{L^1(\Omega)} + b\|q\|_{L^2(\Omega)}^2 + c\int_\Omega |Dq|$$

with $0 < p < \infty$ and nonnegative a, b, c in which at least one of them is nonzero, where

$$\int_\Omega |Dq| := \sup\left\{\int_\Omega q \operatorname{div} f dx : f \in C_0^1(\Omega, \mathbb{R}^n) \text{ and } \|f\|_{L^\infty(\Omega)} \le 1\right\}$$

denotes the total variation of q on Ω. If the sought parameter is a finite or infinite sequence, one may take ([3])

$$\mathcal{R}(q) = a\|q\|_0 + b\|q\|_{\ell^p}^p + c\|q\|_{TV}$$

with nonnegative a, b, c in which at least one of them is nonzero, where $\|q\|_0$ denotes the "0-norm" counting the number of nonzero elements in q and $\|q\|_{TV}$ denotes a discrete total variation of q which can be defined in various ways. One can even take

$$\mathcal{R}(q) = \|q\|_{\ell^1} - \eta\|q\|_{\ell^2},$$

with $0 \le \eta < 1$, for sparsity recovery ([6, 8, 35]).

The choices of the data misfit term $\mathcal{S}$ depend on the modelling of inverse problems and the type of noise which corrupts the data. In case an inverse problem can be formulated in the form (1.2) with $\mathcal{U}$ being a Banach space, one may take $\mathcal{S}(q, u) = \|F(q) - u\|^r$ with $r > 0$. For inverse problems of the form (1.2) with $F : \mathcal{Q} \to \mathcal{U}$ being a mapping from a topological space $\mathcal{Q}$ to a topological space $\mathcal{U}$, one may take

$$\mathcal{S}(q, u) = \Phi(F(q), u),$$

with $\Phi : \mathcal{U} \times \mathcal{U} \to [0, \infty]$ satisfying suitable properties. One may refer to [10, 30] for numerous choices of Φ including the Kullback-Leibler functional.

In the following we will provide further examples by discussing two groups of inverse problems for partial differential equations.

3.1. A convex framework for elliptic parameter estimation.

We will provide a unified framework for treating parameter estimation in a class of elliptic problems which have been considered in [14, 15, 16, 20, 31, 36] individually. To this end, we consider the elliptic problems whose weak formulation takes the form

$$a(u, v; q) + b(u, v) = \ell(v), \quad \forall v \in V, \tag{3.1}$$

where V is a Hilbert space, $\ell \in V^*$, i.e. ℓ is a bounded linear functional on V, q is a parameter belonging to a Banach space Z, and $a(\cdot, \cdot; \cdot)$ and $b(\cdot, \cdot)$ satisfy the following conditions:

(A1) $a(\cdot, \cdot; \cdot)$ is trilinear over $V \times V \times Z$, and $b(\cdot, \cdot)$ is bilinear over $V \times V$. Both a and b are symmetric with respect to the first two arguments.

(A2) There exists a constant C_0 such that

$$|a(u,v;q)| \le C_0 \|u\|_V \|v\|_V \|q\|_Z \quad \text{and} \quad |b(u,v)| \le C_0 \|u\|_V \|v\|_V$$

for all $u, v \in V$ and $q \in Z$.

(A3) There is a bounded convex set $\mathcal{A} \subset Z$ and a constant $c_0 > 0$ such that

$$a(u,u;q) + b(u,u) \ge c_0 \|u\|_V^2$$

for all $u \in V$ and $q \in \mathcal{A}$.

(A4) There is a reflexive Banach space B such that $Z \hookrightarrow B$ and $\mathcal{A}$ is closed in B; moreover for any sequence $\{q_n\} \subset \mathcal{A}$ and $q \in \mathcal{A}$ satisfying $\|q_n - q\|_B \to 0$ there holds

$$\sup_{\|u\|_V \le \beta} |a(u,v;q_n - q)| \to 0 \qquad \text{as } n \to \infty$$

for each $v \in V$ and each constant $\beta > 0$.

We are interested in estimating the parameter $q \in \mathcal{A}$ from a measurement $\tilde{u}$ of u in V. We will use the minimizer

$$\tilde{q}_\alpha = \arg\min_{q \in \mathcal{A}} \{a(u(q) - \tilde{u}, u(q) - \tilde{u}; q) + b(u(q) - \tilde{u}, u(q) - \tilde{u}) + \alpha \mathcal{R}(q)\} \tag{3.2}$$

to approximate the sought parameter $q^\dagger \in \mathcal{A}$, where $\mathcal{R}(q)$ is a suitably chosen regularization functional, and $u(q) \in V$ denotes the unique solution of (3.1) for $q \in \mathcal{A}$, whose existence, according to the conditions (A1)–(A3), is guaranteed by the Lax-Milgram theorem; moreover

$$\|u(q)\|_V \le \frac{1}{c_0} \|\ell\|_{V^*}. \tag{3.3}$$

In order to apply the theory developed in Section 2, we take $\mathcal{U} = V$, $\mathcal{Q} = \mathcal{A}$, $\tau_\mathcal{U} =$ strong topology on V, $\tau_\mathcal{Q} =$ weak topology on B, and define $\mathcal{S} : \mathcal{Q} \to \mathcal{U}$ by

$$\mathcal{S}(q,u) := a(u(q) - u, u(q) - u; q) + b(u(q) - u, u(q) - u),$$

then the identification of $q^\dagger$ becomes the equation (1.1) and (3.2) becomes the form (1.4). Our heuristic rule then choose the regularization parameter $\alpha_* \in \Delta_\gamma$ by Rule 2.1 with

$$\Theta(\alpha, \tilde{u}) = \left(\frac{1}{\alpha} + A\right) [a(u(\tilde{q}_\alpha) - \tilde{u}, u(\tilde{q}_\alpha) - \tilde{u}; \tilde{q}_\alpha) + b(u(\tilde{q}_\alpha) - \tilde{u}, u(\tilde{q}_\alpha) - \tilde{u})]. \tag{3.4}$$

We need to check that $\mathcal{S}$ satisfies Assumption 2.1 (iii) and (iv) and Assumption 2.3.

The verification of Assumption 2.3 can be proceeded as follows. By using (A1), (A2) and the boundedness of $\mathcal{A}$ in Z there is a constant $C_1 \ge 1$ such that

$$\frac{1}{C_1} \|u(q) - u\|_V^2 \le \mathcal{S}(q,u) \le C_1 \|u(q) - u\|_V^2$$

for all $q \in \mathcal{A}$ and $u \in V$. Therefore, for any $q \in \mathcal{A}$ and $u \in V$ we have

$$\begin{aligned} S(q, u^\dagger) &\le C_1 \|u(q) - u^\dagger\|_V^2 = C_1 \|u(q) - u(q^\dagger)\|_V^2 \\ &\le 2C_1 \left(\|u(q^\dagger) - u\|_V^2 + \|u(q) - u\|_V^2\right) \\ &\le 2C_1^2 \left(\mathcal{S}(q^\dagger, u) + \mathcal{S}(q,u)\right) \end{aligned}$$

which shows Assumption 2.3.

To check Assumption 2.1 (iv) for $\mathcal{S}$, we note that

$$\begin{aligned}\mathcal{S}(q,\bar{u}) - \mathcal{S}(q,u) &= a(u-\bar{u}, u-\bar{u}; q) + b(u-\bar{u}, u-\bar{u}) + 2a(u-\bar{u}, u(q)-u; q)\\ &\quad + 2b(u-\bar{u}, u(q)-u)\end{aligned}$$

for any $q \in \mathcal{A}$ and $u, \bar{u} \in V$. Therefore, it follows from (A2) that

$$\begin{aligned}&|\mathcal{S}(q,\bar{u}) - \mathcal{S}(q,u)|\\ &\le C_0(\|q\|_Z + 1)\|u-\bar{u}\|_V^2 + 2C_0(\|q\|_Z+1)\|u-\bar{u}\|_V \|u(q)-u\|_V \end{aligned} \tag{3.5}$$

which immediately implies that, for each fixed $q \in \mathcal{A}$, the function $u \to \mathcal{S}(q,u)$ is $\tau_{\mathcal{U}}$-continuous, and hence Assumption 2.1 (iv) is verified.

The verification of Assumption 2.1 (iii) is more subtle. We will achieve it by showing that for each fixed $u \in V$, the function $q \to \mathcal{S}(q,u)$ on $\mathcal{A}$ is convex and continuous with respect to the strong topology on B. To this end, we need a series of lemmas.

Given $q \in \mathcal{A}$ we can define a mapping $u'(q): Z \to V$ such that, for each $h \in Z$, $\eta := u'(q)h$ is a solution of the problem

$$a(\eta, v; q) + b(\eta, v) = -a(u(q), v; h), \qquad \forall v \in V. \tag{3.6}$$

According to the given conditions (A1)–(A3), this $\eta := u'(q)h$ is uniquely defined with the property

$$\|u'(q)h\|_V \le \frac{C_0}{c_0}\|u(q)\|_V \|h\|_Z \le \frac{C_0}{c_0^2}\|\ell\|_{V^*}\|h\|_Z.$$

Therefore $u'(q): Z \to V$ is a bounded linear map.

Lemma 3.1. *For any $p, q \in \mathcal{A}$ there holds*

$$\|u(p) - u(q) - u'(q)(p-q)\|_V \le \frac{C_0}{c_0^3}\|\ell\|_{V^*}\|p-q\|_Z^2.$$

Proof. To see this, we recall that

$$\begin{aligned} a(u(q), v; q) + b(u(q), v) &= \ell(v), \qquad \forall v \in V,\\ a(u(p), v; p) + b(u(p), v) &= \ell(v), \qquad \forall v \in V.\end{aligned}$$

From these two equations we can derive that

$$a(u(p)-u(q), v; p) + b(u(p)-u(q), v) = -a(u(q), v; p-q).$$

In view of the definition of $u'(q)(p-q)$ we then have

$$\begin{aligned}&a(u(p)-u(q)-u'(q)(p-q), v; p) + b(u(p)-u(q)-u'(q)(p-q), v)\\ &= -a(u'(q)(p-q), v; p-q).\end{aligned}$$

By taking $v = u(q+h) - u(q) - u'(q)(p-q)$ we can obtain

$$\begin{aligned}c_0\|u(q+h)-u(q)-u'(q)(p-q)\|_V &\le C_0\|u'(q)(p-q)\|_V\|p-q\|_Z\\ &\le \frac{C_0}{c_0^2}\|\ell\|_{V^*}\|p-q\|_Z^2.\end{aligned}$$

The proof is thus complete. □

Lemma 3.2. *For each $u \in V$, the function $q \to \mathcal{S}(q,u)$ is a convex function on the convex set $\mathcal{A}$.*

Proof. For any two points $q_0, q_1 \in \mathcal{A}$ let $q_t := (1-t)q_0 + tq_1$ for $0 \le t \le 1$. It follows from Lemma 3.1 that

$$\frac{d}{dt} u(q_t) = u'(q_t)h, \tag{3.7}$$

where $h := q_1 - q_0$. Therefore, by using the definition of $\mathcal{S}(q_t, u)$ and (3.6) we have

$$\begin{aligned}\frac{d}{dt}\mathcal{S}(q_t, u) &= a(u(q_t) - u, u(q_t) - u; h) + 2a(u'(q_t)h, u(q_t) - u; q_t)\\ &\quad + 2b(u'(q_t)h, u(q_t) - u)\\ &= a(u(q_t) - u, u(q_t) - u; h) - 2a(u(q_t), u(q_t) - u; h)\\ &= a(u, u; h) - a(u(q_t), u(q_t); h). \end{aligned}\tag{3.8}$$

Consequently, by using (3.7), (3.6) and (A3) we further have

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{S}(q_t, u) &= -2a(u(q_t), u'(q_t)h; h)\\ &= 2\left[a(u'(q_t)h, u'(q_t)h; h) + b(u'(q_t)h, u'(q_t)h)\right]\\ &\ge 2c_0\|u'(q_t)h\|_V^2 \ge 0.\end{aligned}$$

This implies that $\mathcal{S}((1-t)q_0+tq_1, u) \le (1-t)\mathcal{S}(q_0, u)) + t\mathcal{S}(q_1, u)$ for any $q_0, q_1 \in \mathcal{A}$ and $0 \le t \le 1$. Therefore $q \to \mathcal{S}(q, u)$ is convex over $\mathcal{A}$. □

Lemma 3.3. *For each $u \in V$ the function $q \to \mathcal{S}(q, u)$ is continuous on the set $\mathcal{A}$ with respect to the strong topology in B.*

Proof. Let $\{q_n\} \subset \mathcal{A}$ and $q \in \mathcal{A}$ be such that $\|q_n - q\|_B \to 0$ as $n \to \infty$. Since $\mathcal{A}$ is a bounded set in Z, we may use (3.3) to conclude that $\{\|u(q_n)\|_V\}$ is a bounded sequence. Note that (A2) and (A3) imply that

$$|\!|\!| v |\!|\!| := \sqrt{a(v, v; q) + b(v, v)}, \qquad v \in V$$

is an equivalent norm on V under which $(V, |\!|\!|\cdot|\!|\!|)$ becomes a Hilbert space. Thus, by taking a subsequence if necessary, we may assume that $\{u(q_n)\}$ converges weakly to some u in $(V, |\!|\!|\cdot|\!|\!|)$, i.e.

$$a(u(q_n), v; q) + b(u(q_n), v) \to a(u, v; q) + b(u, v)$$

as $n \to \infty$ for all $v \in V$. In view of the definition of $u(q_n)$ and the linearity of $q \to a(\cdot, \cdot; q)$ we have

$$\begin{aligned}\ell(v) = a(u(q_n), v; q_n) + b(u(q_n), v) &= a(u(q_n), v; q) + b(u(q_n), v)\\ &\quad + a(u(q_n), v; q_n - q).\end{aligned}$$

By taking $n \to \infty$ and using (A4) we can obtain

$$\ell(v) = a(u, v; q) + b(u, v), \qquad \forall v \in V.$$

This shows that $u = u(q)$ and thus $\{u(q_n)\}$ converges weakly to $u(q)$ in $(V, |\!|\!|\cdot|\!|\!|)$, i.e.

$$a(u(q_n), v; q) + b(u(q_n), v) \to a(u(q), v; q) + b(u(q), v), \quad \forall v \in V.$$

Moreover, since $\ell \in V^*$, it is also a bounded linear functional on $(V, \|\!|\cdot\|\!|)$ and hence $\ell(u(q_n)) \to \ell(u(q))$. Using these facts, (A1) and (A4) we can derive that

$$\begin{aligned}
\mathcal{S}(q_n, u) &= a(u(q_n) - u, u(q_n) - u; q_n) + b(u(q_n) - u, u(q_n) - u) \\
&= a(u(q_n), u(q_n) - u; q_n) + b(u(q_n), u(q_n) - u) \\
&\quad - a(u, u(q_n) - u; q) - b(u, u(q_n) - u) \\
&\quad - a(u, u(q_n) - u; q_n - q) \\
&= \ell(u(q_n) - u) - a(u(q_n) - u, u; q) - b(u(q_n) - u, u) \\
&\quad - a(u(q_n) - u, u; q_n - q) \\
&\to \ell(u(q) - u) - a(u(q) - u, u; q) - b(u(q) - u, u) \\
&= \ell(u(q) - u) - a(u, u(q) - u; q) - b(u, u(q) - u).
\end{aligned} \tag{3.9}$$

According to the definition of $u(q)$ we have

$$\ell(u(q) - u) = a(u(q), u(q) - u; q) + b(u(q), u(q) - u).$$

Therefore we can obtain from (3.9) that $\mathcal{S}(q_n, u) \to \mathcal{S}(q, u)$ as $n \to \infty$. Thus $q \to \mathcal{S}(q, u)$ is continuous on $\mathcal{A}$ with respect to the strong topology in B. □

Lemma 3.4. *For each $u \in V$ the function $q \to \mathcal{S}(q, u)$ is $\tau_{\mathcal{Q}}$ lower semi-continuous on $\mathcal{A}$.*

Proof. Since $\mathcal{A}$ is convex and closed in B and since, by Lemma 3.2 and Lemma 3.3, $q \to \mathcal{S}(q, u)$ is convex and continuous on $\mathcal{A}$ with respect to the strong topology in B, the $\tau_{\mathcal{Q}}$ lower semi-continuity of $q \to \mathcal{S}(q, u)$ then follows from a well-known fact in functional analysis. □

Now we are ready to verify Assumption 2.1 (iii) for $\mathcal{S}$. Let $\{q_n\} \subset \mathcal{A}$ and $\{u_n\} \subset V$ be such that $q_n \xrightarrow{\tau_{\mathcal{Q}}} q \in \mathcal{A}$ and $u_n \xrightarrow{\tau_{\mathcal{U}}} u \in V$. We write

$$\mathcal{S}(q_n, u_n) - \mathcal{S}(q, u) = [\mathcal{S}(q_n, u_n) - \mathcal{S}(q_n, u)] + [\mathcal{S}(q_n, u) - \mathcal{S}(q, u)]. \tag{3.10}$$

According to (3.5) we have

$$|\mathcal{S}(q_n, u_n) - \mathcal{S}(q_n, u)| \le C_0(\|q_n\|_Z + 1)\left(\|u_n - u\|_V + \|u(q_n) - u\|_V\right)\|u_n - u\|_V.$$

Since $\{q_n\} \subset \mathcal{A}$ is a bounded, $\{\|u(q_n)\|_V\}$ is also bounded by virtue of (3.3). Thus, by using $u_n \xrightarrow{\tau_{\mathcal{U}}} u$ we can conclude that

$$\lim_{n\to\infty} (\mathcal{S}(q_n, u_n) - \mathcal{S}(q_n, u)) = 0.$$

Consequently, it follows from (3.10) and Lemma 3.4 that

$$\begin{aligned}
&\liminf_{n\to\infty} (\mathcal{S}(q_n, u_n) - \mathcal{S}(q, u)) \\
&\quad = \lim_{n\to\infty} (S(q_n, u_n) - \mathcal{S}(q_n, u)) + \liminf_{n\to\infty} (S(q_n, u) - \mathcal{S}(q, u)) \\
&\quad = \liminf_{n\to\infty} (S(q_n, u) - \mathcal{S}(q, u)) \ge 0,
\end{aligned}$$

i.e. $\mathcal{S}(q, u) \le \liminf_{n\to\infty} \mathcal{S}(q_n, u_n)$ and thus Assumption 2.1 (iii) is verified.

Example 3.1. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$. We consider the determination of q in the diffusion problem

$$\begin{aligned}
-\mathrm{div}(q\nabla u) &= f \quad \text{in } \Omega, \\
u &= 0 \quad \text{on } \partial\Omega
\end{aligned} \tag{3.11}$$

from an measurement $\tilde{u}$ of u in $H_0^1(\Omega)$. If we take $V = H_0^1(\Omega)$, $Z = L^\infty(\Omega)$, $B = L^2(\Omega)$, and define $\mathcal{A} = \{q \in L^\infty(\Omega) : \gamma_0 \le q \le \gamma_1 \text{ a.e. }\}$ for some positive constants γ_0 and γ_1, then the elliptic problem (3.11) can be formulated into the form (3.1) with

$$a(u,v;q) = \int_\Omega q\nabla u \cdot \nabla v dx, \quad b(u,v) = 0, \quad \ell(v) = \int_\Omega f v dx,$$

and the inverse problem takes the form (1.1) with

$$S(q,u) = \int_\Omega q|\nabla(u(q) - u)|^2.$$

It is easy to see that (A1)–(A3) hold. For verifying (A4), we first use the Cauchy-Schwarz inequality and the Poincaré inequality to obtain

$$|a(u,v;q_n - q)| \le C\|u\|_V \left(\int_\Omega |q_n - q|^2 |\nabla v|^2\right)^{1/2}$$

for some universal constant C. Thus

$$\sup_{\|u\|_V \le \beta} |a(u,v;q_n - q)| \le C\beta \left(\int_\Omega |q_n - q|^2 |\nabla v|^2\right)^{1/2}.$$

By virtue of $\|q_n - q\|_{L^2(\Omega)} \to 0$ and $\{q_n\} \subset \mathcal{A}$, we may use the Lebesgue dominated convergence theorem to conclude

$$\int_\Omega |q_n - q|^2 |\nabla v|^2 \to 0 \quad \text{ as } n \to \infty.$$

We thus verify (A4).

Example 3.2. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$. We consider the estimation of the parameters $q = (q_1, q_2)$ in the problem

$$-\text{div}(q_1 \nabla u) + q_2 u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega \tag{3.12}$$

from a measured data $\tilde{u}$ of u in $H^1(\Omega)$. If we take $V = H_0^1(\Omega)$, $Z = L^\infty(\Omega) \times L^\infty(\Omega)$, $B = L^2(\Omega) \times L^2(\Omega)$ and define

$$\mathcal{A} = \{(q_1, q_2) \in L^\infty(\Omega) \times L^\infty(\Omega) : \gamma_0 \le q_1 \le \gamma_1 \text{ and } \gamma_2 \le q_2 \le \gamma_3 \text{ a. e. }\}$$

for some positive constants $\gamma_0, \gamma_1, \gamma_2$ and γ_3, then the elliptic problem has the weak formulation (3.1) with

$$a(u,v;(q_1,q_2)) = \int_\Omega (q_1 \nabla u \cdot \nabla v + q_2 u v) dx, \quad b(u,v) = 0, \quad \ell(v) = \int_\Omega f v dx$$

and the inverse problem takes the form (1.1) with

$$\mathcal{S}((q_1,q_2),u) = \int_\Omega \left(q_1 |\nabla(u(q_1,q_2) - u)|^2 + q_2 |u(q_1,q_2) - u|^2\right),$$

where $u(q_1,q_2) \in H_0^1(\Omega)$ denotes the unique solution of (3.12) for given $(q_1,q_2) \in \mathcal{A}$. It is easy to see that (A1)–(A4) hold.

3.2. **Electrical impedance tomography.** Let $\Omega \subset \mathbb{R}^d$ be a bounded open domain with Lipschitz boundary $\partial\Omega$, let ν denote the unit outward normal on $\partial\Omega$, and let $\rho \in H^{-1}(\Omega)$ be given. We consider the electrical impedance tomography which consists in determining the conductivity q in the elliptic equation

$$-\mathrm{div}(q\nabla u) = \rho \quad \text{in } \Omega \tag{3.13}$$

by virtue of a family of Cauchy data

$$(f_\ell, g_\ell), \qquad \ell = 1, \cdots, L,$$

where $f_\ell = u|_{\partial\Omega}$ and $g_\ell = q\frac{\partial u}{\partial \nu}\big|_{\partial\Omega}$ for some $u \in H^1(\Omega)$ satisfying (3.13). We assume that

$$q \in \mathcal{Q} := \{q \in L^\infty(\Omega) : \underline{q} \le q(x) \le \bar{q} \text{ a.e. in } \Omega\}$$

with known positive constants $\underline{q}$ and $\bar{q}$.

We will adopt the variational approach of Kohn and Vogelius [26, 27, 28] to identify q. To this end, for a given $f \in H^{1/2}(\partial\Omega)$ we use $u_D[q, f] \in H^1(\Omega)$ to denote the unique weak solution of

$$-\mathrm{div}(q\nabla u) = \rho \text{ in } \Omega, \quad u = f \text{ on } \partial\Omega, \tag{3.14}$$

i.e. $u = f$ on $\partial\Omega$ and

$$\int_\Omega q\nabla u \cdot \nabla \phi = \int_\Omega \rho\phi, \quad \forall \phi \in H_0^1(\Omega),$$

where

$$H_0^1(\Omega) = \{u \in H^1(\Omega) : u = 0 \text{ on } \partial\Omega\}.$$

For a given $g \in H^{-1/2}(\partial\Omega)$ we also use $u_N[q, g] \in H^1_\diamond(\Omega)$ to denote the unique weak solution of

$$-\mathrm{div}(q\nabla v) = \rho \text{ in } \Omega, \quad q\frac{\partial v}{\partial \nu} = g \text{ on } \partial\Omega \tag{3.15}$$

with vanishing boundary mean, i.e.

$$\int_\Omega q\nabla v \cdot \nabla \psi = \int_\Omega \rho\psi + \int_{\partial\Omega} g\psi, \quad \forall \psi \in H^1_\diamond(\Omega), \tag{3.16}$$

where

$$H^1_\diamond(\Omega) := \left\{u \in H^1(\Omega) : \int_{\partial\Omega} u = 0\right\}.$$

According to the theory of elliptic equations, $u_D[q, f]$ and $u_N[q, g]$ are well-defined and

$$\begin{aligned} \|u_D[q, f]\|_{H^1(\Omega)} &\le C_D \left(\|\rho\|_{H^{-1}(\Omega)} + \|f\|_{H^{1/2}(\partial\Omega)}\right), \\ \|u_N[q, g]\|_{H^1(\Omega)} &\le C_N \left(\|\rho\|_{H^{-1}(\Omega)} + \|g\|_{H^{-1/2}(\partial\Omega)}\right) \end{aligned} \tag{3.17}$$

for some universal constants C_D and C_N depending only on $\underline{q}$, $\bar{q}$ and Ω.

Now we define the mapping $\mathcal{S} : \mathcal{Q} \times (H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega))^L \to [0, \infty)$ by

$$\mathcal{S}(q, \{(f_\ell, g_\ell)\}_{\ell=1}^L) := \sum_{\ell=1}^L \int_\Omega q|\nabla(u_D[q, f_\ell] - u_N[q, g_\ell])|^2$$

for $q \in \mathcal{Q}$ and $\{(f_\ell, g_\ell)\}_{\ell=1}^L \in (H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega))^L$. Then for the exactly given Cauchy data $\{(f_\ell^\dagger, g_\ell^\dagger)\}_{\ell=1}^L \in (H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega))^L$, the electrical impedance tomography amounts to solving

$$\mathcal{S}(q, \{(f_\ell^\dagger, g_\ell^\dagger)\}_{\ell=1}^L) = 0 \quad \text{in } \mathcal{Q}.$$

In case only measurement data $\{(\tilde{f}_\ell, \tilde{g}_\ell)\}_{\ell=1}^L \in (H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega))^L$ is available, we may construct the conductivity $q \in \mathcal{Q}$ by considering the minimization problem

$$\tilde{q}_\alpha \in \arg\min_{q\in\mathcal{Q}} \left\{ \sum_{\ell=1}^L \int_\Omega q|\nabla(u_D[q, \tilde{f}_\ell] - u_N[q, \tilde{g}_\ell])|^2 + \alpha\mathcal{R}(q) \right\} \tag{3.18}$$

with a suitably chosen regularization functional $\mathcal{R} : \mathcal{Q} \to [0, \infty)$. This minimization problem clearly takes the form (1.4). The regularization property of (3.18) has been analyzed recently in [19] under a priori choice of the regularization parameter. Our heuristic rule chooses the regularization parameter as

$$\alpha_* \in \arg\min_{\alpha\in\Delta_\gamma} \left\{ \left(\frac{1}{\alpha} + A\right) \sum_{\ell=1}^L \int_\Omega \tilde{q}_\alpha |\nabla(u_D[\tilde{q}_\alpha, \tilde{f}_\ell] - u_N[\tilde{q}_\alpha, \tilde{g}_\ell])|^2 \right\}.$$

In order to apply the convergence result for Rule 2.1 in Section 2, we are going to show that if $\{q^{(n)}\} \subset \mathcal{Q}$ and $\{(f_\ell^{(n)}, g_\ell^{(n)})\} \subset H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega)$ satisfying $\|q^{(n)} - q\|_{L^1(\Omega)} \to 0$ and $\|f_\ell^{(n)} - f_\ell\|_{H^{1/2}(\partial\Omega)} + \|g_\ell^{(n)} - g_\ell\|_{H^{-1/2}(\partial\Omega)} \to 0$ as $n \to \infty$ for $\ell = 1, \cdots, L$, then

$$\lim_{n\to\infty} \mathcal{S}(q^{(n)}, \{(f_\ell^{(n)}, g_\ell^{(n)})\}_{\ell=1}^L) = \mathcal{S}(q, \{(f_\ell, g_\ell)\}_{\ell=1}^L)$$

which implies that Assumption 2.1 (iii) and (iv) hold with $\mathcal{U} := (H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega))^L$, $\tau_\mathcal{Q}$:= strong topology on $L^1(\Omega)$ and $\tau_\mathcal{U}$:= strong topology on $\mathcal{U}$. To see this, it suffices to show that if $\{q^{(n)}\} \subset \mathcal{Q}$ and $\{(f^{(n)}, g^{(n)})\} \subset H^{1/2}(\partial\Omega) \times H^{-1/2}(\partial\Omega)$ satisfy $\|q^{(n)} - q\|_{L^1(\Omega)} \to 0$ and $\|f^{(n)} - f\|_{H^{1/2}(\partial\Omega)} + \|g^{(n)} - g\|_{H^{-1/2}(\partial\Omega)} \to 0$ as $n \to \infty$, then

$$\lim_{n\to\infty} \Gamma(q^{(n)}, (f^{(n)}, g^{(n)})) = \Gamma(q, (f, g)), \tag{3.19}$$

where

$$\Gamma(q, (f, g)) := \int_\Omega q|\nabla(u_D[q, f] - u_N[q, g])|^2.$$

We first show that

$$\|u_D[q^{(n)}, f] - u_D[q, f]\|_{H^1(\Omega)} \to 0, \quad \|u_N[q^{(n)}, g] - u_N[q, g]\|_{H^1(\Omega)} \to 0 \tag{3.20}$$

as $n \to \infty$. By the definition of $u_D[q^{(n)}, f]$ and $u_D[q, f]$ we have $u_D[q^{(n)}, f] - u_D[q, f] \in H_0^1(\Omega)$ and

$$\int_\Omega q^{(n)} \nabla u_D[q^{(n)}, f] \cdot \nabla\phi = \int_\Omega \rho\phi = \int_\Omega q\nabla u_D[q, f] \cdot \nabla\phi$$

for any $\phi \in H_0^1(\Omega)$. Therefore

$$\begin{aligned} 0 &= \int_\Omega q^{(n)} \nabla u_D[q^{(n)}, f] \cdot \nabla\phi - \int_\Omega q\nabla u_D[q, f] \cdot \nabla\phi \\ &= \int_\Omega q^{(n)} \nabla \left(u_D[q^{(n)}, f] - u_D[q, f] \right) \cdot \nabla\phi + \int_\Omega (q^{(n)} - q)\nabla u_D[q, f] \cdot \nabla\phi. \end{aligned}$$

By taking $\phi = u_D[q^{(n)}, f] - u_D[q, f]$ in the above equation, using the condition $q^{(n)} \geq \underline{q}$, and the Cauchy-Schwartz inequality we have

$$\begin{aligned}
&\underline{q}\int_\Omega |\nabla(u_D[q^{(n)}, f] - u_D[q, f])|^2 \leq \int_\Omega q^{(n)}|\nabla(u_D[q^{(n)}, f] - u_D[q, f])|^2 \\
&\quad = \int_\Omega (q - q^{(n)})\nabla u_D[q, f] \cdot \nabla(u_D[q^{(n)}, f] - u_D[q, f]) \\
&\quad \leq \left(\int_\Omega |\nabla(u_D[q^{(n)}, f] - u_D[q, f])|^2\right)^{1/2} \left(\int_\Omega |q^{(n)} - q|^2 |\nabla u_D[q, f]|^2\right)^{1/2}
\end{aligned}$$

Consequently

$$\underline{q}^2 \int_\Omega |\nabla(u_D[q^{(n)}, f] - u_D[q, f])|^2 \leq \int_\Omega |q^{(n)} - q|^2 |\nabla u_D[q, f]|^2.$$

Since $\{q^{(n)}\} \subset \mathcal{Q}$ and $\|q^{(n)} - q\|_{L^1(\Omega)} \to 0$, by the dominated convergence theorem we can show that the right hand side converges to zero as $n \to 0$. Therefore, we may use the Poincare inequality to conclude the first result in (3.20). The second result in (3.20) can be proved in a similar way.

Next we will show that

$$\lim_{n\to\infty} \Gamma(q^{(n)}, (f, g)) = \Gamma(q, (f, g)). \tag{3.21}$$

We first write

$$\begin{aligned}
&\Gamma(q^{(n)}, (f, g)) - \Gamma(q, (f, g)) \\
&= \int_\Omega q^{(n)}|\nabla(u_D[q^{(n)}, f] - u_N[q^{(n)}, g])|^2 - \int_\Omega q|\nabla(u_D[q, f] - u_N[q, g])|^2 \\
&= I_1^{(n)} + I_2^{(n)} + I_3^{(n)},
\end{aligned} \tag{3.22}$$

where

$$\begin{aligned}
I_1^{(n)} &= \int_\Omega q^{(n)}|\nabla u_D[q^{(n)}, f]|^2 - \int_\Omega q|\nabla u_D[q, f]|^2, \\
I_2^{(n)} &= -2\int_\Omega q^{(n)}\nabla u_D[q^{(n)}, f] \cdot \nabla u_N[q^{(n)}, g] + 2\int_\Omega q\nabla u_D[q, f] \cdot \nabla u_N[q, g], \\
I_3^{(n)} &= \int_\Omega q^{(n)}|\nabla u_N[q^{(n)}, g]|^2 - \int_\Omega q|\nabla u_N[q, g]|^2.
\end{aligned}$$

Note that

$$\begin{aligned}
I_1^{(n)} = &\int_\Omega (q^{(n)} - q)|\nabla u_D[q, f]|^2 \\
&+ \int_\Omega q^{(n)}\nabla(u_D[q^{(n)}, f] - u_D[q, f]) \cdot \nabla(u_D[q^{(n)}, f] + u_D[q, f]),
\end{aligned}$$

we have

$$\begin{aligned}
|I_1^{(n)}| \leq &\int_\Omega |q^{(n)} - q||\nabla u_D[q, f]|^2 \\
&+ \bar{q}\left(\|u_D[q^{(n)}, f]\|_{H^1(\Omega)} + \|u_D[q, f]\|_{H^1(\Omega)}\right)\|u_D[q^{(n)}, f] - u_D[q, f]\|_{H^1(\Omega)}.
\end{aligned}$$

The first term converges to zero by the dominated converhence theorem and the second term converges to zero by the first result in (3.20). Hence $I_1^{(n)} \to 0$ as

$n \to \infty$. By a similar argument we can also show that $I_3^{(n)} \to 0$ as $n \to \infty$. For the term $I_2^{(n)}$ we can write

$$\begin{aligned} I_2^{(n)} &= 2\int_\Omega q^{(n)}\nabla(u_D[q,f] - u_D[q^{(n)},f]) \cdot \nabla u_N[q^{(n)},g] \\ &\quad + 2\int_\Omega q^{(n)}\nabla u_D[q,f] \cdot \nabla(u_N[q,g] - u_N[q^{(n)},g]) \\ &\quad + 2\int_\Omega (q - q^{(n)})\nabla u_D[q,f] \cdot \nabla u_N[q,g]. \end{aligned}$$

Therefore

$$\begin{aligned} |I_2^{(n)}| &\le 2\bar{q}\|u_N[q^{(n)},g]\|_{H^1(\Omega)}\|u_D[q,f] - u_D[q^{(n)},f]\|_{H^1(\Omega)} \\ &\quad + 2\bar{q}\|u_D[q,f]\|_{H^1(\Omega)}\|u_N[q,g] - u_N[q^{(n)},g]\|_{H^1(\Omega)} \\ &\quad + 2\int_\Omega |q - q^{(n)}||\nabla u_D[q,f]||\nabla u_N[q,g]|. \end{aligned}$$

By using (3.20), the first two terms on the right hand side converge to zero, and, by the dominated convergence theorem, the last term on the right hand side also converges to zero. Therefore $I_2^{(n)} \to 0$ as $n \to \infty$. Combining the above results with (3.22) we thus obtain (3.21).

Finally we show (3.19). According to (3.21), it remains only to show that

$$\lim_{n\to\infty} \Big(\Gamma(q^{(n)},(f^{(n)},g^{(n)})) - \Gamma(q^{(n)},(f,g))\Big) = 0. \tag{3.23}$$

Let $v^{(n)} := u_D[q^{(n)},f^{(n)}] - u_D[q^{(n)},f]$ and $w^{(n)} := u_N[q^{(n)},g^{(n)}] - u_N[q^{(n)},g]$. Then $v^{(n)} \in H^1(\Omega)$ and $w^{(n)} \in H^1_\diamond(\Omega)$ are weak solutions of

$$\begin{cases} -\operatorname{div}(q^{(n)}\nabla v^{(n)}) = 0 & \text{in } \Omega \\ v^{(n)} = f^{(n)} - f & \text{on } \partial\Omega \end{cases} \quad \text{and} \quad \begin{cases} -\operatorname{div}(q^{(n)}\nabla w^{(n)}) = 0 & \text{in } \Omega \\ q^{(n)}\frac{\partial w^{(n)}}{\partial \nu} = g^{(n)} - g & \text{on } \partial\Omega \end{cases}$$

respectively. Thus, it follows from (3.17) that

$$\begin{aligned} \|v^{(n)}\|_{H^1(\Omega)} &\le C_D\|f^{(n)} - f\|_{H^{1/2}(\partial\Omega)}, \\ \|w^{(n)}\|_{H^1(\Omega)} &\le C_N\|g^{(n)} - g\|_{H^{-1/2}(\partial\Omega)}. \end{aligned} \tag{3.24}$$

Note that

$$\begin{aligned} \Gamma(q^{(n)},(f^{(n)},g^{(n)})) - \Gamma(q^{(n)},(f,g)) &= \int_\Omega q^{(n)}\left|\nabla(u_D[q^{(n)},f^{(n)}] - u_N[q^{(n)},g^{(n)}])\right|^2 \\ &\quad - \int_\Omega q^{(n)}\left|\nabla(u_D[q^{(n)},f] - u_N[q^{(n)},g])\right|^2 \\ &= \int_\Omega q^{(n)}\nabla(v^{(n)} - w^{(n)}) \cdot \nabla\eta^{(n)}, \end{aligned}$$

where

$$\eta^{(n)} := u_D[q^{(n)},f^{(n)}] - u_N[q^{(n)},g^{(n)}] + u_D[q^{(n)},f] - u_N[q^{(n)},g].$$

By using (3.17) it is easy to see that $\|\eta^{(n)}\|_{H^1(\Omega)} \le C$ for a universal constant C independent of n. Therefore, we can use (3.24) to conclude that

$$\begin{aligned} &|\Gamma(q^{(n)},(f^{(n)},g^{(n)})) - \Gamma(q^{(n)},(f,g))| \\ &\le C\|\eta^{(n)}\|_{H^1(\Omega)}\|v^{(n)} - w^{(n)}\|_{H^1(\Omega)} \le C\left(\|f^{(n)} - f\|_{H^{\frac{1}{2}}(\partial\Omega)} + \|g^{(n)} - g\|_{H^{-\frac{1}{2}}(\partial\Omega)}\right). \end{aligned}$$

This shows (3.23).

4. Numerical simulations

In this section we will present numerical results to validate the theory developed in Section 2 by considering the inverse problems discussed in Section 3.1 and Section 3.2.

For the inverse problems discussed in Section 3.1, the regularized solutions are defined by (3.2) and our heuristic rule chooses the regularization parameter $\alpha_* \in \Delta_\gamma$ by Rule 2.1 with $\Theta(\alpha, \tilde{u})$ given by (3.4). Therefore we need to solve for $\alpha \in \Delta_\gamma$ the minimization problem (3.2). When B is a Hilbert space and the regularization functional $\mathcal{R}(q) = \phi(q, q)$ for a bounded symmetric bilinear form ϕ, we may use a Newton-type method to solve it. To be more precise, according to (3.8) and the convexity of $\mathcal{S}$, solving (3.2) is equivalent to finding $(u, q) \in V \times \mathcal{A}$ such that

$$\begin{aligned} &a(u, v; q) + b(u, v) = \ell(v), \qquad \forall v \in V, \\ &a(\tilde{u}, \tilde{u}; p - q) - a(u, u; p - q) + 2\alpha\phi(q, p - q) \geq 0, \qquad \forall p \in \mathcal{A}. \end{aligned} \tag{4.1}$$

We then solve (4.1) by applying the projected Newton method or the semi-smooth Newton method [4, 18, 23, 34]) which enables us to consider (4.1) first by replacing the inequality by equality and then by projecting the computational result for q back onto $\mathcal{A}$ after each iteration. This thus leads to Algorithm 1. In case $\mathcal{R}(q)$ is nonsmooth, e.g. $\mathcal{R}(q)$ is the total variation of q, we will use an iteratively reweighted least-squares (IRLS) algorithm [5, 29] and then apply the projected Newton method.

Algorithm 1 Projected Newton method for (4.1)

1: Given initial guess u_0, q_0.
2: **for** $k = 0, 1, 2, ...$ **do**
3: Obtain $u_{k+1}, \tilde{q}_{k+1}$ by solving

$$a(u_{k+1}, v; q_k) + b(u_{k+1}, v) + a(u_k, v; \tilde{q}_{k+1}) = \ell(v) + a(u_k, v; q_k), \ \forall\ v \in V,$$

$$2\alpha\phi(\tilde{q}_{k+1}, p) - 2a(u_k, u_{k+1}, p) = -a(\tilde{u}, \tilde{u}; p) - a(u_k, u_k, p), \ \forall p \in B.$$

4: Project $\tilde{q}_{k+1}$ onto $\mathcal{A}$ to obtain q_{k+1}.
5: Check stop condition.
6: **end for**

We remark that the projected Newton method is robust and efficient and enjoys a local superlinear convergence. Since our heuristic rule requires to compute $\tilde{q}_\alpha$ for a number of $\alpha \in \Delta_\gamma$, we adopt a path-following strategy: let $\alpha_n = \alpha_0\gamma^n$ and use Algorithm 1 on the decreasing sequences $\{\alpha_n\}$ with the solution $\tilde{q}_{\alpha_n}$ as an initial guess for computing $\tilde{q}_{\alpha_{n+1}}$. According to a remark after Rule 2.1 in Section 2, the constant A appearing in $\Theta(\alpha, \tilde{u})$ need to be large enough in order to avoid getting an unwanted large regularization parameter. In our computations, we will always choose $A = 10^5$.

In order to implement Algorithm 1 numerically, we need to discretize the variational equations. We may replace V and B by suitably chosen finite-dimensional spaces V_h and B_h and replace $\mathcal{A}$ by $\mathcal{A}_h := \mathcal{A} \cap B_h$ in Algorithm 1. When V and B are spaces consisting of functions defined on a bounded domain $\Omega \subset \mathbb{R}^n$, we may construct V_h and B_h by finite elements. Let $\mathcal{T}_h$ be a regular triangulation of Ω with maximum mesh size $h := \max_{T\in\mathcal{T}_h} \operatorname{diam}(T)$ and suppose that $\bar{\Omega}$ is the union of the elements of $\mathcal{T}_h$. We may take

$$\begin{aligned} V_h &:= \{v_h : v_h \in C(\bar{\Omega}), v_h|_T \in P_2(T)\ \forall\, T \in \mathcal{T}_h\}, \\ B_h &:= \{v_h : v_h \in C(\bar{\Omega}), v_h|_T \in P_1(T)\ \forall\, T \in \mathcal{T}_h\}, \\ \mathcal{A}_h &:= \mathcal{A} \cap B_h, \end{aligned} \tag{4.2}$$

where $P_l(T)$ denotes the spaces of polynomials of degree l defined on T.

In the following we will provide some numerical results for which the experiments are performed using FreeFem++ [17] with 128 grid points on $\partial\Omega$.

Example 4.1. We consider the inverse problem given in Example 3.1, where

$$\Omega = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}$$

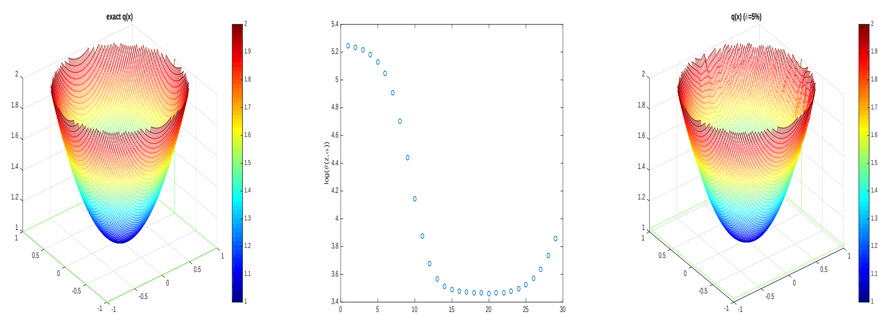
is the unit disk in $\mathbb{R}^2$. We first consider the case that the sought solution is smooth given by

$$q^\dagger(x) = 1 + x_1^2 + x_2^2.$$

Assume that $u^\dagger(x) = \sin(\pi(x_1^2 + x_2^2))$. Then f can be obtained via the elliptic equation. The noisy data $\tilde{u}$ is produced from $u^\dagger$ by adding random Gaussian noise with $\|\tilde{u} - u^\dagger\|_{L^2(\Omega)}/\|u^\dagger\|_{L^2(\Omega)} = 5\%$. Now we use $\tilde{u}$ to reconstruct $q^\dagger$. Since the sought solution is smooth, we take

$$\mathcal{R}(q) = \int_\Omega |\nabla q|^2 dx.$$

Then the corresponding minimization problem (3.2) can be solved by Algorithm 1.



(a) (b) (c)

Figure 1. The reconstruction for Example 4.1 with (a) exact parameter. (b) the relation between $\Theta(\alpha_n, \tilde{u})$ and n. (c) reconstruction result

In our computation, we take $\gamma_0 = 0.1$ and $\gamma_1 = 20$ in the definition of $\mathcal{A}$. In order to determine the regularization parameter by our heuristic rule, we use the set Δ_γ of grid points with $\alpha_0 = 10$ and $\gamma = 0.6$. Let $\alpha_n := 10 \times 0.6^n$. Then the relation between $\Theta(\alpha_n, \tilde{u})$ and n is plotted in Figure 1 (b). The plot demonstrates that $\Theta(\alpha, \tilde{u})$ achieves its minimum at $\alpha_* = 10 \times 0.6^{20}$. The corresponding regularized solution $\tilde{q}_{\alpha_*}$ is plotted in Figure 1 (c). Comparing with the exact parameter plotted in Figure 1 (a), the reconstruction result is clearly satisfactory.

Example 4.2. We consider again the inverse problem given in Example 3.1 with Ω being the unit disk in $\mathbb{R}^2$. In this example we assume the sought solution is piecewise constant given by

$$q^\dagger(x) = \begin{cases} 5, & x \in [-0.7, -0.2] \times [-0.2, 0.2], \\ 2, & x \in [0.2, 0.5] \times [-0.5, 0.1], \\ 1, & \text{otherwise.} \end{cases}$$

Assuming $f(x) = x_1 + x_2$, we may solve the elliptic problem to obtain the exact data $u^\dagger$. We then add random Gaussian noise to $u^\dagger$ to produce a noisy data $\tilde{u}$ with $\|\tilde{u} - u^\dagger\|_{L^2(\Omega)}/\|u^\dagger\|_{L^2(\Omega)} = 5\%$. We will use $\tilde{u}$ to reconstruct $q^\dagger$. Due to the a priori information on $q^\dagger$, we take the regularization functional $\mathcal{R}(q)$ to be the total variation of q over Ω, i.e.

$$\mathcal{R}(q) = \int_\Omega |Dq|.$$

Now the corresponding minimization problem (3.2) can not be solved by Algorithm 1 directly. Instead we use an IRLS strategy to reduce (3.2) to a sequence of minimization problems in which each subproblem can be solved by Algorithm 1. More precisely, we take a sequence of small positive numbers $\{\varepsilon_j\}$; with an initial guess q_0, we then define

$$q_{j+1} \in \arg\min_{q \in \mathcal{A}} \left\{ \int_\Omega q |\nabla(u(q) - \tilde{u})|^2 dx + \alpha \int_\Omega \omega_j |\nabla q|^2 dx \right\}, \tag{4.3}$$

where $\omega_j := (|\nabla q_j|^2 + \varepsilon_j^2)^{-1/2}$. In summary, we have the following algorithm.

Algorithm 2 The IRLS algorithm for Example 4.2

1: Given an initial guess q_0, let $\omega_0 = (|\nabla q_0|^2 + \varepsilon_0^2)^{-\frac{1}{2}}$.
2: **for** $j = 0, 1, ...$ **do**
3: Find q_{j+1} from (4.3) by Algorithm 1;
4: Let $\omega_{j+1} = (|\nabla q_{j+1}|^2 + \varepsilon_{j+1}^2)^{-\frac{1}{2}}$;
5: **end for**

When using our heuristic rule to determine the regularization parameter, we use the set Δ_γ with $\alpha_0 = 1.0$ and $\gamma = 0.5$. For $\alpha_n := 1.0 \times 0.5^n$ the regularized solutions $\tilde{q}_{\alpha_n}$ are computed by Algorithm 2 with $q_0 = 1$ and $\varepsilon_j = 0.01$ for all j. The constants γ_0 and γ_1 appearing in $\mathcal{A}$ are taken to be 0.1 and 20 respectively. The relation between $\Theta(\alpha_n, \tilde{u})$ and n is plotted in Figure 2 (a) which demonstrates that the regularization parameter determined by our heuristic rule is $\alpha_* = 1.0 \times 0.5^{18}$. The exact parameter $q^\dagger$ and the reconstruction result $\tilde{q}_{\alpha_*}$ are plotted in Figure 2 (b) and (c) respectively. It is clear that the reconstructed parameter is a good approximation of the exact parameter. For further comparison, we also plot in Figure 2 (d) the cross section along $\{x_1 = -0.1\}$ for $q^\dagger$ and $\tilde{q}_{\alpha_*}$. We remark that Algorithm 2 consists of two components: the outer IRLS iteration and the inner projected Newton iteration. Due to the superlinear convergence of the projected Newton method, in our numerical experiments we only need to take a few IRLS iterations and Newton iterations to find $\tilde{q}_\alpha$ for each $\alpha \in \Delta_\gamma$.

Example 4.3. We next consider the electrical impedance tomography discussed in Section 3.2 with $\Omega = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$ and $\rho(x) = \frac{3}{2}\chi_D - \frac{1}{2}\chi_{\Omega \setminus D}$, where

$$D := \{(x_1, x_2) \in \Omega : |x_1| \le 0.5 \text{ and } |x_2| \le 0.5\}$$

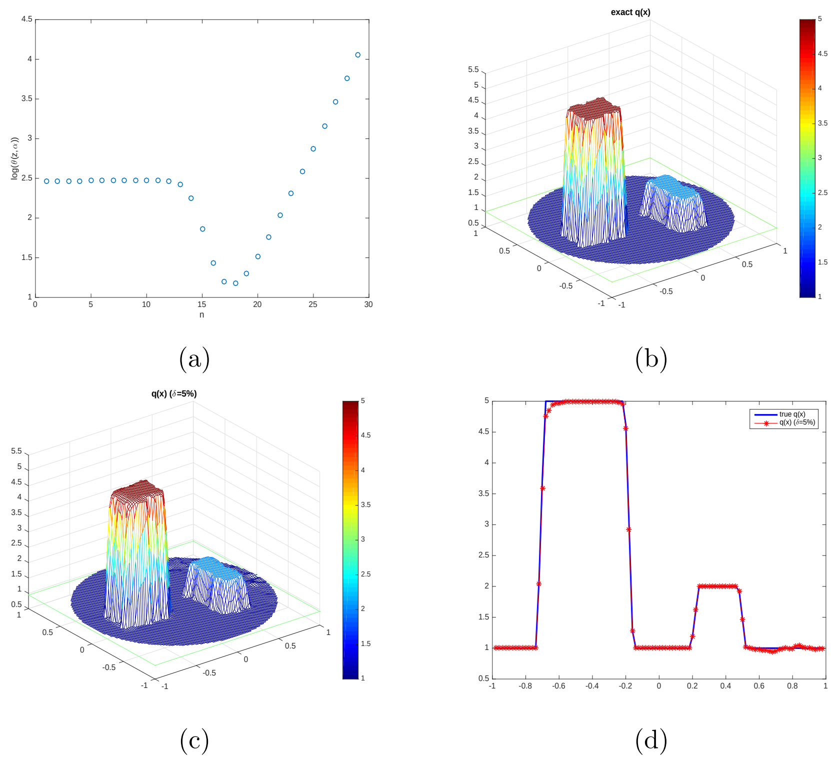


FIGURE 2. The reconstruction for Example 4.2 with (a) The relation between $\Theta(\alpha_n, \tilde{u})$ and n. (b) the exact parameter. (c) the reconstructed result. (d) cross sections along $\{x_1 = -0.1\}$.

and χ_D denotes the characteristic function of D. We assume that the sought parameter is given by $q^\dagger = 3\chi_{\Omega_1} + 2\chi_{\Omega_2} + \chi_{\Omega\setminus(\Omega_1\cup\Omega_2)}$, where

$$\Omega_1 := \{(x_1, x_2) \in \Omega : 9(x_1 + 0.5)^2 + 16(x_2 - 0.5)^2 \le 1\},$$
$$\Omega_2 := \{(x_1, x_2) \in \Omega : (x_1 - 0.5)^2 + (x_2 + 0.5)^2 \le 1/16\}.$$

We will use multiple measurements of Cauchy data $(\tilde{f}_\ell, \tilde{g}_\ell), \ell = 1, \cdots, L$, to reconstruct $q^\dagger$. Since $q^\dagger$ is piecewise constant, we take $\mathcal{R}(q) = \int_\Omega |Dq|$, the total variation of q over Ω, as the regularization term. Thus the regularized solution $\tilde{q}_\alpha$ is determined by the minimizer of the functional

$$q \to \sum_{\ell=1}^{L} \int_\Omega q(x)|\nabla(u_D[q, \tilde{f}_\ell] - u_N[q, \tilde{g}_\ell])|^2 dx + \alpha \int_\Omega |Dq|$$

over $\mathcal{Q}$. Due to the nonsmoothness of $\mathcal{R}(q)$, we use an IRLS scheme and the projected Newton method to find an approximation of the regularized solution, i.e. by taking a sequence of small positive numbers $\{\varepsilon_j\}$ and an initial guess q_0, we produce an approximate solution by solving the minimisation problem

$$q_{j+1} \in \arg\min_{q\in\mathcal{Q}} \left\{ \sum_{\ell=1}^{L} \int_\Omega q(x)|\nabla(u_D[q, \tilde{f}_\ell] - u_N[q, \tilde{g}_\ell])|^2 dx + \alpha \int_\Omega \omega_j |\nabla q|^2 \right\} \quad (4.4)$$

iteratively, where $\omega_j = (|\nabla q_j|^2 + \varepsilon_j^2)^{-1/2}$. According to the Karush-Kuhn-Tucker theory, for (4.4) we have the KKT conditions

$$\begin{aligned}
&\int_\Omega q\nabla u_\ell \cdot \nabla \phi_\ell = \int_\Omega \rho\phi_\ell, \quad \forall \phi_\ell \in H_0^1(\Omega),\ \ \ell = 1, \cdots, L;\\
&\int_\Omega q\nabla v_\ell \cdot \nabla \psi_\ell = \int_{\partial\Omega} g\psi_\ell + \int_\Omega \rho\psi_\ell, \quad \forall \psi_\ell \in H_\diamond^1(\Omega),\ \ \ell = 1, \cdots, L;\\
&\int_\Omega \left[(\tilde{q} - q)\sum_{\ell=1}^{L}(|\nabla u_\ell|^2 - |\nabla v_\ell|^2) + 2\alpha\omega_j \nabla q \cdot \nabla(\tilde{q} - q)\right] \geq 0, \quad \forall \tilde{q} \in \mathcal{Q}.
\end{aligned} \tag{4.5}$$

We may use the projected/semismooth Newton method to solve (4.5) which leads to Algorithm 3.

Algorithm 3 Projected Newton algorithm for (4.4)

1: Given initial guess q^0 and u_ℓ^0, v_ℓ^0 for $\ell = 1, \cdots, L$.
2: **for** $k = 0, 1, 2, ...$ **do**
3: Obtain the increments $\xi_\ell^k, \eta_\ell^k, \sigma^k$ of u_ℓ^k, v_ℓ^k, q^k by solving

$$\begin{cases}
\displaystyle\int_\Omega q^k \nabla \xi_\ell \cdot \nabla \phi_\ell + \int_\Omega \sigma \nabla u_\ell^k \cdot \nabla \phi_\ell \\
\qquad \displaystyle = \int_\Omega q^k \nabla u_\ell^k \cdot \nabla \phi_\ell - \rho\phi_\ell, \quad \forall \phi_\ell \in H_0^1(\Omega),\ \ \ell = 1, \cdots, L\\
\displaystyle\int_\Omega q^k \nabla \eta_\ell \cdot \nabla \psi_\ell + \int_\Omega \sigma_\ell \nabla v_\ell^k \cdot \nabla \psi_\ell \\
\qquad \displaystyle = \int_\Omega q^k \nabla v_\ell^k \cdot \nabla \psi_\ell - \rho\psi_\ell - \int_{\partial\Omega} \tilde{g}_\ell \psi_\ell, \quad \forall \psi_\ell \in H_\diamond^1(\Omega),\ \ \ell = 1, \cdots, L\\
\displaystyle\int_\Omega 2\mu \sum_{\ell=1}^{L}(\nabla u_\ell^k \cdot \nabla \xi_\ell - \nabla v_\ell^k \cdot \nabla \eta_\ell) + 2\alpha \int_\Omega \omega_j \nabla \sigma \cdot \nabla \mu \\
\qquad \displaystyle = \int_\Omega \mu \sum_{\ell=1}^{L}(|\nabla u_\ell^k|^2 - |\nabla v_\ell^k|^2) + 2\alpha \int_\Omega \omega_j \nabla q^k \cdot \nabla \mu, \quad \forall \mu \in \mathcal{Q}
\end{cases}$$

4: Update u_ℓ^k, v_ℓ^k, q^k by $u_\ell^{k+1} = u_\ell^k - \xi_\ell^k$, $v_\ell^{k+1} = v_\ell^k - \eta_\ell^k$ and $q^{k+1} = q^k - \sigma^k$.
5: Projection q^{k+1} onto $\mathcal{Q}$ to get new q^{k+1}.
6: Check stop condition.
7: **end for**

In our numerical simulations we use four groups of Cauchy data $(f_\ell^\dagger, g_\ell^\dagger)$, $\ell = 1, \cdots, 4$, where

$$\begin{aligned}
g_1^\dagger &= \chi_{(0,1]\times\{-1\}} - \chi_{[-1,0]\times\{1\}} + 2\chi_{(0,1]\times\{1\}} - 2\chi_{[-1,0]\times\{-1\}}\\
&\quad + 3\chi_{\{-1\}\times(-1,0]} - 3\chi_{\{1\}\times(0,1)} + 4\chi_{\{1\}\times(-1,0]} - 4\chi_{\{-1\}\times(0,1)},\\
g_2^\dagger &= x_1 + x_2, \quad g_3^\dagger = x_1 - x_2, \quad g_4^\dagger = x_1^2 - x_2^2,
\end{aligned}$$

and each $f_\ell^\dagger$ is produced by solving (3.16) with $q = q^\dagger$ and $g = g_\ell^\dagger$ and then taking the trace of the solution on $\partial\Omega$. Now we add random Gaussian noise on $(f_\ell^\dagger, g_\ell^\dagger)$ to produce noisy data $(\tilde{f}_\ell, \tilde{g}_\ell)$ with

$$\frac{\|\tilde{f}_\ell - f_\ell^\dagger\|_{L^2(\partial\Omega)}}{\|f_\ell^\dagger\|_{L^2(\partial\Omega)}} = 0.1\% \quad \text{and} \quad \frac{\|\tilde{g}_\ell - g_\ell^\dagger\|_{L^2(\partial\Omega)}}{\|g_\ell^\dagger\|_{L^2(\partial\Omega)}} = 0.1\%$$

for $\ell = 1, \cdots, 4$. We then use $(\tilde{f}_\ell, \tilde{g}_\ell)$, $\ell = 1, \cdots, 4$, to reconstruct $q^\dagger$. In our numerical computation, we take $\varepsilon_j = 0.01$ for all j and solve the variational problems in Algorithm 3 by FreeFem++ [17]; in the definition of $\mathcal{Q}$ we take $\underline{q} = 0.1$ and $\bar{q} = 20$. When using our heuristic rule for choosing the regularization parameter, we use $A = 10^5$ and use Δ_γ with $\alpha_0 = 1.0$ and $\gamma = 0.6$. According to the computations, our heuristic rule gives the regularization parameter $\alpha_* = 1.0 \times 0.6^{26}$. The relation between $\Theta(\alpha_n, \cdot)$, with $\alpha_n = 1.0 \times 0.6^n$, and n is plotted in Figure 3 (a). The exact parameter $q^\dagger$ and the reconstructed parameter $\tilde{q}_{\alpha_*}$ are plotted in Figure 3 (b) and (c). For further comparison, we also plot the cross section of $q^\dagger$ and $\tilde{q}_{\alpha_*}$ along $\{x_2 = -0.5\}$. We can see the reconstructed parameter is quite close to the exact parameter.

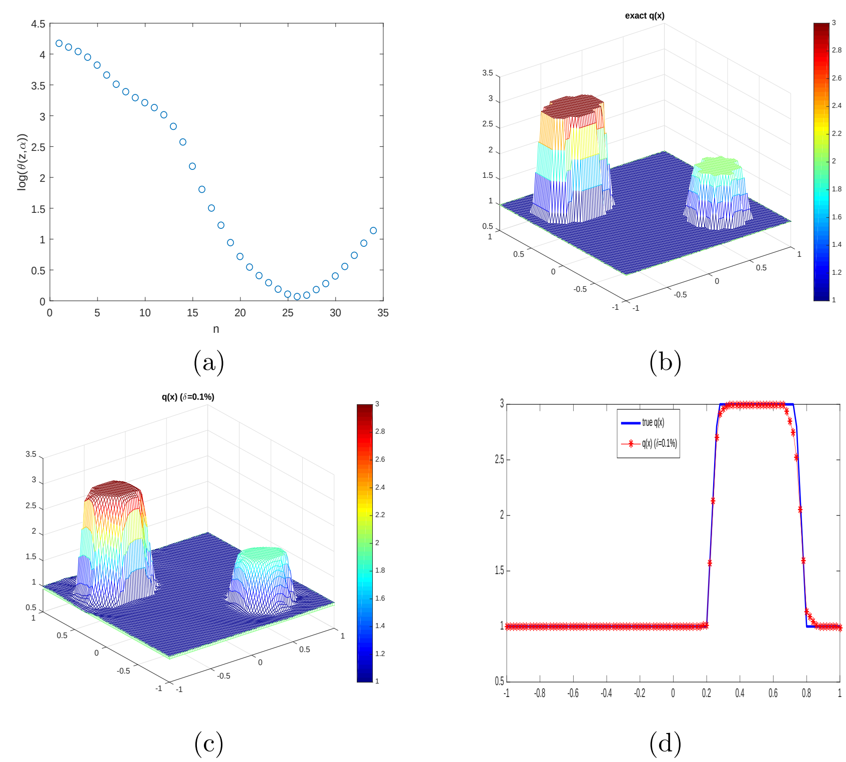


FIGURE 3. The reconstruction for Example 4.3 with (a) the relation between $\Theta(\alpha_n, \cdot)$ and n. (b) exact parameter. (c) reconstruction result. (d) cross sections along $\{x_2 = -0.5\}$.

Acknowledgement

X. Lu is partially supported by the National Natural Science Foundation of China (No. 91630313 & No. 11871385); X. Jia is partially supported by the National Natural Science Foundation of China (No. 11871313); and Q. Jin is partially supported by the Future Fellowship of the Australian Research Council.

References

[1] A. B. Bakushinskii, *Remarks on choosing a regularization parameter using the quasioptimality and ratio criterion*, U.S.S.R. Comp. Maths. Math. Phys. 24, No. 4, 181–182.

[2] R. I. Boţ and B. Hofmann, *An extension of the variational inequality approach for nonlinear ill-posed problems*, J. Integral Equations Appl., 22 (2010), 369–392.

[3] E. J. Cands, J. K. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), no. 8, 1207–1223.

[4] X. Chen, Z. Nashed and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), no. 4, 1200–1216.

[5] I. Daubechies, R. Devore, M. Fornasier and C. S. N. Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Comm. Pure Appl. Math., 63 (2010), 1–38.

[6] L. Ding and W. Han, *$\alpha\ell_1 - \beta\ell_2$ regularization for sparse recovery*, Inverse Problems, to appear, 2019.

[7] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Vol. 375, Springer Science & Business Media, 1996.

[8] E. Esser, Y. Lou and J. Xin, *A method for finding structured sparse solutions to non-negative least squares problems with applications*, SIAM J. Imaging Sci., 6 (2013), 2010–2046.

[9] J. Flemming, *Existence of variational source conditions for nonlinear inverse problems in Banach spaces*, J. Inverse Ill-posed Problems, 26 (2018), no. 2, 277–286.

[10] J. Flemming and B. Hofmann, *A new approach to source conditions in regularization with general residual term*, Numer. Funct. Anal. Optim., 31 (2010), no. 1-3, 254–284.

[11] K. Frick and M. Grasmair, *Regularization of linear ill-posed problems by the augmented Lagrangian method and variational inequalities*, Inverse Problems, 28 (2012), 104005.

[12] M. Grasmair, *Generalized Bregman distances and convergence rates for non-convex regularization methods*, Inverse Problems, 26(2010), 115014 (16pp).

[13] M. Hanke and T. Raus, *A general heuristic for choosing the regularization parameter in ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), no. 4, 956–972.

[14] D. N. Hao and T. N. T. Quyen, *Convergence rates for Tikhonov regularization of coefficient identification problems in Laplace-type equations*, Inverse Problems, 26 (2010), 125014.

[15] D. N. Hao and T. N. T. Quyen, *Convergence rates for total variation regularization of coefficient identification problems in elliptic equations I*, Inverse Problems 27 (2011), no. 7, 075008, 28 pp.

[16] D. N. Hao and T. N. T Quyen, *Convergence rates for Tikhonov regularization of a two-coefficient identification problem in an elliptic boundary value problem*, Numer. Math., 120 (2012), no. 1, 45–77.

[17] F. Hecht, *New development in freefem++*, J. Numer. Math. 20 (2012), 251–265.

[18] M. Hintermüller, K. Ito and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), no. 3, 865–888.

[19] M. Hinze, B. Kaltenbacher and T. N. T. Quyen, *Identifying conductivity in electrical impedance tomography with total variation regularization*, Numer. Math. 138 (2018), 723–765.

[20] M. Hinze and T. N. T. Quyen, *Matrix coefficient identification in an elliptic equation with the convex energy functional method*, Inverse Problems, 32 (2016), no. 8, 085007, 29 pp.

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

[22] B. Hofmann and P. Mathé, *Parameter choice in Banach space regularization under variational inequalities*, Inverse Problems, 28 (2012), no. 10, 104006, 17 pp.

[23] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM. Philadelphia, 2008.

[24] B. Jin and D. A. Lorenz, *Heuristic parameter-choice rules for convex variational regularization based on error estimates*, SIAM J. Numer. Anal., 48 (2010), 1208–1229.

[25] Q. Jin, *Hanke-Raus heuristic rule for variational regularization in Banach spaces*, Inverse Problems, 32 (2016), 085008 (19pp).

[26] I. Knowles, *A variational algorithm for electrical impedance tomography*, Inverse Problems, 14 (1998), 1513–1525.

[27] R. V. Kohn and A. McKenney, *Numerical implementation of a variational method for electrical impedance tomography*, Inverse Problems, 6 (1990), 389–414.
[28] R. V. Kohn and M. Vogelius, *Relaxation of a variational method for impedance computed tomography*, Commun. Pure Appl. Math. 40 (1987), 745–777.
[29] M. Osborne, *Finite Algorithms in Optimization and Data Analysis*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, Chichester, 1985.
[30] C. Pöschl, *Tikhonov regularization with general residual term.* PhD thesis, University of Innsbruck, Austria, 2008.
[31] T. N. T. Quyen, *Variational method for multiple parameter identification in elliptic PDEs*, J. Math. Anal. Appl., 461 (2018), no. 1, 676–700.
[32] L. Rudin, S. Osher and C. Fatemi, *Nonlinear total variation based noise removal algorithm*, Phys. D, 60 (1992), 259–268.
[33] T. Schuster, B. Kaltenbacher, B. Hofmann and K. S. Kazimierski, *Regularization Methods in Banach Spaces*, Radon Series on Computational and Applied Mathematics 10 Walter de Gruyter, Berlin 2012.
[34] Z. Sun, Y. Jiao, B. Jin and X. Lu, *Numerical identification of a sparse Robin coefficient*, Adv. Comput. Math., 41 (2015), 131–148.
[35] P. Yin, Y. Lou, Q. He and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*, SIAM J. Sci. Comput., 37 (2015), A536–A563.
[36] J. Zou, *Numerical methods for elliptic inverse problems*, Int. J. Comput. Math. 70 (1998), 211–232.

(Huan Liu) School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
E-mail address: huanliu@whu.edu.cn

(Rommel Real) Mathematical Sciences Institute, Australian National University, Canberra, ACT 2601, Australia
E-mail address: Rommel.Real@anu.edu.au

(Xiliang Lu) School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China & Hubei Key Laboratory of Computational Science (Wuhan University), Wuhan 430072, China
E-mail address: xllv.math@whu.edu.cn

(Xianzheng Jia) School of Mathematics and Statistics, Shandong University of Technology, Zibo, Shandong Province, China
E-mail address: jxz_1@hotmail.com

(Qinian Jin) Mathematical Sciences Institute, Australian National University, Canberra, ACT 2601, Australia
E-mail address: qinian.jin@anu.edu.au