IMA Journal of Applied Mathematics (2024) **89**, 12–43
https://doi.org/10.1093/imamat/hxae008
Advance Access publication on 12 March 2024

Trust your source: quantifying source condition elements for variational regularisation methods

Martin Benning*
Department of Computer Science, University College London, London WC1E 6BT, UK
*Corresponding author: martin.benning@ucl.ac.uk

Tatiana A. Bubba
Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

Luca Ratti
Department of Mathematics, University of Bologna, Bologna 40126, Italy

and

Danilo Riccio
School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK

[Received on 28 February 2023; revised on 8 February 2024; accepted on 11 February 2024]

Source conditions are a key tool in regularisation theory that are needed to derive error estimates and convergence rates for ill-posed inverse problems. In this paper, we provide a recipe to practically compute source condition elements as the solution of convex minimisation problems that can be solved with first-order algorithms. We demonstrate the validity of our approach by testing it on two inverse problem case studies in machine learning and image processing: sparse coefficient estimation of a polynomial via LASSO regression and recovering an image from a subset of the coefficients of its discrete Fourier transform. We further demonstrate that the proposed approach can easily be modified to solve the machine learning task of identifying the optimal sampling pattern in the Fourier domain for a given image and variational regularisation method, which has applications in the context of sparsity promoting reconstruction from magnetic resonance imaging data.

Keywords: Inverse problems; Regularisation theory; Source conditions; Range conditions; Variational regularisation; Machine learning; Optimal sampling; Compressed sensing; MRI.

1. Introduction

The canonical example of an inverse problem is the resolution of the linear equation

$$Ku^\dagger = f \tag{1}$$

where $u^\dagger \in X$ is the quantity we wish to recover from the data $f \in Y$, and $K : X \to Y$ is a (bounded, linear) operator. In order for (1) to be ill-posed, the operator K is usually assumed to be compact, such that the Moore–Penrose inverse of K is unbounded. In most practical applications, the exact data f are not accessible; instead, we only have access to a perturbed version $f^\delta \in Y$ and we are required to combat the

ill-posedness with regularisation methods. Over the past decades, variational regularisation methods have been extensively used to approximate the solution of ill-posed inverse problems (Scherzer *et al.*, 2009; Benning & Burger, 2018a). The key idea of variational regularisation methods consists of approximating $u^\dagger$ in (1) as the solution of the minimisation problem

$$\arg\min_{u\in X} \left\{\mathcal{D}(Ku,f^\delta) + \alpha J(u)\right\}, \tag{2}$$

where $\mathcal{D}(Ku,f^\delta)$ is a data fidelity term expressing the gap between the computed and measured data, $J(u)$ is a regularisation functional favouring a priori information on the solution or, equivalently, penalising solutions with undesired structures, and $\alpha > 0$ is the regularisation parameter controlling the influence of the two terms on the minimiser. In this work, the data fidelity term is simply the least-squares term $\mathcal{D}(Ku,f^\delta) = \frac{1}{2}\|Ku - f^\delta\|_Y^2$. The choice of the regularisation term is less straightforward and it has been an active field of research: from the pioneering work of Tikhonov (Andrey, 1943; Benning & Burger, 1963), to more recent papers on sparsity promoting regularisation (Donoho, 1992; Candès *et al.*, 2006; Daubechies *et al.*, 2004), passing through the seminal work of Rudin et al. Leonid *et al.* (1992) and Mumford & Shah (1989) where nonlinear variational models were first introduced. We want to emphasise that the concepts of inverse problems and variational regularisation also play a key role in machine learning tasks such as regression and logistic regression, with the LASSO (Tibshirani, 1996) being one of the most prominent examples of a sparsity-promoting regression approach of the form of (2).

One reason for the popularity of variational regularisation methods is that, along with an intuitive approach to modelling, these provide a framework for their basic analysis, particularly in the case of convex regularisation functionals. In this context, once the existence of a minimiser for (2) is guaranteed by suitable assumptions, a compelling matter is establishing error estimates and quantify rates of convergence. These are generally smoothness conditions on the solution that are required, which go under the name of *source conditions*.

There exist several different types of source conditions and a detailed survey of their role to derive convergence rates can be found in (Schuster *et al.*, 2012, Chapter 3). Classical source conditions for nonlinear variational methods trace back to Chavent & Kunisch (1997) and the seminal work of Burger & Osher (2004), where the authors derive convergence rates in the Bregman distance (Bregman, 1967) setting for general convex functionals in Banach spaces, extending the results already known in the quadratic case (Engl *et al.*, 1989). By that time, convergence rates under source conditions were established in the Hilbert setting (Tautenhahn, 1998), also for the nonlinear case (Engl *et al.*, 1996).

An alternative – but often equivalent (see Benning & Burger 2018a, Theorem 5.12) – formulation to classical source conditions is provided by *range conditions* (see Burger & Osher, 2004, Proposition 1 and Benning & Burger 2018a, Definition 5.8), which ensure that the unknown solution $u^\dagger$ of the inverse problem (1) is in the range of the variational regularisation operator. For example, range conditions (together with a restricted injectivity assumption) allow to establish linear convergence rates for ℓ^1 regularisation, avoiding stronger assumptions like the restricted isometry property (Grasmair *et al.*, 2011). More recently, range conditions have also been deployed in the context of data-driven regularisation (Mukherjee *et al.*, 2021).

While we will focus on source and range conditions in this work, we want to emphasise that stronger source conditions exist (like those in Resmerita, 2005) that yield better convergences rates compared to those derived in Burger & Osher (2004).

Other noticeable source conditions formulated as a variational inequality (instead of an equation or inclusion) are the so-called variational source conditions first introduced in Hofmann *et al.* (2007) in the context of nonlinear ill-posed inverse problems. These have been widely used to derive convergence rates

(Flemming, 2012): in particular, in the context of sparsity promoting regularisers (Grasmair *et al.*, 2008; Hohage & Weidling, 2017; Hohage & Miller, 2019). Last but not least, approximate source conditions have been used to quantify errors of regularisation methods and were originally introduced in the Hilbert setting (Hofmann *et al.*, 2006), but quickly adapted to the Banach space setting (Hein & Hofmann, 2009). They are generally weaker compared to classical source conditions, but allow to measure effectively how well the range condition can be approximated (Burger *et al.*, 2018).

Over the years, source conditions have been regarded as a purely theoretical tool to quantify convergence rates. In the papers cited above, there is little or no discussion on practical problems or applications where source conditions are proven to be met. Even for simple (non-trivial) inverse problems, systematic numerical studies of convergence rates for which the solutions provably verify a source condition are rare. Such examples include Ramlau & Resmerita (2010), where the authors consider a convolution operator and show that the convergence rates derived in their work are only verified in the case of a solution fulfilling the source condition. In Benning & Burger (2011, Example 6.1 and Section 6.2.1), the authors compare their theoretical error estimates with the computational errors. In order to do so, they propose two inverse problem solutions of which one can be shown to analytically satisfy the source condition while the other violates it (Example 6.1), and numerically estimate a source condition element to verify error estimates (Section 6.2.1.). More recently, in the context of statistical inverse problems (Bubba *et al.*, 2023; Tatiana & Ratti, 2022), the authors derive convergence rates in expected Bregman distance, using both classical and approximate source conditions. These are numerically verified for a tomographic problem using a solution that fulfils the source condition. Another exception is the special case of generalised eigenfunctions and singular vectors, whose theoretical and numerical computation has attracted significant interest over previous years (cf. Benning & Burger, 2013; Gilboa, 2014; Benning & Burger, 2014; Burger *et al.* 2015, 2016; Gilboa *et al.*, 2016; Schmidt *et al.*, 2018; Benning & Burger 2018a,b; Nossek & Gilboa, 2018; Bozorgnia *et al.*, 2020; Bungert *et al.*, 2021 and references therein).

Motivated by the lack of a generic, practical strategy to reliably estimate source and range condition elements, in this paper, **we propose a novel approach to compute source condition elements as the solution of a convex and differentiable minimisation problem**. We prove that the solution is equivalent to fulfilling a classic source (or range) condition when K in (1) is injective. We demonstrate the validity of our strategy by testing it numerically for two case studies: the machine learning problem of sparse polynomial regression and the image processing problem of recovering an image from a subset of its Fourier samples. Having at disposal explicit knowledge of the source condition elements has indubitable significance from a theoretical perspective: (1) the source condition element provides exact quantitative error estimates (i.e. the constants appearing in the convergence rates bounds can be computed explicitly), and (2) computing a range condition element allows us to determine the data that is required for a variational regularisation approach to approximate the solution of an inverse problem. However, it also opens up interesting lines of research in the context of learning for inverse problems. In this work, we show how estimating a sparse source condition element can be used to devise a strategy for optimally sampling in the Fourier domain in the context of variational regularisation. Applications for such a problem can for instance be found in magnetic resonance imaging (MRI; cf. Sherry *et al.*, 2020).

Our contributions. The main contribution of this paper is a novel, practical approach to compute source conditions elements, with the aim of depicting the source conditions as quantitative tools in variational regularisation, rather than merely theoretical requirements. To do so, we consider a rather general Banach space setup, and we employ tools from convex analysis to provide an alternative formulation for the source condition, together with an iterative procedure for the numerical approximation of the source condition element. To demonstrate the potential of this strategy in applications, we provide a set of numerical experiments where we focus on two ill-posed inverse problems: 1D interpolation and 2D

imaging. Finally, we show how a slight modification of the proposed approach forms a supervised optimal design technique by suitably identifying samples in the Fourier domain to promote desired properties of the associated source condition elements.

Structure of the paper. The remainder of the paper is organised as follows. Section 2 is devoted to briefly reviewing relevant theoretical concepts from convex analysis. In Section 3, we describe what source conditions are, why they are necessary and detail the key idea of our approach, namely, how to compute source condition elements as the solution of a convex minimisation problem. We demonstrate the performance of our methodology by a series of numerical experiments in Section 4 with focus on machine learning. In particular, in Section 4.3, we show how our strategy can be used to obtain the optimal sampling pattern in the Fourier domain for a fixed image and given variational regularisation method. Concluding remarks and future prospects are summarised in Section 5.

2. Mathematical preliminaries

In this section, we set the notation and give a brief overview of some key results from convex analysis which are useful for the rest of the paper (see Ekeland & Temam, 1999).

For a functional $F\colon X \to \mathbb{R}$, the subdifferential evaluated at a point $\hat{u} \in X$ is defined as

$$\partial F(\hat{u}) = \left\{x^* \in X^* : F(u) - F(\hat{u}) \geq \langle x^*, u - \hat{u}\rangle_*, \ \forall u \in X\right\},$$

where $\langle \cdot, \cdot \rangle_*$ denotes the canonical pairing between the dual space X^* and X. If F is also differentiable, the only element of the subdifferential $\partial F(\hat{u})$ is the Fréchet-derivative $\nabla F(\hat{u})$. The convex conjugate of a functional $F\colon X \to \mathbb{R}$ is defined as $F^\star\colon X^* \to \mathbb{R}$ such that

$$F^\star\colon p \in X^* \mapsto F^\star(p) = \sup_{u \in X} \left\{\langle p, u\rangle_* - F(u)\right\}.$$

If X is a Hilbert space, we can interpret $F^\star$ as a functional from X to $\mathbb{R}$, by identifying X with X^* via the Riesz's representation theorem. For a continuous convex functional F, the Bregman distance between u and w, associated with $p \in \partial F(w)$ is defined as

$$D_F^p(u, w) = F(u) - F(w) - \langle p, u - w\rangle_*.$$

Such a mapping is actually not a distance: it is non-negative but, in general, it is not symmetric and it does not satisfy the triangle inequality. To recover the symmetry, we usually rely on the symmetric Bregman distance, also known as the Jeffrey's distance, namely:

$$D_F^{\text{symm},q,p}(u, w) = D_F^q(w, u) + D_F^p(u, w), \qquad q \in \partial F(u),\ p \in \partial F(w).$$

Notice that if F is differentiable the subdifferentials are single-valued, hence we can drop the dependence of p, q and simply denote by $D_F(u, w)$ the symmetric Bregman distance between u and w.

3. Computing source condition elements

We consider the inverse problem (1) of recovering $u^\dagger \in X$ from $f^\delta \in Y$, a perturbation of the noiseless measurements $f = Ku^\dagger$ satisfying $\|f^\delta - f\|_Y \leq \delta$ for some given noise level $\delta \geq 0$. The forward map $K\colon X \to Y$ is a bounded linear operator, but we allow it to be compact and hence, do not assume that

it admits a bounded inverse in general. As a consequence, retrieving $u^\dagger$ from f^δ is an ill-posed problem, and solving (1) can suffer from instability in particular. The goal of regularisation theory for inverse problems is to construct a (family of) continuous operators that provide a satisfactory approximation of the potentially discontinuous inverse map. One of the most prominent paradigms is represented by variational regularisation, where a family of (potentially set-valued) operators R_α, parameterised by $\alpha \in (0, \infty)$, is defined as

$$R_\alpha : Y \rightrightarrows X, \quad R_\alpha : f \mapsto u_\alpha \in \operatorname*{arg\,min}_{u \in X} \left\{ \frac{1}{2} \|Ku - f\|_Y^2 + \alpha J(u) \right\}, \tag{3}$$

with $J : X \to \mathbb{R} \cup \{\infty\}$ being a so-called regularisation functional. For the remainder of this work, we make the following set of assumptions.

Assumption 3.1. In line with Benning & Burger (2018a), we assume that the following conditions are satisfied:

- Y is a Hilbert space;
- X is the dual of a normed space $\mathcal{X}$ such that its weak-star topology on X is metrisable on bounded sets;
- K is the adjoint of a bounded linear operator from Y to $\mathcal{X}$;
- J is the conjugate of a proper functional from $\mathcal{X}$ to $\mathbb{R} \cup \{\infty\}$ and is non-negative;
- for every $f \in Y$ and $\alpha > 0$ there exists a constant $c = c(a, b, \|f\|)$ that depends monotonically non-decreasing on all arguments such that $\|u\|_X \leq c$ if $\|Ku - f\| \leq a$ and $J(u) \leq b$.

Notice in particular that the assumption on J implies that it is a convex functional. Under those assumptions, it is possible to prove that (3) is well defined: namely, that for every $f \in Y$ and $\alpha > 0$ the set $R_\alpha(f)$ is non-empty (see Benning & Burger 2018a, Theorem 5.6, which also shows that $R_\alpha(f)$ is convex). Moreover, for $\alpha > 0$, each operator R_α is stable in the sense of the Kuratowski limit superior: namely, for every sequence $f_n \to f$ in Y there exists a subsequence $u_{n_k} \in R_\alpha(f_{n_k})$ converging to an element $u^* \in R_\alpha(f)$ in the weak-star topology on X (see Benning & Burger 2018a, Theorem 5.7). To conclude that R_α is a family of regularisers, one needs to ensure that, if the noise level δ converges to zero, i.e. $\delta \searrow 0$, there exists a choice $\alpha(\delta)$ such that $R_{\alpha(\delta)}(f^\delta)$ converges to $u^\dagger$. More advanced results in this direction also provide convergence rates, usually at the price of an additional assumption involving the unknown solution $u^\dagger$. In particular, an important tool to obtain quantitative error estimates between the solution $u^\dagger$ and the regularized solution u_α is represented by *source conditions*. In their most classical formulation, they require to assume that there exists $v \in Y$ such that

$$K^* v \in \partial J(u^\dagger), \tag{SC}$$

where $K^* : Y \to X^*$ is the adjoint of K (thanks to Riesz representation lemma, we identify Y with its dual), and $\partial J \subset X^*$ is the subdifferential of J (see Section 2). The object $v \in Y$ is referred to as the source condition element: in the following, we want to recall why elements satisfying the source condition allow us to derive error estimates with convergence rates (Section 3.1) and then discuss how to compute v for given $u^\dagger$, if such v exists, in Section 3.2.

3.1 *Error estimates for variational regularisation methods*

In Benning & Burger (2018a) and Burger & Osher (2004), error estimates measured in Bregman distances have been established for convex but non-smooth variational regularisation methods with the help of source conditions. For reasons of self-containment, we recall one of the key results.

We start by an equivalent formulation of (3) by means of first-order optimality conditions, exploiting the convexity of J: namely, $u_\alpha \in R_\alpha(f^\delta)$ if and only if

$$\exists\, p_\alpha \in \partial J(u_\alpha): \quad K^*(Ku_\alpha - f^\delta) + \alpha p_\alpha = 0,$$

where the previous equality holds in X^*. Assuming that (SC) is satisfied, we can subtract $\alpha K^* v$ on both sides of the equation to obtain

$$K^*(Ku_\alpha - f^\delta) + \alpha\left(p_\alpha - K^* v\right) = -\alpha K^* v.$$

Taking a dual product of both sides of this equation with $u_\alpha - u^\dagger$ then yields

$$\langle Ku_\alpha - f^\delta, Ku_\alpha - f\rangle_Y + \alpha\langle u_\alpha - u^\dagger, p_\alpha - K^* v\rangle_* = -\alpha\langle K^* v, u_\alpha - u^\dagger\rangle_*\,.$$

Please note that $\langle\cdot,\cdot\rangle_Y$ indicates the inner product in Y, whereas $\langle\cdot,\cdot\rangle_*$ denotes the pairing between a functional in X^* and an element of X. We can interpret the second term on the left-hand side by means of the symmetric Bregman distance. Since $p_\alpha \in \partial J(u_\alpha)$ and $K^* v \in \partial J(u^\dagger)$, the quantity $\langle u_\alpha - u^\dagger, p_\alpha - K^* v\rangle_*$ equals $D_J^{\text{symm},p_\alpha,K^*v}(u_\alpha, u^\dagger)$ (in the following we only refer to this term as D_J^{symm} for ease of notation). In addition, we can reformulate $\langle Ku_\alpha - f^\delta, Ku_\alpha - f\rangle_Y$ as

$$\langle Ku_\alpha - f + f - f^\delta, Ku_\alpha - f\rangle_Y = \frac{1}{2}\|Ku_\alpha - f\|_Y^2 + \frac{1}{2}\|Ku_\alpha - f^\delta\|_Y^2 - \frac{1}{2}\|f - f^\delta\|_Y^2,$$

to obtain

$$\frac{1}{2}\|Ku_\alpha - f\|_Y^2 + \frac{1}{2}\|Ku_\alpha - f^\delta\|_Y^2 + \alpha D_J^{\text{symm}}(u_\alpha, u^\dagger) = \frac{1}{2}\|f - f^\delta\|_Y^2 - \alpha\langle v, Ku_\alpha - f\rangle_Y,$$

where we have made use of (1), which assumes the identity $Ku^\dagger = f$. Using the identity

$$\langle \alpha v, f - Ku_\alpha\rangle_Y = \frac{\alpha^2}{2}\|v\|_Y^2 + \frac{1}{2}\|Ku_\alpha - f\|_Y^2 - \frac{1}{2}\|\alpha v - f + Ku_\alpha\|_Y^2$$

and subsequently eliminating $\frac{1}{2}\|Ku_\alpha - f\|_Y^2$ on both sides leads to the equation

$$\frac{1}{2}\|Ku_\alpha - f + \alpha v\|_Y^2 + \frac{1}{2}\|Ku_\alpha - f^\delta\|_Y^2 + \alpha D_J^{\text{symm}}(u_\alpha, u^\dagger) = \frac{1}{2}\|f - f^\delta\|_Y^2 + \frac{\alpha^2}{2}\|v\|_Y^2.$$

Dividing by $\alpha > 0$ and using the estimate $\|f - f^\delta\|_Y \leq \delta$ then yields the well-known error estimate (cf. Burger *et al.*, 2007)

$$\frac{1}{2\alpha}\|Ku_\alpha - f + \alpha v\|_Y^2 + \frac{1}{2\alpha}\|Ku_\alpha - f^\delta\|_Y^2 + D_J^{\text{symm}}(u_\alpha, u^\dagger) \leq \frac{\alpha}{2}\|v\|_Y^2 + \frac{\delta^2}{2\alpha}.$$

With the a priori choice $\alpha(\delta) = \delta/\|v\|_Y$ we minimise the right-hand side of this inequality and obtain

$$\begin{aligned} 0 &\leq \frac{\|v\|_Y}{2\delta}\left\|Ku_{\alpha(\delta)} - f + \delta\frac{v}{\|v\|_Y}\right\|_Y^2 + \frac{\|v\|_Y}{2\delta}\|Ku_{\alpha(\delta)} - f^\delta\|_Y^2 + D_J^{\text{symm}}(u_{\alpha(\delta)}, u^\dagger) \\ &\leq \|v\|_Y\delta. \end{aligned} \tag{4}$$

Hence, the symmetric Bregman distance between $u_{\alpha(\delta)}$ and $u^\dagger$ is bounded by the worst-case data error bound δ multiplied by the norm of the source condition element v. To quantify a priori error estimates such as (4), it is therefore vital to quantify the norm of the source condition element.

3.2 *Casting the computation of source condition elements as convex minimisation problems*

Having established the need for source condition elements and estimates for their norm in order to quantify a priori error estimates of regularised inverse problems solutions, we now discuss how we can formulate the computation of v verifying (SC) as a variational minimisation problem for a suitable convex and differentiable functional.

We first assume, in analogy to Burger *et al.* (2007), that the space X is continuously embedded into a Hilbert space Z. By means of the Hahn-Banach theorem, K can be extended to a bounded linear operator from Z to Y (which, with an abuse of notation, we still denote as K), whereas we interpret the functional J as acting on the whole space Z by extending it to ∞ outside of X. By this modification, we can equivalently solve the minimisation problem (3) in X or in Z.

REMARK 3.1. To rigorously verify the last statement, assume that $\tilde{K}\colon Z \to Y$ and $\tilde{J}\colon Z \to \mathbb{R}$ are the proposed extensions of K and J. Then, for any $f \in Y$, the problem

$$\tilde{u}_\alpha \in \tilde{R}_\alpha(u) = \operatorname*{arg\,min}_{u\in Z}\left\{\frac{1}{2}\|\tilde{K}u - f\|_Y^2 + \alpha\tilde{J}(u)\right\}$$

is such that $\tilde{R}_\alpha(f) = R_\alpha(f)$. Moreover, the optimality conditions are equivalent, since it is possible to show that, for $\hat{u} \in X$, the set $\partial\left\{\frac{1}{2}\|\tilde{K}\cdot - f\|_Y^2 + \alpha\tilde{J}\right\}(\hat{u}) \subset Z$ can be identified (via bounded extension) with the set $\partial\left\{\frac{1}{2}\|K\cdot - f\|_Y^2 + \alpha J\right\}(\hat{u}) \subset X^*$.

Motivated by Wang & Benning (2023), we provide a characterisation of the source condition element involving the use of proximity operators. For a proper, convex, lower semi-continuous functional F defined on the Hilbert space Z, the proximity operator (or proximal map) is defined as

$$\text{prox}_F\colon Z \to Z, \qquad \text{prox}_F(z) := \operatorname*{arg\,min}_{u\in Z}\left\{\frac{1}{2}\|u - z\|_Z^2 + F(u)\right\}. \tag{5}$$

The Hilbert space structure is, in principle, not needed for the definition of the proximal operator: for example, if X is assumed to be a Banach space, its norm can be used inside (5), which would not lead to an equivalent operator. We nevertheless rely on definition (5) and on the use of the norm $\|\cdot\|_Z$. The following result provides an alternative formulation of (SC).

PROPOSITION 3.1. The source condition (SC) can be rewritten as

$$u^\dagger = \operatorname{prox}_J\left(u^\dagger + K^* v\right), \tag{6}$$

Proof. By Assumption 3.1, the functional J is convex, which implies that $\frac{1}{2}\|\cdot - z\|_Z^2 + J$ is strictly convex, for every $z \in Z$. Therefore, its minimiser can be determined by means of first-order optimality conditions, yielding

$$u = \operatorname{prox}_J(z) \quad \Leftrightarrow \quad z - u \in \partial J(u). \tag{7}$$

Choosing $u = u^\dagger$ and $z = u^\dagger + K^* v$ (since K has been extended to an operator from Z to Y, it holds $K^*\colon Y \to Z$), we conclude (6). □

The advantage of the Hilbert space structure mainly resides in the possibility to use the expression (7). In Banach spaces, a similar result holds, but it requires the use of duality mappings, which would prevent the development of the following results. From this moment on, for the ease of notation, we denote by $\|\cdot\|$ the norm on the Hilbert space Z, previously denoted as $\|\cdot\|_Z$, and as $\langle\cdot,\cdot\rangle$ the inner product in Z.

We now reformulate the source condition by means of the following functional, which, following Wang & Benning (2023), we refer to as the *Bregman loss*:

$$B_J(u,p) := \left(\frac{1}{2}\|\cdot\|^2 + J\right)(u) + \left(\frac{1}{2}\|\cdot\|^2 + J\right)^\star(p) - \langle p, u\rangle, \tag{8}$$

where $(\frac{1}{2}\|\cdot\|^2 + J)^\star$ denotes the convex conjugate of $\frac{1}{2}\|\cdot\|^2 + J$.

We provide a summary of results related to B_J, which provide possible interpretations as well as important properties which will be used later. A proof of the following proposition can be found in (Wang & Benning, 2023, Theorem 10).

PROPOSITION 3.2. The functional $B_J\colon Z \times Z \to \mathbb{R}$ satisfies the following properties:

1. $B_J(u,p) = D^p_{\frac{1}{2}\|\cdot\|^2 + J}(u, \operatorname{prox}_J(p))$;
2. $B_J(u,p) = \frac{1}{2}\|u - \operatorname{prox}_J(p)\|^2 + D_J^{p - \operatorname{prox}_J(p)}(u, \operatorname{prox}_J(p))$;
3. B_J is bi-convex (separately in the variables u, p);
4. for any fixed $u \in \operatorname{dom}(J)$, the function $B_J(u,\cdot)$ is continuously Fréchet-differentiable: in particular,
$$\nabla_p B_J(u,p) = \operatorname{prox}_J(p) - u;$$
5. if moreover $\partial J(u) \neq \varnothing$, then $u = \operatorname{prox}_J(p)$ is a global minimiser of $B_J(u,\cdot)$.

REMARK 3.2. Note that, if $u \in \mathrm{dom}(J)$ but $J(u) = \varnothing$, an element $\hat{p}$ such that $\nabla_p B_J\left(u, \hat{p}\right) = 0$ may not exist. This scenario never occurs if $u \in \mathrm{int}(\mathrm{dom}(J))$, and in particular when the domain of J is a closed subset of Z.

In view of Proposition 3.2, we focus on the problem of minimising $B_J(u, \cdot)$ for a fixed p. According to *a*) and *b*), this accounts to finding an object $\hat{p}$ such that $\mathrm{prox}_J(\hat{p})$ is close to u and in particular, if the minimum is attained that $u = \mathrm{prox}_J(\hat{p})$. Moreover, thanks to the convexity and Fréchet-differentiability of $B_J(u, \cdot)$, its minimisation of $B_J(u, \cdot)$ can be performed by means of first-order optimisation methods. Combining ideas from Propositions 3.1 and 3.2, we introduce the functional

$$G_J \colon Y \to \mathbb{R}, \quad G_J(v) = B_J(u^\dagger, u^\dagger + K^* v).$$

Since G_J is a composition of an affine map with $B_J(u^\dagger, \cdot)$, it inherits the convexity and differentiability, and in particular

$$\nabla G_J(v) = K\mathrm{prox}_J(u^\dagger + K^* v) - K u^\dagger. \tag{9}$$

We immediately deduce the following result, related to the source condition:

PROPOSITION 3.3. If there exists v which satisfies (SC), then such v is a global minimiser of G_J. Viceversa, if v is a global minimiser of G_J, it satisfies

$$K u^\dagger = K\mathrm{prox}_J\left(u^\dagger + K^* v\right), \tag{10}$$

which automatically implies (SC) if K only has a trivial null space.

REMARK 3.3. Please note that Proposition 3.3 does not guarantee the existence of a global minimiser of G_J.

REMARK 3.4. We want to emphasise that if K is not injective and has a non-trivial null space, an element v that sets (9) to zero can still satisfy (SC) as long as long as $\mathrm{prox}(u^\dagger + K^* v) - u^\dagger$ does not lie in the null space of K. Furthermore, we can always verify if v satisfies (SC) a posteriori.

In order to check if $u^\dagger$ satisfies the source condition, and to compute the associated v, we are interested in solving

$$\hat{v} = \arg\min_v G_J(v) = \arg\min_v B_J\left(u^\dagger, u^\dagger + K^* v\right).$$

We can solve this minimisation problem with various different first-order methods, the simplest one being gradient descent. Thanks to (9), the iterative step in this case reads

$$v^{k+1} = v^k - \tau K\left(\mathrm{prox}_J\left(u^\dagger + K^* v^k\right) - u^\dagger\right), \tag{11}$$

which is globally convergent for $\tau \leq 1/\|K\|^2$ for arbitrary initial value $v^0 \in Y$, since every proximal map is 1-Lipschitz.

The descent algorithm described in (11) allows to approximate a source condition element by means of simple iterations. Nevertheless, it requires the knowledge of the proximal map prox_J, which has a closed-form solution for many popular choices of the functional J. However, many interesting functionals do not have closed-form proximal maps, such as the popular class of functionals of the form $J(u) = H(Au)$, where H is a functional with closed-form proximal map but where A is a linear operator. In the next section, we provide a strategy to deal with this larger family of regularisation functionals. An alternative approach, at least from a numerical perspective, is to approximate the evaluation of prox_J by means of a suitable optimisation sub-routine.

3.3 *Extension to more general functionals and range conditions*

In this section, we consider composite functionals of the form

$$J(u) = H(Au + b), \tag{12}$$

with $b \in Z$ and $A\colon Z \to Z$, and for which we assume that prox_H is known. This allows to take into account a large family of regularisation strategies, including generalised Tikhonov, sparsity promotion with respect to orthogonal bases or frames, and Total Variation, as we are going to recall in Section 4.2.

In order to estimate source condition elements in this scenario, we start by relating the source condition with the *range condition*. Referring to Benning & Burger (2018a, Definition 4.9), we say that $u^\dagger$ satisfies the range condition for $\alpha > 0$ if there exist data $g_\alpha \in Y$ such that

$$u^\dagger \in \underset{u\in X}{\arg\min} \left\{ \frac{1}{2}\|Ku - g_\alpha\|^2 + \alpha J(u) \right\}, \tag{RC1}$$

which means that $u^\dagger$ is in the range of the regulariser R_α, which motivates the name of the assumption. Thanks to Assumption 3.1 and to the definition of R_α in (3), we can apply Benning & Burger (2018a, Theorem 5.12) and conclude that $u^\dagger$ satisfies the range condition (RC1) for all $\alpha > 0$ if and only if it verifies the source condition (SC). In particular, considering the equivalent optimality condition associated with (RC1),

$$\frac{1}{\alpha}K^*\left(g_\alpha - Ku^\dagger\right) \in \partial J(u^\dagger), \tag{RC2}$$

we immediately obtain the following expression connecting the source condition element v with the pre-image g_α:

$$g_\alpha = Ku^\dagger + \alpha v. \tag{13}$$

Even though this equivalence holds true for any feasible J, it is particularly useful when applied to functionals of the form (12). Indeed, in this case
the range condition (RC2) can be written as

$$K^*(Ku^\dagger - g_\alpha) + \alpha A^* q^\dagger = 0,$$

for $q^\dagger \in \partial H(Au^\dagger + b)$. Multiplying by $1/\alpha$ and using (13), we obtain the conditions

$$K^* v = A^* q^\dagger, \tag{14a}$$

$$q^\dagger \in \partial H(Au^\dagger + b). \tag{14b}$$

In analogy to Section 3.2, we can rewrite (14b) as $Au^\dagger + b + q^\dagger \in \partial \left(\frac{1}{2}\|\cdot\|^2 + H\right)(Au^\dagger + b)$, respectively $Au^\dagger + b = \operatorname{prox}_H(Au^\dagger + b + q^\dagger)$, set up a Bregman loss functional and define the corresponding functional $G_H \colon Y \to \mathbb{R}$ as

$$G_H(q) = B_H(Au^\dagger + b, q + Au^\dagger + b).$$

Then, we can formulate estimating the source condition element v and the subgradient $q^\dagger$ as the minimisation of the objective functional

$$E_H(v, q^\dagger) = \frac{1}{2}\|K^* v - A^* q^\dagger\|^2 + G_H(q^\dagger). \tag{15}$$

Indeed, setting the partial Fréchet derivatives of (15) to zero yields

$$K\left(K^* v - A^* q^\dagger\right) = 0,$$
$$A\left(A^* q^\dagger - K^* v\right) + \operatorname{prox}_H\left(Au^\dagger + b + q^\dagger\right) - Au^\dagger - b = 0.$$

The above expressions are not equivalent to the range condition (RC1). Nevertheless, as in Proposition 3.3, if we assume that K has a trivial null space, then the first expression implies that (14a) is satisfied, whereas the second one implies (14b). Moreover, we can always check a posteriori if v and $q^\dagger$ satisfy (14).

In order to approximate a minimiser of E_H, we again can use first order methods. Since minimising (15) is a minimisation problem in two variables, it seems logical to use algorithms such as explicit coordinate descent (Beck & Tetruashvili, 2013; Wright, 2015), i.e.

$$\begin{aligned} v^{k+1} &= v^k - \tau K(K^* v^k - A^* q^k), \\ q^{k+1} &= q^k - \sigma\left(A(A^* q^k - K^* v^{k+1}) + \operatorname{prox}_H(Au^\dagger + b + q^k) - Au^\dagger - b\right), \end{aligned} \tag{16}$$

which guarantees the (global) convergence to a minimiser, provided that the the step sizes σ, τ are suitably chosen. Alternatively, one can formulate the augmented Lagrangian

$$\mathcal{L}_\delta(v, q^\dagger; \mu) = \frac{\delta}{2}\|K^* v - A^* q^\dagger\|^2 + G_H(q^\dagger) + \langle \mu, K^* v - A^* q^\dagger \rangle$$

and employ algorithmic approaches such as the alternating direction method of multipliers (Lions & Mercier, 1979) and variants to find the saddle-point, or directly formulate the saddle-point problem

$$\inf_{v,q^\dagger} \sup_{\mu} G_H(q^\dagger) + \langle \mu, K^* v - A^* q^\dagger \rangle$$

and compute the saddle-point with algorithms such as the primal-dual hybrid gradient method or variants of it (see Chambolle & Pock, 2016; Benning & Riis, 2021 and references therein).

In the following section, we provide some numerical examples to discuss the effectiveness of the reconstruction of source condition elements, both via the minimisation of G_J or of E_H, according to the different choices of J.

4. Numerical results

In this section, we demonstrate the validity of our strategy by testing it for two case studies. For each of our experiments, we present two examples. One, where we assume that the source condition is satisfied with a corresponding source condition element with small norm. And one, where we assume that the source condition is either not satisfied or only satisfied with a corresponding source condition element with large norm.

In Subsection 4.1, we consider the machine learning problem of polynomial LASSO regression, while in Subsection 4.2, we consider the problem of estimating a two-dimensional image from a subset of its Fourier samples. Finally, in Subsection 4.3, we show how our strategy can be used to learn the optimal sampling pattern in the Fourier domain for a fixed image and given variational regularisation method.

4.1 *1D example: polynomial LASSO regression*

As a first example, we show how to estimate source condition elements for a polynomial regression model. We start by considering the polynomial

$$\varphi(u) = 5u^2 - 3u^5 - 1, \qquad u \in \mathbb{R}. \tag{17}$$

We assume that we cannot directly access $\varphi(u)$, but instead have N pairs of samples (u_i, f_i^δ) for $i \in \{1, \dots, N\}$ with

$$f_i^\delta = \varphi(u_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{18}$$

i.e. measured data are affected by additive noise ε_i sampled from a Gaussian distribution with zero mean and standard deviation σ, and the superscript δ refers to the resulting data error $\delta = \|f - f^\delta\|$. We choose $\sigma = 0.1$ for our experiments and an example of sampled values is shown in Fig. 1. A polynomial regression model of order d that matches the N sampled data pairs (18) can be written as a system of linear algebraic equations in the form

$$\Phi w = f^\delta, \tag{19}$$

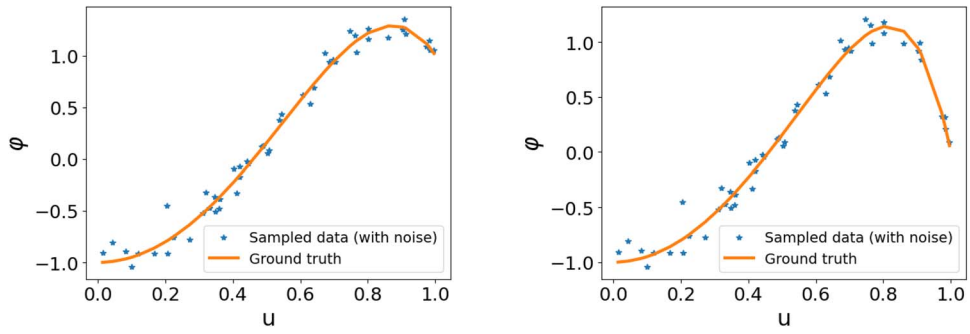


FIG. 1. Noisy sampled data (18), for the ground truth φ defined by (17) (left) and (22) (right).

where $w \in \mathbb{R}^{(d+1)}$ are the coefficients of the polynomial, $f^\delta \in \mathbb{R}^N$ is the vector containing the noisy samples f_i^δ, and $\Phi \in \mathbb{R}^{N\times(d+1)}$ is the Vandermonde matrix obtained from the values of x_i^p for $p \in \{0,\dots,d\}$.

We use $d = 75$ and $N = 50$ in our experiments.

Traditionally, the goal is finding the sparse set of coefficients $w = (w_0, w_1, \dots, w_d)$ of the dth order polynomial model via the LASSO model

$$w_\alpha \in \operatorname*{arg\,min}_{w\in\mathbb{R}^{1+d}} \left\{ \frac{1}{2}\|\Phi w - f^\delta\|^2 + \alpha\|w\|_1 \right\}, \tag{20}$$

where $\|w\|_1 = \sum_{j=0}^d |w_j|$ is the one-norm of the coefficient vector w.

Now instead of finding w_α, our goal is to find data g_α such that the true coefficients $w^\dagger$ are a solution of (20) for data g_α (instead of f^δ), respectively, the source condition element v using $\Phi^\top v \in \partial\|w^\dagger\|_1$, where the subdifferential of the one-norm is defined (component-wise) as

$$(\partial\|w^\dagger\|_1)_j := \begin{cases} 1 & \text{if } w_j > 0 \\ [-1,1] & \text{if } w_j = 0 \\ -1 & \text{if } w_j < 0 \end{cases}.$$

For our example, the coefficients $w^\dagger$ are zero except for $w_0^\dagger = -1$, $w_2^\dagger = 5$ and $w_5^\dagger = -3$. To compute the source condition element v, we minimise $G_J(v)$ iteratively as shown in (11), with

$$G_{\|\cdot\|_1}(v) = B_{\|\cdot\|_1}(w^\dagger, w^\dagger + \Phi^\top v). \tag{21}$$

Following (9), we observe

$$\nabla G_{\|\cdot\|_1}(v) = \Phi\left(\operatorname{prox}_{\|\cdot\|_1}\left(w^\dagger + \Phi^\top v\right) - w^\dagger\right).$$

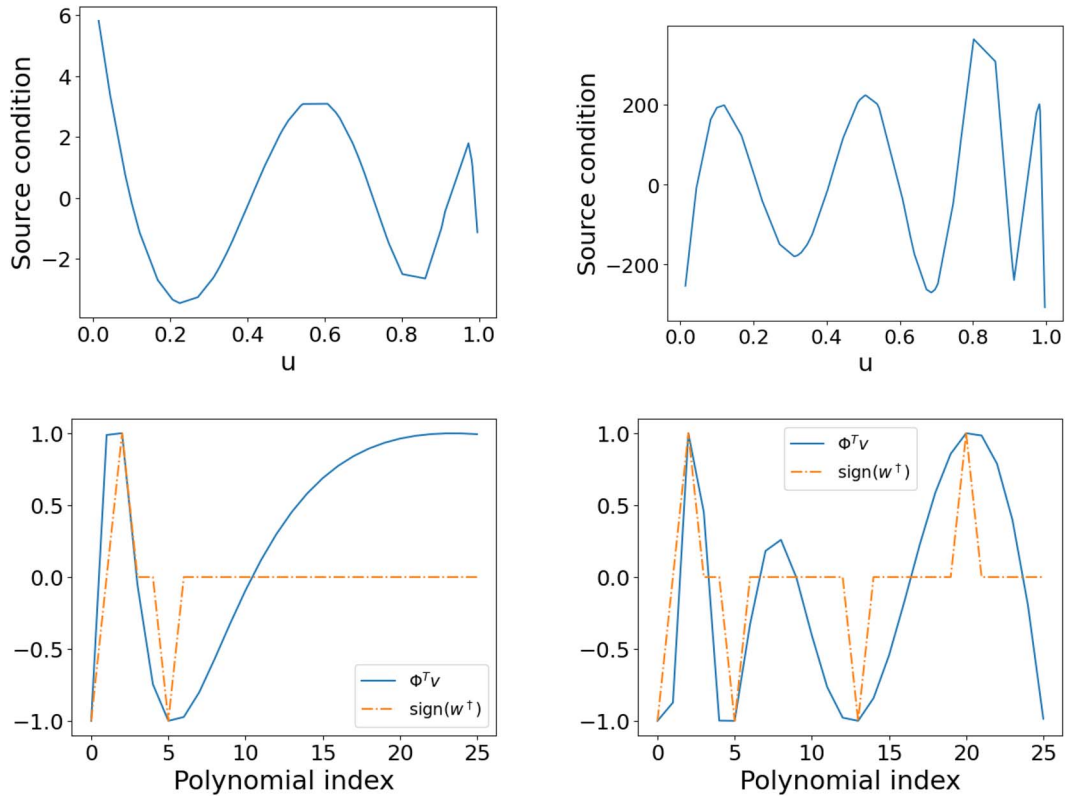


FIG. 2. Top row: source condition element v computed for examples (17) (left) and (22) (right). Note the difference in the ordinate scales: the norm of the source condition element for the higher order polynomial (22) is much higher and has larger oscillations than the ones for the lower order polynomial (17). Bottom row: comparison between $\Phi^\top v$ and the sign of the true coefficients $w^\dagger$. For both plots, every time $\text{sign}(w^\dagger) = \pm 1$, then also $\Phi^\top v = \pm 1$. This indicates that the estimated source condition v is correct.

In this particular instance, the proximal map $\text{prox}_{\|\cdot\|_1}$ has the closed-form solution

$$\text{prox}_{\|\cdot\|_1} = \text{sign}(\cdot)\max(\cdot - 1, 0)\,,$$

which is also known as the soft thresholding or soft shrinkage operator.

In order to compute the source condition element via (11), we initialise $v^0 = 0$, and then we iteratively compute $v^{k+1} = v^k - \tau \nabla G_{\|\cdot\|_1}(v^k)$, with step-size $\tau = 1/\|\Phi\|^2$. Further, we accelerate this procedure by using a Nesterov accelerated version (Nesterov, 1983) as described in Benning & Riis (2021). We stop iterating when either 2×10^8 iterations have passed, or when the Euclidean norm of ∇G_J is smaller than 10^{-12}. The norm of the computed source condition element is $\|v\| \approx 15.32$. The small norm of the source condition element indicates that we can retrieve the weights w reasonably well with the LASSO model even in the presence of noise, as the worst-case error in the estimate (4) is not amplified too strongly. To check that we have really computed a source condition element v, we need to compare $\Phi^\top v$ and $\text{sign}(w^\dagger)$. This comparison, together with the source condition element, is shown in the left column of Fig. 2. Finally, we compute the range condition $g_\alpha = \alpha v + \Phi w^\dagger$, where $\alpha = \delta/\|v\|$, and we show the

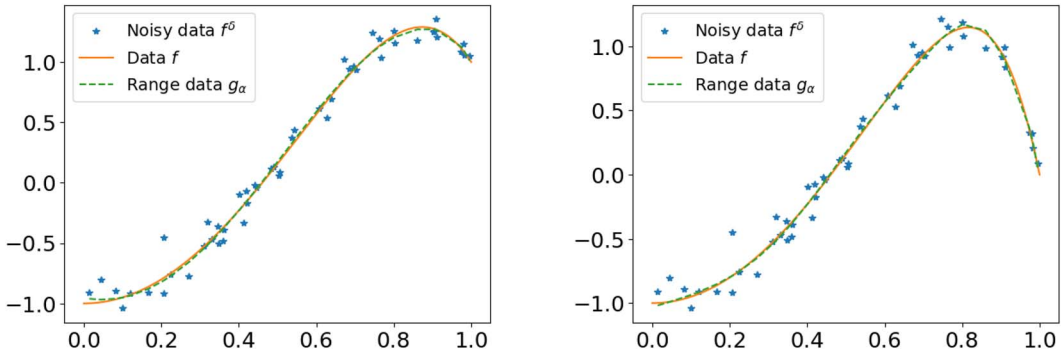


FIG. 3. Comparison between noisy samples f^δ, data f and range condition element g_α for $\alpha = \delta/\|v\|$, computed for examples (17) (left) and (22) (right).

result in the left column of Fig. 3. Remember that the data g_α is the data that, when fed into the LASSO operator (20) (instead of f^δ), ensures that the LASSO operator returns the coefficients $u^\dagger$.
The same experiment is now repeated with the higher-order polynomial

$$\varphi(u) = 5u^2 - 3u^5 - \frac{3}{2}u^{13} + \frac{1}{2}u^{20} - 1, \qquad u \in \mathbb{R}, \tag{22}$$

and the same Vandermonde matrix $\Phi \in \mathbb{R}^{N\times(d+1)}$ as before. As outlined in the introduction to this section, this second example is chosen because the problem of identifying coefficients of a polynomial of higher degree is more ill-conditioned compared to the previous example, and ℓ^1-regularisation has a much more difficult task at hand. We therefore assume that either the source condition will be violated for this example, or there exists a source condition element, but with much larger norm compared to the previous example.

In what follows, $w^\dagger$ must be changed accordingly to include the additional coefficients of this polynomial. Following the previous steps, we compute the source condition element v from the same initial (zero) vector, using the same iterative algorithm. Once the iteration is completed, we retrieve a source condition element with norm $\|v\| \approx 11359$. Convergence is obtained after approximately 32.66×10^6 iterations, which is much more than the iterations required for the simpler model to converge (which took less than 10^4 iterations). This indicates that the new problem is indeed harder to solve, as it takes more iterations and more time to converge to the solution. However, as the algorithm has converged, we expect that v satisfies the source condition, i.e. $\Phi^\top v \in \partial w^\dagger$. This comparison, together with the source condition v, is shown in the right column of Fig. 2. We can see that for both models we are able to retrieve the source condition as the two solutions both verify (SC), but we can see larger oscillations and a larger norm in the 20th-order polynomial, which indicates that the problem is harder than the 5th-order polynomial because of ill-conditioning. Using the inequality (4), we can conclude that we get a better bound for the fifth-order polynomial since the noise level $\delta = \|y - y^\delta\|$ is the same for both models. Similar to the previous example, we also visualise the corresponding range condition data g_α in Fig. 3.

4.2 *2D example: Fourier sub-sampling*

As a next example, we consider the inverse problem of sub-sampling the Fourier domain of two-dimensional images. If we restrict ourselves to Cartesian grids, we can describe this inverse problem

mathematically via the operator equation

$$S\mathcal{F}u^\dagger = f,$$

where $u^\dagger \in \mathbb{R}^{n_y \times n_x}$ denotes the unknown, two-dimensional discrete image, $f \in \mathbb{C}^m$ the sub-sampled Fourier data, $\mathcal{F}\colon \mathbb{R}^{n_y \times n_x} \to \mathbb{C}^{n_y \times n_x}$ the two-dimensional discrete Fourier transform, i.e.

$$(\mathcal{F}u)_{pq} = \frac{1}{\sqrt{n_x n_y}} \sum_{l=0}^{n_y-1} \sum_{j=0}^{n_x-1} u_{lj}\, e^{-i\frac{2\pi pl}{n_y}} e^{-i\frac{2\pi qj}{n_x}},$$

for $p \in \{0, \dots, n_y - 1\}$ and $q \in \{0, \dots, n_x - 1\}$, and $S\colon \mathbb{C}^{n_y \times n_x} \to \mathbb{C}^m$ the sampling operator that selects samples from the Cartesian grid. Since the operator $\mathcal{F}$ is orthogonal, it holds that $\mathcal{F}^\top = \mathcal{F}^{-1}$. For our numerical experiments we choose J to be the (discretised) isotropic total variation, i.e. $J = \mathrm{TV}\colon \mathbb{R}^{n_y \times n_x} \to \mathbb{R}$ with

$$\mathrm{TV}(u) = \sum_{i=1}^{n_y-1} \sum_{j=1}^{n_x-1} \sqrt{\left|u_{(i+1)j} - u_{ij}\right|^2 + \left|u_{i(j+1)} - u_{ij}\right|^2}.$$

In order to compute source condition elements $\mathcal{F}^{-1}S^\top v \in \partial \mathrm{TV}(u^\dagger)$, we make use of the range condition reformulation described in Section 3.3 and minimise (15) via explicit coordinate descent as described in (16), which, for our problem, reads

$$\begin{aligned} v^{k+1} &= v^k - \tau S\mathcal{F}\left(\mathcal{F}^{-1}S^\top v^k - A^\top q^k\right), \\ q^{k+1} &= q^k - \sigma\left(A\left(A^\top q^k - \mathcal{F}^{-1}S^\top v^{k+1}\right) + \mathrm{prox}_{\|\cdot\|_{2,1}}\left(Au^\dagger + q^k\right) - Au^\dagger\right). \end{aligned} \tag{23}$$

Here $A\colon \mathbb{R}^{n_y \times n_x} \to \mathbb{R}^{(n_y-1)\times(n_x-1)\times 2}$ is the (forward) finite-difference discretisation of the gradient, i.e.

$$(Au)_{ijp} = \begin{cases} u_{(i+1)j} - u_{ij} & \text{for } p = 1, \\ u_{i(j+1)} - u_{ij} & \text{for } p = 2, \end{cases}$$

for $i \in \{1, \dots, n_y - 1\}$ and $j \in \{1, \dots, n_x - 1\}$, and $\mathrm{prox}_{\|\cdot\|_{2,1}}$ denotes the proximal map with respect to the function $\|\cdot\|_{2,1}\colon \mathbb{R}^{(n_y-1)\times(n_x-1)\times 2} \to \mathbb{R}$ defined as

$$\|q\|_{2,1} = \sum_{i=1}^{n_y-1} \sum_{j=1}^{n_x-1} \sqrt{\left|q_{ij1}\right|^2 + \left|q_{ij2}\right|^2}.$$

The proximal map for this function reads

$$\left(\operatorname{prox}_{\|\cdot\|_{2,1}}(z)\right)_{ijp} = \frac{z_{ijp}}{\sqrt{|z_{ij1}|^2 + |z_{ij2}|^2}} \max\left(\sqrt{|z_{ij1}|^2 + |z_{ij2}|^2} - 1, 0\right),$$

for $i \in \{1, \ldots, n_y - 1\}$, $j \in \{1, \ldots, n_x - 1\}$ and $p \in \{1, 2\}$. We choose the positive step-size parameters $\tau = 1$ and $\sigma = 1/8$, in order to guarantee that (23) converges to a global minimiser of (15) for any initialisation, assuming that one exists. Note that throughout this section, we will always initialise the variables v^0 and q^0 with zeros of the correct dimensions.

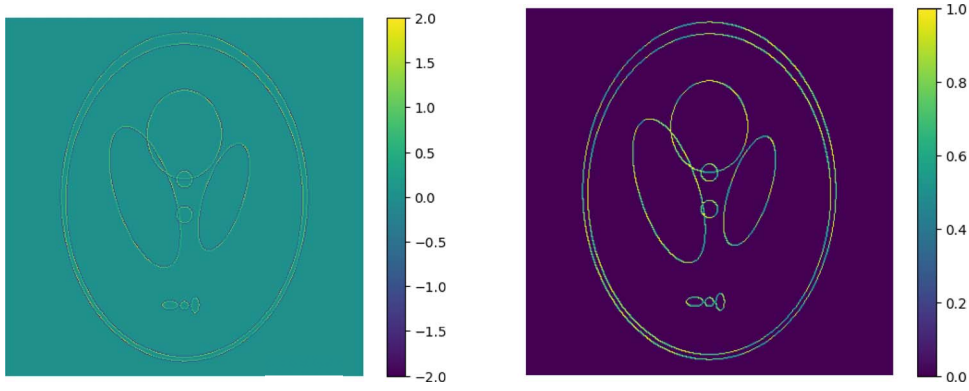
In the following, we show the results for two choices of $u^\dagger$: the famous Shepp–Logan phantom depicted in Fig. 7(b), and (a gray-scale version of) the image of astronaut Eileen Collins depicted in Fig. 10(b). Intuitively, we expect the Shepp–Logan phantom to satisfy the source condition with reasonably low norm of the corresponding source condition element, since it is piecewise constant and therefore well suited for total variation regularisation. The astronaut image of Eileen Collins, on the other hand, is not piecewise constant but textured and contains both jumps and gradual increases in intensity and therefore ill suited for total variation regularisation, so that we expect that a source condition is either not satisfied, or that, if it is satisfied in this discrete setting, it is only satisfied with a large norm for the corresponding source condition element.

4.2.1 *Shepp–Logan phantom.* We begin our discussion of numerical results in this section with the Shepp–Logan phantom. We a use a gray-scale version with $n_y = n_x = 400$ pixels. Before we begin solving the inverse problem of recovering the phantom from low-pass filtered Fourier data, we verify empirically that the Shepp–Logan phantom $u^\dagger$ satisfies the source condition $v \in \partial \mathrm{TV}(u^\dagger)$. In order to compute v, we evaluate (23) with S being the identity operator, without any Fourier operator and with the parameter choices $\tau = 1$ and $\sigma = 1/9$ since $\|A\| < 8$ (cf. Chambolle, 2004), and stop the iteration once the Euclidean norm of the partial derivatives satisfies $\frac{1}{2}(\|A(A^\top q^K - v^K) + \operatorname{prox}_{\|\cdot\|_{2,1}}(Au^\dagger + q^K) - Au^\dagger\| + \|v^K - A^\top q^K\|) \leq 3.84 \times 10^{-14}$, which is chosen because it is very close to machine accuracy and therefore effectively zero. The corresponding iterates v^K and q^K are visualised in Fig. 4. The Euclidean norm of v^K is approximately $\|v^K\| \approx 101.78$, which means that we can accurately quantify error estimates of the form (4) for the Shepp–Logan phantom for the inverse problem of denoising.

We now want to move on to the inverse problem of estimating the Shepp–Logan phantom from low-pass filtered Fourier data. We choose a square low-pass filter of size 130×130 around the centre frequency. The Fourier transform of the Shepp–Logan phantom and the corresponding low-pass filter are visualised in Fig. 5.

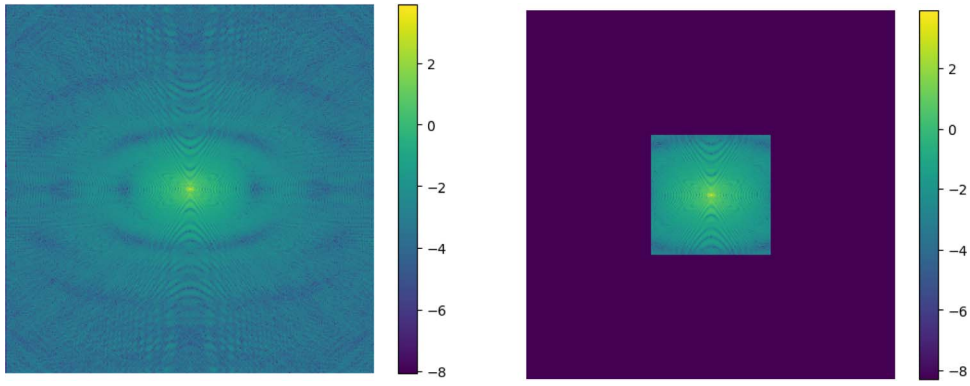
We now employ the same algorithmic strategy, namely (23), but where S corresponds to the low-pass sub-sampling. In contrast to the previous example, the decrease of the norm of the partial derivatives is much slower, and we stop the iteration after 1000 iterations, with an approximate value of 0.12. This indicates that the problem is computationally much harder to solve or that a source condition element does not exist and can therefore not be computed. However, we can certainly use the output v^K and q^K for $K = 1000$ as an approximate source condition element, which we visualise in Fig. 6.

We immediately observe that the scaling of $\mathcal{F}^{-1}S^\top v^K$ is fairly different from v^K in the denoising case, with values mostly in the range of $[-1, 1]$ instead of $[-2, 2]$. Interestingly, the norm of v^K after $K = 1000$ iterations for this low-pass filter example is $\|v^K\| \approx 72.79$, which is significantly smaller than the 101.78 that we encountered in the denoising case. A reason for this could be that we did not solve the optimisation problem to the necessary accuracy. Another reason could be that the small 130×130



(a) v^K (b) $\sqrt{|q_1^K|^2 + |q_2^K|^2}$

FIG. 4. (a) depicts the source condition element v^K that satisfies $v^K \in \partial \mathrm{TV}(u^\dagger)$ for the Shepp–Logan phantom $u^\dagger$. (b) shows the corresponding vector q^K such that $A^\top q^K = v^K$. As one would expect this for an element in the subgradient, the Euclidean norm of q^K with respect to the vector components is bounded by 1.



(a) Fourier transformed data (b) Sub-sampled Fourier transformed data

FIG. 5. The Fourier transform of the Shepp–Logan phantom depicted in Fig. 7(b), and the sub-sampled Fourier transform that emulates a simple low-pass filter. Note that we depict the logarithm of the absolute value of the Fourier transformed data plus the constant $1/4000$ for better visualisation.

window of the low-pass filter, which allows v^k to only have 16900 instead of 160000 non-zero values, has an impact on the norm of the (approximate) source condition element. The latter also explains the difference in scale of the projection $\mathcal{F}^{-1}S^\top v^K$ shown in Fig. 6(b). We also want to emphasise that the Euclidean norm of the vector-field q^K is not strictly bounded by one (even though the visualisation in Fig. 6(c) seems to suggest this), but that some values exceed this threshold. This is proof that for this example a source condition element is approximated, but not found.

Another proof that v^K is only an approximate source condition element can be found with the help of the relation between source and range condition (RC1). With the range condition we can immediately characterise data $g_\alpha = \alpha v + Ku^\dagger$, given a source condition element v, or $g_\alpha^K = \alpha v^K + S\mathcal{F}u^\dagger$ for our example. We set α to the arbitrary value of $\alpha = 1/2$ and compute a solution of (3) from the data g_α^K

(a) v^K (b) Backprojection $\mathcal{F}^{-1}S^\top v^K$ (c) $\sqrt{|q_1^K|^2 + |q_2^K|^2}$

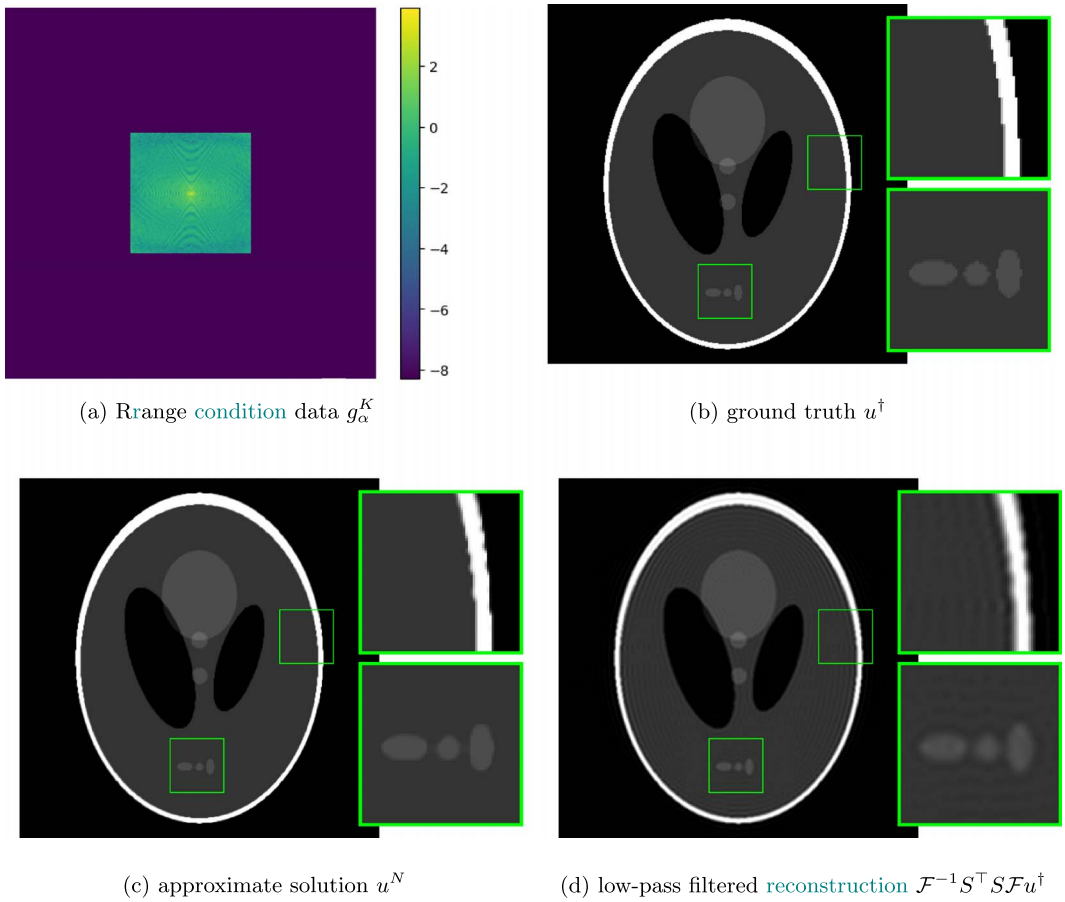
FIG. 6. The approximate source condition element v^K computed with (23) after $K = 1000$ iterations (left), the application of $\mathcal{F}^{-1}S^\top$ to v^K (middle) and the Euclidean vector norm of the vector-field q^K of the corresponding subgradient $A^\top q^K$ (right).

with an implementation of the PDHG method (Chambolle & Pock, 2016). We choose the PDHG version described in Benning & Riis (2021) with step-sizes $\tau = 1/8$ and $\sigma = 1$, and iterate for $N = 1000$ iterations after which the iterates of the primal (u^N) and dual variable (q^N) of the PDHG method satisfy $1/2 * (\|u^N - u^{N-1}\|/\|u^N\| + \|q^N - q^{N-1}\|/\|q^K\|) < 6.85 \times 10^{-5}$. The number of iterations have been capped at 1000 because improvements in reconstruction beyond 1000 iterations is marginal. The results together with the data g_α are visualised in Fig. 7. Please note that while it is straight-forward to verify that an element computed with either (11) or (16) satisfies the source condition (as we have done in Section 4.1 and also for the fully-sampled example in this section), it is not equally straight-forward to disprove existence of a source condition element when the result of either algorithm does not satisfy the source condition. This could also be a result of very slow convergence, and faster optimisation algorithms with guaranteed convergence after a finite number of iterations may be required.

We see from the close-up in Fig. 7 that the reconstruction does not exactly match the Shepp–Logan phantom, which is what one would expect if v^K was a source condition element. However, we nevertheless observe that the reconstruction u^N for data g_α^K is a good approximation of the Shepp–Logan phantom and reasonably better than the traditional low-pass filter as one would expect. Another interesting observation is that if we were to define $u^\dagger = u^N$, we can guarantee that $u^\dagger$ satisfies the source condition with source condition element v that satisfies $\|v\| \approx 72.79$. This is lower than the value of 101.78 that we obtained in the denoising case for the Shepp–Logan phantom, and likely a result of this new $u^\dagger$ being smoother than the Shepp–Logan phantom. However, the nature of the inverse problem with the low-pass filter forward operator might also play a role for the lower value of the norm as it is plausible that for this type of filter errors are amplified less strongly, since the transpose operation $\mathcal{F}^{-1}S^\top$ filters high-frequency errors quite effectively.

We are going to see in Section 4.3 that we can find a more data-adaptive sampling strategy (compared to the low-pass filter) for which we can recover $u_\alpha \in \mathcal{R}_\alpha(g_\alpha)$ almost perfectly from a smaller number of samples.

4.2.2 *Eileen Collins.* We perform identical experiments as in the previous section, but this time we choose an element $u^\dagger$ for which satisfying the source condition is highly unlikely: an image with textures and fine-scale details. We pick the astronaut image of Eileen Collins depicted in Fig. 10(b). We use a gray-scale version that is down-scaled to $n_y = 400$ and $n_x = 400$ pixels as our image $u^\dagger$, and we try to empirically verify if the astronaut image satisfies a source condition of the form $\mathcal{F}^{-1}S^\top v \in \partial \mathrm{TV}(u^\dagger)$.



(a) Rrange condition data g_α^K

(b) ground truth $u^\dagger$

(c) approximate solution u^N

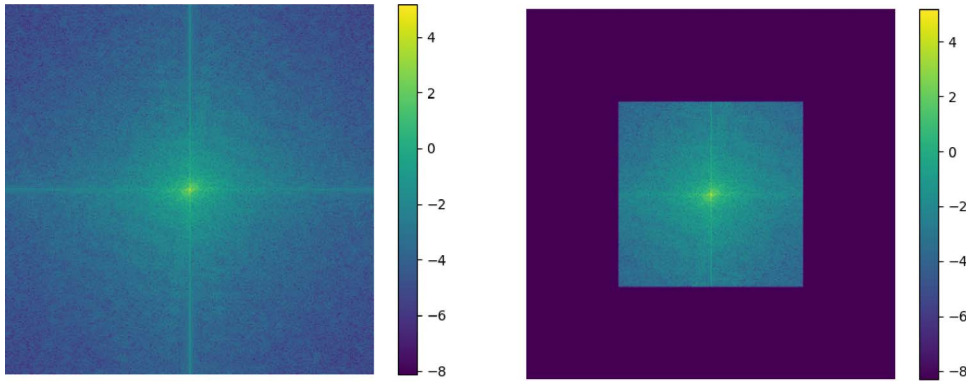
(d) low-pass filtered reconstruction $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$

FIG. 7. This figure shows the (approximate) range condition data $g_\alpha^K = \alpha v^K + S\mathcal{F}u^\dagger$ for $\alpha = 1/2$ in (a). In (b), we see the original Shepp–Logan phantom $u^\dagger$. In (c), the approximate solution u^N of (3) computed with the primal-dual hybrid gradient (PDHG) method for $\alpha = 1/2$ and input data g_α^K. In (d), the linear low-pass filtered reconstruction $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$. For the last three figures, a close-up of two regions of interest is provided (green squares).

Similar to the previous example, we choose a square low-pass filter, but this time of size 200×200 around the centre frequency. The Fourier transform of the Eileen Collins image and the corresponding low-pass filter are visualised in Fig. 8(a) and (b), respectively.

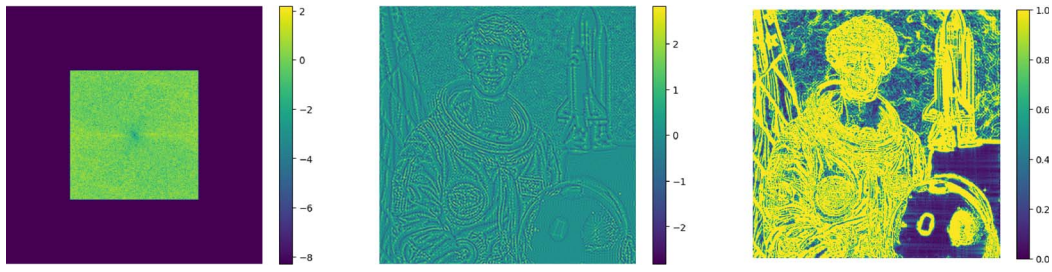
In analogy to the previous section, we evaluate (23) with the parameter choices $\tau = 1$ and $\sigma = 1/8$, initialise with zero arrays and also stop the iteration after $K = 1000$ iterations when the Euclidean norm of the partial derivatives satisfies $\frac{1}{2}(\|A(A^\top q^K - \mathcal{F}^{-1}S^\top v^K + \mathrm{prox}_{\|\cdot\|_{2,1}}(Au^\dagger + q^K) - Au^\dagger\| + \|\mathcal{F}^{-1}S^\top v^K - A^\top q^K\|) \leq 0.59$. The corresponding iterates v^K and q^K are visualised in Fig. 9(a) and (c), respectively. We want to emphasise that the Euclidean norm of the vector-field q^K is not strictly bounded by one (even though the visualisation in Fig. 9c seems to suggest this), but that some values exceed this threshold, which is proof that for this example a source condition element is only approximated, but not found.

In addition, the Euclidean norm of v^K is approximately $\|v^K\| \approx 255.15$ and much larger compared to the Shepp–Logan example, which is what we would expect from a textured image $u^\dagger$ with fine details.



(a) Fourier transformed data (b) Sub-sampled Fourier transformed data

FIG. 8. The Fourier transform of the astronaut image depicted in Fig. 10(b), and the sub-sampled Fourier transform that emulates a simple low-pass filter. Note that we depict the logarithm of the absolute value of the Fourier transformed data plus the constant $1/4000$ for better visualisation.

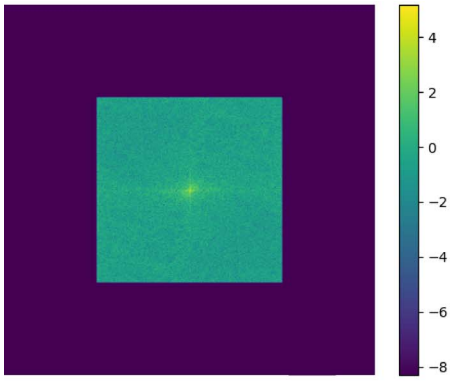


(a) v^K (b) Backprojection $\mathcal{F}^{-1}S^\top v^K$ (c) $\sqrt{|q_1^K|^2+|q_2^K|^2}$

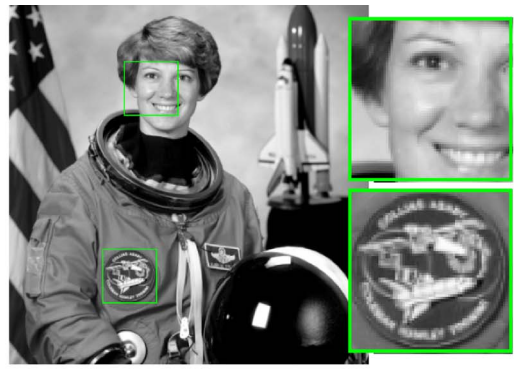
FIG. 9. The approximate source condition element v^K computed with (23) after $K = 1000$ iterations (left), the application of $\mathcal{F}^{-1}S^\top$ to v^K (middle) and the Euclidean vector norm of the vector-field q^K of the corresponding subgradient $A^\top q^K$ (right).

Similarly to the Shepp–Logan phantom, we check solutions of (3) for the range data $g_\alpha = \alpha v + S\mathcal{F}u^\dagger$ for $\alpha = 1/2$. For our example, we define $g_\alpha^K = \alpha v^K + S\mathcal{F}u^\dagger$ and compute a solution of (3) with the PDHG method as described in the previous section (for identical initialisation and parameter choices). After $N = 1000$ iterations, we compute primal (u^N) and dual (q^N) iterates that satisfy $(\|u^N - u^{N-1}\|/\|u^N\| + \|q^N - q^{N-1}\|/\|q^N\|)/2 < 10^{-4}$. The result u^N together with the data g_α^N are visualised in Fig. 10.

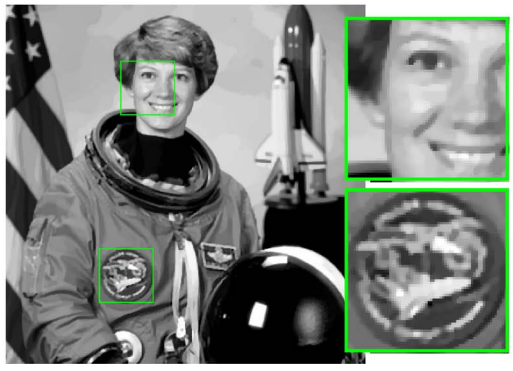
We clearly see that the reconstruction u^N is not identical to $u^\dagger$ but rather a cartoon-like approximation that misses high-scale features such as textures. This does not come as a surprise as we would expect this from a total-variation approximation that does not have access to the high-scale features that the low-pass filter suppresses. We are going to see in Section 4.3 that we can improve the reconstruction by identifying a more adaptive sampling pattern when estimating the approximate source condition element.



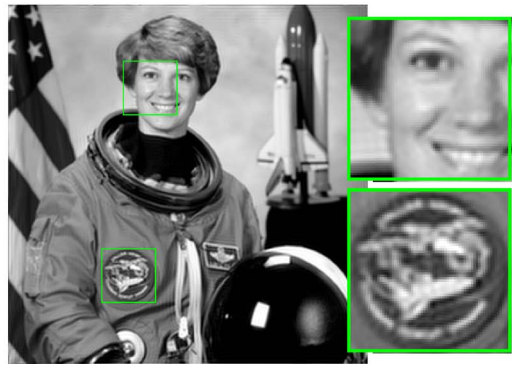
(a) Rrange condition data g_α^K



(b) ground truth $u^\dagger$



(c) approximate solution u^N



(d) low-pass filtered reconstruction $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$

FIG. 10. This figure shows the (approximate) range condition data $g_\alpha^K = \alpha v^K + S\mathcal{F}u^\dagger$ for $\alpha = 1/2$ in (a). In (b), we see the image $u^\dagger$ of Eileen Collins. In (c), the approximate solution u^N of (3) computed with the PDHG method for $\alpha = 1/2$ and input data g_α^K. In (d), the linear low-pass filtered reconstruction $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$. For all three figures, a close-up of two regions of interest is provided (green squares).

4.3 *Optimal sampling in the Fourier domain*

In Section 4.2, we studied the empirical computation of source condition and approximate source condition elements for the variational regularisation of the form

$$u_\alpha \in \operatorname*{arg\,min}_{u\in\mathbb{R}^{n_y\times n_x}} \left\{ \frac{1}{2} \left\| S\mathcal{F}u - f^\delta \right\|^2 + \alpha \mathrm{TV}(u) \right\}.$$

In this section we want to take a step further and also estimate the sampling pattern that defines the sampling operator S. This idea is not new and has applications, for example, in MRI (cf. Sherry *et al.*, 2020). However, in this section, we present a much simpler approach for estimating S compared to works such as Sherry *et al.* (2020). Most importantly, the approach can be phrased as a convex

optimisation problem depending on how we choose to enforce sparsity of $\tilde{v}$, while most alternative approaches constitute non-convex optimisation problems.

The approach is summarised as follows. We assume that S no longer maps onto $\mathbb{C}^m$, but that S maps onto $\mathbb{C}^{n_y \times n_x}$ and that it is a diagonal operator with zero-entries on the diagonal wherever the sampling mask is zero. If we consider the standard source condition $\mathcal{F}^{-1}S^\top v \in \partial \mathrm{TV}(u^\dagger)$ for the above problem, we can define $\tilde{v} := S^\top v \in \mathbb{C}^{n_y \times n_x}$, and estimate $\tilde{v}$ instead of v. Since $\tilde{v}$ has to be sparse by nature in order to emulate sub-sampling, we can estimate $\tilde{v}$ by solving

$$\min_{\tilde{v},q^\dagger} \left\{ \frac{1}{2}\|\mathcal{F}^{-1}\tilde{v} - A^\top q^\dagger\|^2 + G_{\|\cdot\|_{2,1}}(q^\dagger) + \beta J(\tilde{v}) \right\}, \tag{24}$$

where J is a sparsity-inducing regularisation function, e.g. $J(\tilde{v}) = \|\tilde{v}\|_1 = \sum_{i=1}^{n_y}\sum_{j=1}^{n_x} |\tilde{v}_{ij}|$ where $|\cdot|$ denotes the complex modulus, and $\beta > 0$ is a penalisation parameter that controls the level of sparsity of $\tilde{v}$. We then minimise (24) via the proximal alternating linearised minimisation (PALM) algorithm (Yangyang & Yin, 2013; Bolte *et al.*, 2014), i.e. we approximate a solution of (24) by iterating

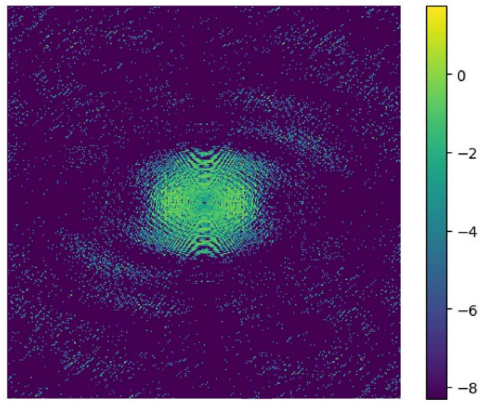
$$\begin{aligned} \tilde{v}^{k+1} &= \mathrm{prox}_{\beta J}\left(\tilde{v}^k - \tau\left(\tilde{v}^k - \mathcal{F}A^\top q^k\right)\right), \\ q^{k+1} &= q^k - \sigma\left(A\left(A^\top q^k - \mathcal{F}^{-1}\tilde{v}^{k+1}\right) + \mathrm{prox}_{\|\cdot\|_{2,1}}\left(q^k + Au^\dagger\right) - Au^\dagger\right), \end{aligned} \tag{25}$$

for $k \in \mathbb{N}$, initial values $\tilde{v}^0$ and q^0 and suitable step-size parameters τ and σ.

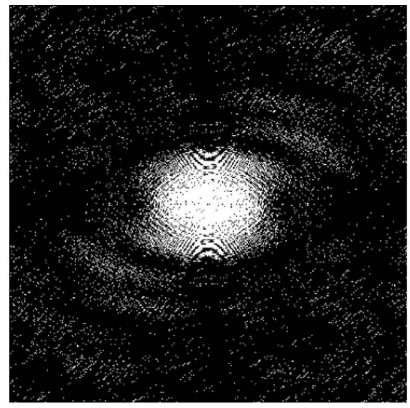
In the following, we approximate $\tilde{v}^K$ and q^K for the arbitrary choice $K = 1000$. We then estimate the mask that determines the zero and non-zero entries on the diagonal of S by simply identifying the zero and non-zero entries of $\tilde{v}$. Please note that we manually enforce that the lowest frequency is included in the mask, as it otherwise would be excluded since total variation subgradients have zero mean. Subsequently, we estimate v by solving (23) with the estimated sub-sampling operator S from the previous step, for $K = 1000$ iterations. We have conducted experiments for the same examples visualised in Fig. 7(b) and (b), which are described in the next two sections, and we compare both sampling pattern and reconstructions to those of Sherry *et al.* (2020). For the latter, we use the code provided here. Please note that due to the high computational cost and runtime of the method in Sherry *et al.* (2020), we cap the outer iterations of the bilevel optimisation algorithm at 250, and only operate on down-scaled 256×256 version of the images.

4.3.1 *Shepp–Logan phantom.* We begin with the Shepp–Logan phantom as seen in Fig. 7(b) and compute the corresponding element $\tilde{v}^K$ via (25) with zero-initialisation, $J = \beta\|\cdot\|_1$ and the parameter $\beta = 0.1$. The corresponding $\tilde{v}^K$, the mask of all non-zero coefficients and the mask of the 16900 largest Fourier coefficients (in magnitude) are visualised in Fig. 12. Please note that for $\beta = 0.1$ the number of non-zero coefficients of $\tilde{v}^K$ after $K = 1000$ iterations is 14553, which is comparable but even slightly lower than the 16900 non-zero coefficients that were used in the low-pass filter example in Section 4.2.1. The number of non-zero coefficients of the approach in Sherry *et al.* (2020) after 250 iterations with the same sparsity parameter β is 18642, which is much larger and likely a result of the early termination as a consequence of the high computational cost.

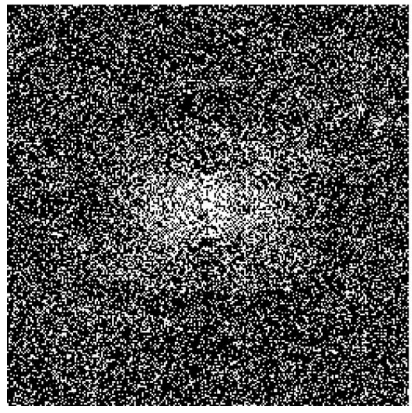
We want to emphasise that the learned mask differs substantially from the mask that stems from the largest Fourier coefficients (in magnitude). In particular, low-frequency information is traded in for high-frequency information since the total variation regularisation is very effective at establishing information



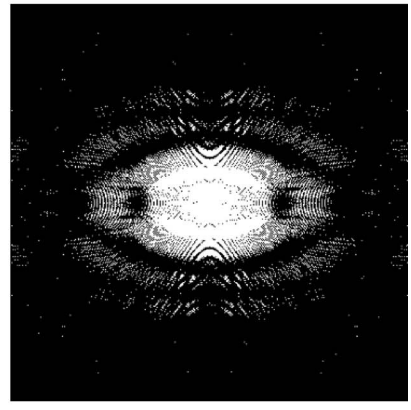
(a) $\tilde{v}^K$



(b) Sampling pattern



(c) Learned sampling pattern with [Sherry *et al.* (2020)]



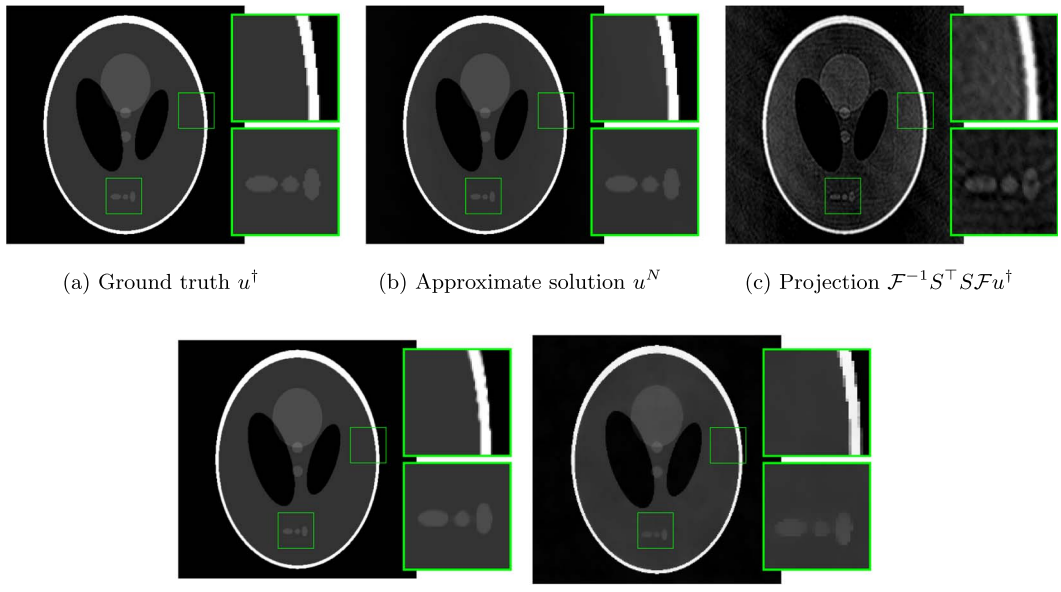
(d) Largest Fourier coefficients

FIG. 11. (a) shows the (approximate) source condition element for the Shepp–Logan phantom computed with (25). (b) shows the corresponding mask with a 9% ratio of non-zero to zero coefficients. For comparison, (c) shows the mask learned with the approach proposed in Sherry *et al.* (2020) with a 28.45% ratio, and (d) shows the mask of the Fourier coefficients with largest magnitude with a 10.5% ratio of non-zero to zero coefficients.

from limited low-frequency data but very ineffective at generating high-frequency information. What is a little bit surprising is that the learned sampling pattern with Sherry *et al.* (2020) also differs from the learned mask with the proposed approach; it appears to be more random, albeit with a heavy weighting of centre frequencies.

Next, given the mask, we estimate v via (23) and obtain v^K for $K = 1000$. Note that the corresponding norm is $\|v^K\| \approx 113.93$, which is slightly larger but fairly comparable to the norm that we established in the denoising setting. Subsequently, we perform another sanity-check and compute an approximation of (3) for data $g_\alpha^K = \alpha v^K + S\mathcal{F}u^\dagger$ for $\alpha = 1/2$ via the PDHG method with the same initialisation and parameter configurations as described in Section 4.2.1 and visualise the results in Fig. 12.

In comparison to Section 4.2.1, we observe that we recover the Shepp–Logan phantom almost perfectly whilst using a sampling operator that uses fewer samples than the low-pass sampling operator.



(a) Ground truth $u^\dagger$ (b) Approximate solution u^N (c) Projection $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$

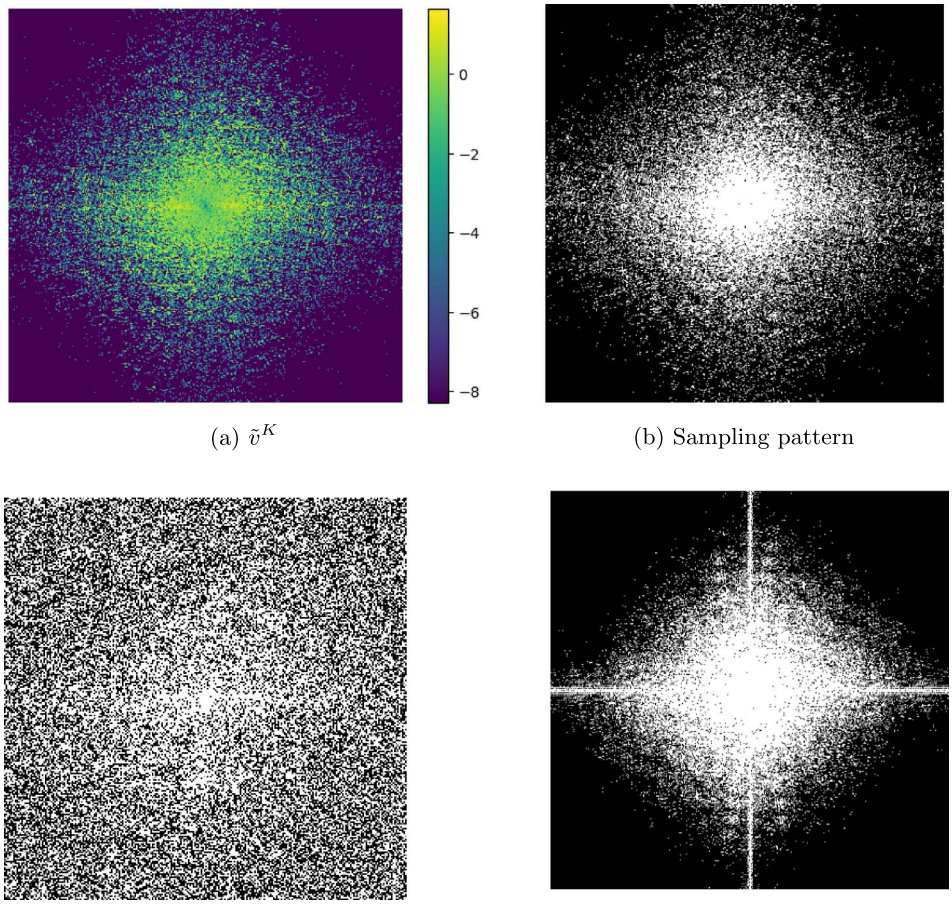
(d) Low-pass filtered reconstruction (e) Reconstruction with [Sherry *et al.* (2020)]

FIG. 12. In (a), we show the Shepp–Logan phantom ; in (b) the approximation u^N of (3) with the learned mask and corresponding range data ; in (c), the projection $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$; in (d), the approximation u^N of (3) with the low-pass filter from Section 4.2.1; in (e), the reconstruction based on Sherry *et al.* (2020). For all figures, a close-up of two regions of interest is provided (green squares).

Please note that the reconstruction also outperforms the reconstruction with Sherry *et al.* (2020), despite the lower sampling ratio. This is likely a consequence of the Huber approximation of the total variation that has to be used in Sherry *et al.* (2020).

4.3.2 *Eileen Collins.* We now proceed to the image of Eileen Collins as depicted in Fig. 14(a) and perform the same set of tasks as described in the previous section, i.e. we compute the element $\tilde{v}^K$ via (25) with zero-initialisation and for $J = \beta\|\cdot\|_1$ for which we choose $\beta = 0.24$ this time. The corresponding $\tilde{v}^K$, the mask of all non-zero coefficients and the mask of the 40000 largest Fourier coefficients (in magnitude) are visualised in Fig. 14. Please note that for $\beta = 0.24$, the number of non-zero coefficients of $\tilde{v}^K$ after $K = 1000$ iterations is 38262, which is comparable but even slightly lower than the 40000 non-zero coefficients that were used in the low-pass filter example in Section 4.2.2. The number of non-zero coefficients of the approach in Sherry *et al.* (2020) after 250 iterations with the same sparsity parameter β is 59197, which is much higher and, like for the Shepp–Logan phantom, likely a result of the early termination of the algorithm.

We see that the learned mask in Fig. 13(b) differs from that obtained from the Fourier coefficients with largest magnitude in Fig. 13(d). In particular, the information that corresponds to coarse edges in the image $u^\dagger$ is less important for the total variation-based model, since it can interpolate this type of missing information rather well. Instead, more higher frequencies corresponding to texture information are being sampled as it is impossible for a total variation-based model to generate textures.



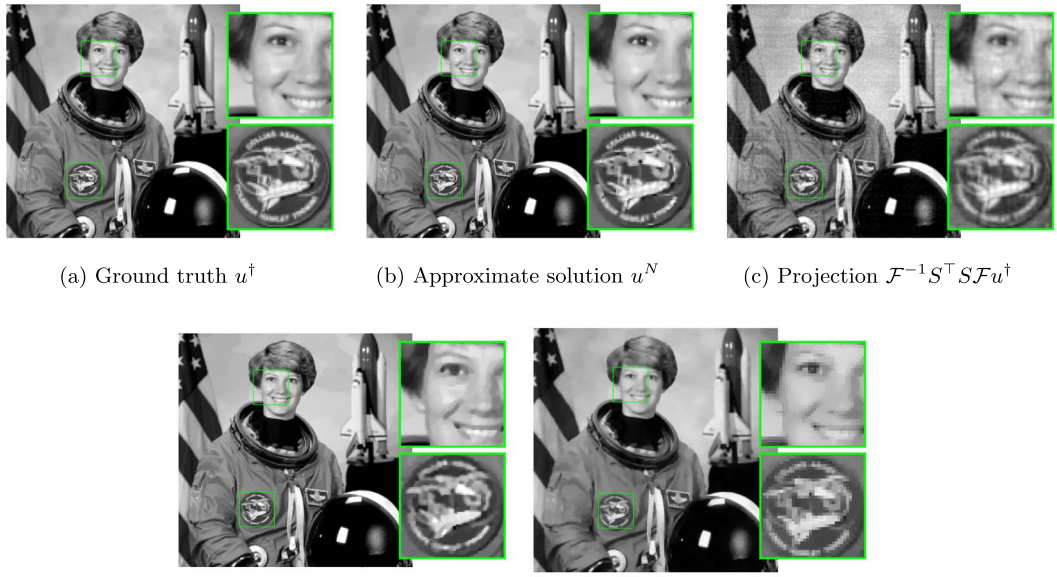
(a) $\tilde{v}^K$ (b) Sampling pattern

(c) Learned sampling pattern with [Sherry *et al.* (2020)] (d) Largest Fourier coefficients

FIG. 13. (a) shows the (approximate) source condition element for the Eileen Collins image computed with (25). (b) shows the corresponding mask with a ratio of 24% non-zero vs zero coefficients. For comparison, (c) shows the mask learned with the approach proposed in Sherry *et al.* (2020) with a ratio of 90.3%, and (d) shows the mask of the Fourier coefficients with largest magnitude with a ratio of 25%.

Given the mask, we estimate v via (23) and obtain v^K for $K = 1000$. Note that the corresponding norm is $\|v^K\| \approx 302.47$, which is larger than the norm that we established in the low-pass filter setting. Subsequently, we perform another sanity-check and compute an approximation of (3) for data $g_\alpha^K = \alpha v^K + S\mathcal{F}u^\dagger$ for $\alpha = 1/2$ via the PDHG method with the same initialisation and parameter configurations as described in Section 4.2.2 and visualise the results in Fig. 14.

In comparison to Section 4.2.2, we observe that we still have not found a source condition element, which is not surprising. We do observe, however, that the recovery of (3) with the range data for the learned sampling operator is much better compared to the recovery with the low-pass sampling operator. We observe several high-frequency features such as the specular highlights in the eyes that are not present in the low-pass filter reconstruction. Please note that the reconstruction with the learned sampling operator uses a smaller amount of samples than the reconstruction with the low-pass sampling operator.



(a) Ground truth $u^\dagger$ (b) Approximate solution u^N (c) Projection $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$

(d) Low-pass filtered reconstruction (e) Reconstruction with [Sherry *et al.* (2020)]

FIG. 14. In (a), we show the image of Eileen Collins; in (b), the approximation u^N of (3) with the learned mask and corresponding range data; in (c), the projection $\mathcal{F}^{-1}S^\top S\mathcal{F}u^\dagger$; in (d), the approximation u^N of (3) with the low-pass filter from Section 4.2.2; in (e), the reconstruction based on Sherry *et al.* (2020). For all figures, a close-up of two regions of interest is provided (green squares).

5. Conclusions and outlook

We conclude this work with a brief summary of its main findings, and provide an outlook of research topics that we have not been able to address but that make for interesting future research.

5.1 *Conclusions*

In this paper, we have pursued the challenging task of estimating source condition elements as tools for the quantitative analysis of variational regularisation of linear inverse problems. Specifically,

- we considered a rather general Banach space setup, which encompasses a large class of regularisation methods of interest in applications;
- we reformulated the source conditions (and the closely related range conditions) as the solution of convex minimisation problems by means of tools from convex analysis;
- we provided iterative algorithms for the numerical approximation of source (and range) condition elements, which pave the way to quantitative error estimates;
- to demonstrate the performance of the proposed approach, we enclosed a significant set of numerical experiments for processing both 1D and 2D signals;

- we described how the introduced framework can provide insightful inspiration for novel approaches in optimal measurement design: in particular, the proposed procedure of optimal subsampling in the Fourier domain provided promising results.

5.2 *Outlook*

There are numerous research aspects that we have either addressed only briefly or have not addressed at all. One obvious aspect is that source and range conditions are not only relevant for convergence rates for variational regularisations but other regularisations such as iterative regularisations, too (cf. Burger *et al.*, 2007; Benning & Burger 2018a). Hence, in order to quantify error estimates for such regularisations, identical strategies as proposed in this paper can be deployed. The same also holds true for regularisation methods that are based on more general data fidelity terms as described in (2), for which error estimates also rely on source or range conditions (Benning & Burger, 2011).

Further, we did not look into stronger source conditions or variational source conditions as outlined in the introduction, but it should be straight forward to design convex minimisation problems similar to the ones presented in this work for the computation of, e.g. strong source condition elements. We also maintained focus solely on linear inverse problems, while source conditions play a pivotal role for convergence rates of nonlinear inverse problems, too.

Another interesting direction for research is the computation of generalised eigenfunctions or singular vectors as addressed in the introduction. Suppose we are given a function f and we would like to find a function u_λ that satisfies $\lambda u_\lambda \in J(u_\lambda)$ such that it is close to f, then we can formulate the constrained minimisation problem

$$\min_u \; B_J(u, (1+\lambda)u) \qquad \text{subject to} \qquad \|u - f\| \leq \delta,$$

for hyperparameters $\lambda > 0$ and $\delta \geq 0$. If a solution u_λ satisfies $B_J(u_\lambda, (1+\lambda)u_\lambda) = 0$, we can conclude that it is a generalised eigenfunction with eigenvalue λ and closest to f within the ball of radius δ.

Last but not the least, we want to emphasise that extensions of the proposed minimisation problems can find applications in a wide range of data-driven inverse problems applications such as operator correction and learning data-driven regularisation functionals that we want to briefly describe in the following two sections.

5.2.1 *Operator correction.* In analogy to the optimal sampling example presented in Section 4.3, one can modify the proposed approaches to perform more general operator corrections beyond sampling. Taking the polynomial regression problem from Section 4.1 as an example, one can consider the following extension of the classical LASSO approach:

$$u_\alpha \in \operatorname*{arg\,min}_{u \in \ell^2 \cap \ell^1} \left\{ \frac{1}{2} \left\| A \left(Ku - f^\delta \right) \right\|^2 + \alpha \|u\|_1 \right\},$$

where $\ell^2 \cap \ell^1$ denotes the intersection of the spaces ℓ^2 and ℓ^1. In this example, the goal is to estimate not only the source condition element v, but to also estimate a pre-conditioner matrix A. We could do so

by formulating the non-convex optimisation problem

$$\min_{A,\{v_i\}_{i=1}^s} \left[\sum_{i=1}^{s} B_{\|\cdot\|_1} \left(u_i^\dagger, u_i^\dagger + K^\top A^\top v_i \right) + \chi_{\|\cdot\| \leq \beta}(v_i) \right],$$

where $\chi_{\|\cdot\| \leq \beta}$ denotes the characteristic function that for an argument v is zero if $\|v\| \leq \beta$ and infinity otherwise. Here we minimise an empirical risk for s vectors of polynomial coefficients $\{u_i^\dagger\}_{i=1}^s$ subject to the constraint that the norm of the source condition elements v_i should not exceed the threshold $\beta > 0$. If the data are representative, minimising this empirical risk can help preconditioning the forward model and the data to lower the norms of the corresponding source condition elements, which in return ensures better convergence rates of the solution of u_α towards true coefficients $u^\dagger$.

5.2.2 *Construction of data-driven regularisations.* One can extend the findings from Section 3 to design data-driven variational regularisation operators R_α with convergence rate $\mathcal{O}(\delta)$ and favourable error amplification constant $\|v\|$ in the error estimate (4). In order to do so, we can make the assumption that we have a parametrised variational regularisation operator R_α of the form (3) with a regularisation function of the form $J(u) = H(Au + b)$. We can then aim at estimating the linear operator A and the function b by minimising an empirical risk of the form

$$\frac{1}{s} \sum_{i=1}^{s} \left[\Gamma(v_i) + \frac{1}{2} \|K^* v_i - A^* q_i^\dagger\|^2 + B_H \left(A u_i^\dagger + b, q_i^\dagger + A u_i^\dagger + b \right) \right],$$

for some suitable regularisation function Γ. In order to keep the error amplification constants $\|v_i\|$ bounded, one intuitive choice for Γ is

$$\Gamma(v) = \chi_{\|\cdot\| \leq \beta}(v) \colon = \begin{cases} 0 & \text{if } \|v\| \leq \beta \\ \infty & \text{if } \|v\| > \beta \end{cases},$$

for a positive constant $\beta > 0$, similar to the operator correction example in Section 5.2.1. Then one can formulate the constrained, non-convex minimisation problem

$$\min_{\{v_i\}_{i=1}^s, \{q_i^\dagger\}_{i=1}^s, A, b} \left\{ \sum_{i=1}^{s} \left[\chi_{\|\cdot\| \leq \alpha}(v_i) + \frac{1}{2} \|K^* v_i - A^* q_i^\dagger\|^2 + B_H \left(A u_i^\dagger + b, q_i^\dagger + A u_i^\dagger + b \right) \right] \right\}.$$

Similar to the idea proposed in Section 5.2.1, minimising such an empirical risk can help identifying a suitable operator A (and shift b) to lower the norms of the corresponding source condition elements, which in return ensures better convergence rates of the solution of the variational regularisation method towards the solution $u^\dagger$ of the inverse problem (1).

Acknowledgements

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme 'Mathematics of Deep Learning' where work on this paper was undertaken.

Funding

EPSRC (Grant No. EP/R014604/1); INdAM-GNCS, INdAM-GNAMPA; Alan Turing Institute (to M.B.). Air Force Office of Scientific Research (award number FA8655-20-1-7027 to L.R.), Fondazione Compagnia di San Paolo. PNRR - M4C2 - Investimento 1.3. Partenariato Esteso PE00000013 - 'FAIR - Future Artificial Intelligence Research' - Spoke 8 'Pervasive AI', funded by the European Commission under the NextGeneration EU programme; EPSRC (grant EP/R513106/1 to D.R.).

Data availability statement

The Python codes for this paper are available at https://github.com/reacho/source-condition.

References

Andrey, N. (1943) Tikhonov, on the stability of inverse problems. *Dokl. Akad. Nauk SSSR*, **39**, 195–198.

Beck, A. & Tetruashvili, L. (2013) On the convergence of block coordinate descent type methods. *SIAM J. Optim.*, **23**, 2037–2060.

Benning, M. & Burger, M. (1963) Solution of incorrectly formulated problems and the regularization method. *Soviet Math.*, **4**, 1035–1038.

Benning, M. & Burger, M. (2011) Error estimates for general fidelities. *Electron. Trans. Numer. Anal.*, **38**, 77.

Benning, M. & Burger, M. (2013) Ground states and singular vectors of convex variational regularization methods. *Methods Appl. Anal.*, **20**, 295–334.

Benning, M. & Burger, M. (2014) A total variation spectral framework for scale and texture analysis. *SIAM J. Imaging Sci.*, **7**, 1937–1961.

Benning, M. & Burger, M. (2018a) Modern regularization methods for inverse problems. *Acta Numer.*, **27**, 1–111.

Benning, M. & Burger, M. (2018b) *Nonlinear Eigenproblems in Image Processing and Computer Vision*. Springer. https://doi.org/10.1007/978-3-319-75847-3.

Benning, M. & Riis, E. S. (2021) Bregman methods for large-scale optimisation with applications in imaging. In: Chen, Keand, S., Tai, C-B., Younces, X-C., Laurent, (eds), *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*. Cham: Springer, pp. 1–42. https://doi.org/10.1007/978-3-030-03009-4_62-1.

Bolte, J., Sabach, S. & Teboulle, M. (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Programming*, **146**, 459–494.

Bozorgnia, F., Bungert, L. & Tenbrinck, D. (2020) The infinity Laplacian eigenvalue problem: reformulation and a numerical scheme. *Journal of Scientific Computing*, **98**(2), 40. https://doi.org/10.1007/s10915-023-02425-w.

Bregman, L. M. (1967) The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.*, **7**, 200–217.

Bubba, T. A., Burger, M., Helin, T. & Ratti, L. (2023) Convex regularization in statistical inverse learning problems. *Inverse Probl. Imaging*, **17**, 1193–1225.

Bungert, L., Hait-Fraenkel, E., Papadakis, N. & Gilboa, G. (2021) Nonlinear power method for computing eigenvectors of proximal operators and neural networks. *SIAM J. Imaging Sci.*, **14**, 1114–1148.

Burger, M. & Osher, S. (2004) Convergence rates of convex variational regularization. *Inverse Probl.*, **20**, 1411.

Burger, M., Eckardt, L., Gilboa, G. & Moeller, M. (2015) Spectral representations of one-homogeneous functionals. In: Aujol, J-F., Nikolova, M., Papadakis, N. (eds), *Scale Space and Variational Methods in Computer Vision: 5th International Conference, SSVM, Lège-cap ferret, France, May 31-June 4, 2015, Proceedings*. Cham: Springer, Vol. **9087**, pp. 16–27. https://doi.org/10.1007/978-3-319-18461-6_2.

Burger, M., Gilboa, G., Moeller, M., Eckardt, L. & Cremers, D. (2016) Spectral decompositions using one-homogeneous functionals. *SIAM J. Imaging Sci.*, **9**, 1374–1408.

Burger, M., Helin, T. & Kekkonen, H. (2018) Large noise in variational regularization. *Trans. Math. Appl.*, **2**, tny002.

Burger, M., Resmerita, E. & He, L. (2007) Error estimation for Bregman iterations and inverse scale space methods in image restoration. *Computing*, **81**, 109–135.

Candès, E. J., Romberg, J. & Tao, T. (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, **52**, 489–509.

Chambolle, A. & Pock, T. (2016) An introduction to continuous optimization for imaging. *Acta Numer.*, **25**, 161–319.

Chambolle, A. (2004) An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, **20**, 73–87.

Chavent, G. & Kunisch, K. (1997) Regularization of linear least squares problems by total bounded variation. *ESAIM*, **2**, 359–376.

Daubechies, I., Defrise, M. & De Mol (2004) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, **57**, 1413–1457.

Donoho, D. L. (1992) Superresolution via sparsity constraints. *SIAM J. Math. Anal.*, **23**, 1309–1331.

Ekeland, I. & Temam, R. (1999) *Convex Analysis and Variational Problems*. SIAM. https://doi.org/10.1137/1.9781611971088.

Engl, H. W., Hanke, M. & Neubauer, A. (1996) *Regularization of Inverse Problems*, vol. **375**. Springer Science & Business Media.

Engl, H. W., Kunisch, K. & Neubauer, A. (1989) Convergence rates for Tikhonov regularisation of non-linear ill-posed problems. *Inverse Probl.*, **5**, 523.

Flemming, J. (2012) *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*. Shaker-Verlag.

Gilboa, G., Moeller, M. & Burger, M. (2016) Nonlinear spectral analysis via one-homogeneous functionals: overview and future prospects. *J. Math. Imaging Vision*, **56**, 300–319.

Gilboa, G. (2014) Nonlinear band-pass filtering using the tv transform. In: *22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal: IEEE, pp. 1696–1700.

Grasmair, M., Haltmeier, M. & Scherzer, O. (2008) Sparse regularization with ℓ^q penalty term. *Inverse Probl.*, **24**, 055020.

Grasmair, M., Scherzer, O. & Haltmeier, M. (2011) Necessary and sufficient conditions for linear convergence of ℓ^1*-regularization*. *Comm. Pure Appl. Math.*, **64**, 161–182.

Hein, T. & Hofmann, B. (2009) Approximate source conditions for nonlinear ill-posed problems–chances and limitations. *Inverse Probl.*, **25**, 035003.

Hofmann, B., Düvelmeyer, D., Krumbiegel, K., Düvelmeyer, D. & Krumbiegel, K. (2006) Approximate source conditions in Tikhonov regularization-new analytical results and some numerical studies. *Math. Model. Anal.*, **11**, 41–56.

Hofmann, B., Kaltenbacher, B., Pöschl, C. & Scherzer, O. (2007) A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Probl.*, **23**, 987.

Hohage, T. & Miller, P. (2019) Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces. *Inverse Probl.*, **35**, 065005.

Hohage, T. & Weidling, F. (2017) Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *SIAM J. Numer. Anal.*, **55**, 598–620.

Leonid, I., Rudin, S. O. & Fatemi, E. (1992) Nonlinear total variation based noise removal algorithms. *Phys. D*, **60**, 259–268.

Lions, P.-L. & Mercier, B. (1979) Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, **16**, 964–979.

Mukherjee, S., Schönlieb, C.-B. & Burger, M. (2021) Learning convex regularizers satisfying the variational source condition for inverse problems. *NeurIPS*, Workshop on Deep Learning and Inverse Problems. https://openreview.net/forum?id=32oLu7d1vlT.

Mumford, D. B. & Shah, J. (1989) Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, **42**, 577–685.

Nesterov, Y. (1983) A method of solving a convex programming problem with convergence rate$\mathcal{O}\left(\frac{1}{k^2}\right)$. *Dokl. Akad. Nauk USSR*, **269**, 543–547.

Nossek, R. Z. & Gilboa, G. (2018) Flows generating nonlinear eigenfunctions. *J. Sci. Comput.*, **75**, 859–888.

Ramlau, R. & Resmerita, E. (2010) Convergence rates for regularization with sparsity constraints. *Electron. Trans. Numer. Anal.*, **37**, 87–104.

Resmerita, E. (2005) Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Probl.*, **21**, 1303.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M. & Lenzen, F. (2009) *Variational Methods in Imaging*. New York: Springer.

Schmidt, M. F., Benning, M. & Schönlieb, C.-B. (2018) Inverse scale space decomposition. *Inverse Probl.*, **34**, 045008.

Schuster, T., Kaltenbacher, B., Hofmann, B. & Kazimierski, K. S. (2012) *Regularization Methods in Banach Spaces*, vol. **10**. Berlin, Boston: Walter de Gruyter.

Sherry, F., Benning, M., de los Reyes, J. C., Graves, M. J., Maierhofer, G., Williams, G., Schonlieb, C.-B. & Ehrhardt, M. J. (2020) Learning the sampling pattern for MRI. *IEEE Trans. Med. Imaging*, **39**, 4310–4321.

Tatiana, A. B. & Ratti, L. (2022) Shearlet-based regularization in statistical inverse learning with an application to X-ray tomography. *Inverse Probl.*, **38**, 054001.

Tautenhahn, U. (1998) Optimality for ill-posed problems under general source conditions. *Numer. Funct. Anal. Optim.*, **19**, 377–398.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *J. R. Statist. Soc.: Ser. B (Methodol.)*, **58**, 267–288.

Wang, X. & Benning, M. (2023) Lifted Bregman training of neural networks. *J. Mach. Learn. Res.*, **24**, 1–51.

Wright, S. J. (2015) Coordinate descent algorithms. *Math. Programming*, **151**, 3–34.

Yangyang, X. & Yin, W. (2013) A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.*, **6**, 1758–1789.