GROUND STATES AND SINGULAR VECTORS OF CONVEX VARIATIONAL REGULARIZATION METHODS[*]

MARTIN BENNING[†] AND MARTIN BURGER[‡]

Abstract. Singular value decomposition is the key tool in the analysis and understanding of linear regularization methods in Hilbert spaces. Besides simplifying computations it allows to provide a good understanding of properties of the forward problem compared to the prior information introduced by the regularization methods. In the last decade nonlinear variational approaches such as ℓ^1 or total variation regularizations became quite prominent regularization techniques with certain properties being superior to standard methods. In the analysis of those, singular values and vectors did not play any role so far, for the obvious reason that these problems are nonlinear, together with the issue of defining singular values and singular vectors in the first place.

In this paper however we want to start a study of singular values and vectors for nonlinear variational regularization of linear inverse problems, with particular focus on singular one-homogeneous regularization functionals. A major role is played by the smallest singular value, which we define as the ground state of an appropriate functional combining the (semi-)norm introduced by the forward operator and the regularization functional. The optimality condition for the ground state further yields a natural generalization to higher singular values and vectors involving the subdifferential of the regularization functional, although we shall see that the Rayleigh principle may fail for higher singular values.

Using those definitions of singular values and vectors, we shall carry over two main properties from the world of linear regularization. The first one is gaining information about scale, respectively the behavior of regularization techniques at different scales. This also leads to novel estimates at different scales, generalizing the estimates for the coefficients in the linear singular value expansion. The second one is to provide classes of exact solutions for variational regularization methods. We will show that all singular vectors can be reconstructed up to a scalar vector by the standard Tikhonov-type regularization approach even in the presence of (small) noise. Moreover, we will show that they can even be reconstructed without any bias by the recently popularized inverse scale space method.

Key words. Inverse problems, variational regularization, singular values, ground states, total variation regularization, Bregman distance, inverse scale space method, compressed sensing.

AMS subject classifications. 35P30, 45Q05, 49N45, 47A52.

1. Introduction. Regularization methods and their analysis are a major topic in inverse problems and image processing. In the last century mainly linear regularization methods for problems in Hilbert spaces have been studied and analyzed, and it seems that for such methods a quite complete theory is now available based on singular value decomposition (cf. [64]) of the forward operator in the norm defined by the regularization (cf. [36]), respectively generalizations to spectral decompositions in the rare cases of non-compact forward operators (cf. [36]).

The research in the 21st century has significantly shifted from linear regularizations to nonlinear approaches, in particular variational methods generalizing Tikhonov regularization, where singular regularization functionals such as ℓ^1-norms or the total variation are used. In many examples the above mentioned functionals have shown to yield improved properties with respect to the incorporation of prior knowledge and the quality of reconstructions, and have also been a key tool in the adjacent theory of compressed sensing (cf. [25, 26, 33, 32]). Various advances in the analysis of such regular-

ization methods have been made over the last years, ranging from basic regularization properties (cf. e.g. [1, 28]) over error estimation (cf. [22, 23, 56, 57, 11, 43, 49]) to corrections of inherent bias by iterative and time-flow techniques (cf. [60, 54, 19, 18, 71]). Singular values and vectors did so far not play any role in the analysis of such methods and it is common belief that their use is restricted to linear regularization methods. This is not surprising, since first of all it is not trivial to define a notion of singular values and to characterize it in the nonlinear case. Moreover, it is obvious that due to missing linearity no decomposition into singular values can be achieved. For these reasons the study of singular values (or eigenvalues of regularization functionals) has been mainly abandoned in the inverse problems community, studies of related nonlinear eigenvalue problems rather exist in nonlinear partial differential equations and functional inequalities (cf. [2, 12, 16, 29, 31, 38, 41, 45, 46, 47, 68]), in control theory (cf. [39]), in image processing (cf. [4, 5, 9, 27]) and surprisingly in machine learning (cf. [17, 66]). Motivated by those as well as general approaches to nonlinear eigenvalue problems we will define a ground state by a Rayleigh-type principle and further singular values and singular vectors by considering the first-order optimality condition for the non-convex variational problem defining ground states. The main results we shall derive are the following:

- First of all, our definition of singular values and singular vectors is studied and demonstrated to be a meaningful extension of the linear case, although some properties can be lost in extreme cases, e.g. the discreteness of the spectrum and the Rayleigh principle for higher singular values (Section 3).
- With the singular values and singular vectors we derive error estimates for appropriate linear functionals of the solution, which provide information about the behavior at different scales. This is made explicit for an example in total variation denoising (Section 4).
- An important part is to verify that singular vectors are exact solutions of variational regularization schemes. This means that if the image of a singular vector under the forward operator is used for the reconstruction, the solution is a multiple of the singular vector. Surprisingly, under certain conditions particularly met for singular regularizations, the same holds true if a certain amount of noise is added. For inverse scale space methods we can further verify that singular vectors are reconstructed without bias, i.e. after finite time (depending on the singular value) the solution of the inverse scale space equals exactly the singular vector, without a multiplicative change (Sections 5 and 7).
- We derive estimates on the bias of variational regularization schemes, which show that the minimal bias is somehow defined by the ground state, respectively by the smallest singular value in our definition (Section 6).
- We provide a variety of examples of inverse problems and regularization functionals, for which singular values and singular vectors can be computed explicitly. This allows to draw various conclusions about the behavior of the regularization and the typical shape of preferred solutions (Section 8).

2. Notations and assumptions. To fix notation, we consider linear inverse problems of the form

$$(2.1) \qquad\qquad\qquad Ku = f\,,$$

with $K : \mathcal{U} \to \mathcal{H}$ being a linear operator mapping from a Banach space $\mathcal{U}$ to a Hilbert space $\mathcal{H}$, with the goal to recover u from (2.1) with given data f that is potentially

being corrupted by noise. We are mainly interested in the case of K being compact, in particular continuous from the weak or weak-* topology of $\mathcal{U}$ to the strong topology of $\mathcal{H}$, which creates ill-posedness of the inverse problems.

Nonlinear variational regularization methods for computing robust approximate solutions of (2.1) are of the form

$$(2.2) \qquad \hat{u} \in \operatorname*{arg\,min}_{u \in \operatorname{dom}(J)} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{H}}^2 + \alpha J(u) \right\} ,$$

with $J : \operatorname{dom}(J) \subseteq \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$ being a so-called regularization functional that incorporates the a-priori knowledge, and $\alpha \in \mathbb{R}_{>0}$ denoting the regularization parameter that controls the impact of J on the solution $\hat{u}$ of (2.2). Note that in the variational approach linear regularizations methods are related to quadratic regularization functionals like

$$J(u) = \frac{1}{2}\|Du\|_{\mathcal{U}}^2$$

for linear operators $D : \mathcal{U} \to \mathcal{U}$, since they lead to linear optimality conditions. Another classical choice motivated from statistical mechanics and information theory is the Boltzmann (Shannon) entropy regularization functional, in $\mathcal{U} = L^1(\Omega)$,

$$J(u) = \int_\Omega u \log(u) - u \ dx,$$

leading to the so-called maximum entropy regularization (cf. [34]). Recently popular functionals are non-differentiable regularization energies like the one-norm $J(u) = \|u\|_{\ell^1}$ in $\mathcal{U} = \mathbb{R}^N$ or the total variation $J(u) = \mathrm{TV}(u)$, being defined as

$$(2.3) \qquad \mathrm{TV}(u) := \sup_{\substack{\varphi \in C_0^\infty(\Omega;\mathbb{R}^n) \\ \|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)} \leq 1}} \int_\Omega u \ \mathrm{div}\varphi \ dx \,.$$

Total variation regularization became popular in the Rudin-Osher-Fatemi (ROF) model [59]

$$(2.4) \qquad \hat{u} \in \operatorname*{arg\,min}_{u \in \mathrm{BV}(\Omega)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha \mathrm{TV}(u) \right\} \,.$$

The space $\mathrm{BV}(\Omega)$ is the space of all function $u \in L^1(\Omega)$ such that $\mathrm{TV}(u)$ is bounded. In case of $\Omega \subseteq \mathbb{R}^n$ for $n \in \{1, 2\}$ the space $\mathrm{BV}(\Omega)$ can be embedded into $L^2(\Omega)$.

Since we are going to deal with rather large classes of convex functionals J, let us recall some basic facts from convex analysis (cf. [58, 35] for detailed discussions). As usual for a Banach space $\mathcal{U}$, the Banach space of bounded linear mappings from $\mathcal{U}$ to $\mathbb{R}$ is called the dual space of $\mathcal{U}$ and is denoted by $\mathcal{U}^*$, with norm

$$\|p\|_{\mathcal{U}^*} := \sup_{\|u\|_{\mathcal{U}}=1} |p(u)| = \sup_{u \in \mathcal{U} \setminus \{0\}} \frac{|p(u)|}{\|u\|_{\mathcal{U}}} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |p(u)| \,.$$

The functional $p(u) = \langle p, u \rangle_{\mathcal{U}^* \times \mathcal{U}}$ is called the dual product. Throughout this work we are going to denote the dual product simply by $\langle p, u \rangle_{\mathcal{U}}$. In case that $\mathcal{U}$ is even a Hilbert space, the dual product can be identified with the scalar product of $\mathcal{U}$.

The characterization of dual spaces and its elements allows us to define the sub-differential of a convex functional.

DEFINITION 1 (Subdifferential). *Let $\mathcal{U}$ be a Banach space with dual space $\mathcal{U}^*$, and let the proper functional $J : \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$ be convex. Then, J is called subdifferentiable at $u \in \mathcal{U}$, if there exists an element $p \in \mathcal{U}^*$ such that*

$$J(v) - J(u) - \langle p, v - u \rangle_{\mathcal{U}} \geq 0$$

holds, for all $v \in \mathcal{U}$. Furthermore, we call p a subgradient at position u. The collection of all subgradients at position u, i.e.

$$\partial J(u) := \{ p \in \mathcal{U}^* \mid J(v) - J(u) - \langle p, v - u \rangle_{\mathcal{U}} \geq 0 \, , \forall v \in \mathcal{U} \} \subset \mathcal{U}^* \, ,$$

is called subdifferential of J at u.

We further mention that the subdifferential of one-homogeneous functionals can be further characterized as (cf. [73])

(2.5) $$\partial J(u) := \{ p \in \mathcal{U}^* \mid \langle p, u \rangle = J(u), \ \langle p, v \rangle \leq J(v) \, , \forall v \in \mathcal{U} \} \subset \mathcal{U}^* \, .$$

Another concept we shall use in several arguments is the notion of (generalized) Bregman distances, defined as

$$D_J^p(v, u) = J(v) - J(u) - \langle p, v - u \rangle_{\mathcal{U}} \, ,$$

with $p \in \partial J(u)$. Bregman distances are not common distance functionals, since they do not satisfy a triangle inequality and are not symmetric in general. However, for J being convex they are non-negative and satisfy $D_J^p(u, u) = 0$. Symmetry can be restored by using symmetric Bregman distances, i.e.

$$D_J^{p,q,\mathrm{symm}}(v, u) = D_J^p(u, v) + D_J^q(v, u) = \langle q - p, v - u \rangle_{\mathcal{U}}$$

for $q \in \partial J(v)$.

Before we continue with the definition of ground states and singular vectors for general convex and subdifferentiable regularization functionals, we want to precisely define the class of operators and functionals we are going to investigate. Thus, for the remainder of this work we will assume the following properties without further notice:

ASSUMPTION 1 (Setup).
- *$\Omega \subseteq \mathbb{R}^d$, $\Sigma \subseteq \mathbb{R}^k$ are bounded domains.*
- *$\mathcal{U}$ is a Banach space, being the dual of some other Banach space.*
- *$\mathcal{H}$ is a Hilbert space.*
- *$K : \mathcal{U} \to \mathcal{H}$ is a bounded linear operator mapping between these spaces*
- *$J : \mathrm{dom}(J) \subseteq \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$ is a proper non-negative convex functional*

3. Ground states and singular vectors. In this section we want to define ground states of regularization functionals as well as an analogue of singular vectors for nonlinear functionals.

3.1. Ground states. We start with a definition of a ground state, which is motivated by similar properties in partial differential equations, e.g. the classical one of a ground-state in the Schrödinger equation and related problems (cf. [12, 68, 2]). In order to obtain a ground state we normalize the element u and minimize the regularization functional among those elements, i.e. u_0 is defined as (see Theorem 1 below for well-definedness in an appropriate setting)

$$(3.1) \qquad u_0 \in \underset{\substack{u \in \mathrm{dom}(J) \\ \|Ku\|_{\mathcal{H}} = 1}}{\arg\min} \{J(u)\} .$$

In the context of variational schemes like (2.2) we are particularly interested in non-trivial ground states of one-homogeneous regularization functional. A trivial ground state appears if $J(u_0) = 0$, and we can immediately provide a well-known example for such:

EXAMPLE 1. Let $\Omega \subset \mathbb{R}^d$ with $d \in \{1, 2\}$. Then we know that $BV(\Omega) \subset L^2(\Omega)$ holds. Thus, for $K = I$ being the identity operator $I : L^2(\Omega) \to L^2(\Omega)$, a trivial ground state of $J = \mathrm{TV}$ is the constant function $u_0 = 1/\sqrt{|\Omega|}$, since we have $\mathrm{TV}(u_0) = 0$ and $\|u_0\|_{L^2(\Omega)} = 1$. Note that as usual the ground state is not unique, since $-u_0$ is a ground state as well.

However, in many cases trivial ground states do not give interesting insights into the nature of a regularization energy, as the previous example shows. Thus, we would like to investigate non-trivial ground states that are orthogonal to the trivial ones in a reasonable sense. Let us therefore define some preliminary notions first. The K-product of two elements $u, v \in \mathcal{U}$ is defined as

$$\langle u, v \rangle_K := \langle Ku, Kv \rangle_{\mathcal{H}} .$$

Furthermore we are going to write $\|u\|_K$ as an abbreviation for $\sqrt{\langle u, u \rangle_K}$. This particular definition of a scalar product for elements of a Banach space allows us to define a useful orthogonal complement of a kernel of a regularization functional, which we define as usual via

$$(3.2) \qquad \ker(J)^{\perp} := \{u \in \mathrm{dom}(J) \mid \langle u, v \rangle_K = 0, \forall v \in \ker(J)\} .$$

For completeness we also introduce

$$\ker(J) := \{u \in \mathrm{dom}(J) \mid J(u) = 0\} .$$

In the case of convex non-negative one-homogeneous functionals we are mainly interested in, the kernel and also its complement can further be characterized as linear subspaces:

LEMMA 1. *Let J be convex, non-negative and one-homogeneous. Then $\ker(J)$ is a linear subspace.*

Proof. Let $u, v \in \ker(J)$ and $a, b \in \mathbb{R}$ such that $|a| + |b| \neq 0$. With $\alpha = \frac{|a|}{|a|+|b|} \in [0, 1]$, $\tilde{u} = \mathrm{sign}(a)u$ and $\tilde{v} = \mathrm{sign}(b)v$ we obtain

$$0 \leq J(au + bv) = J((|a| + |b|)(\alpha\tilde{u} + (1 - \alpha)\tilde{v}))$$
$$= (|a| + |b|)J(\alpha\tilde{u} + (1 - \alpha)\tilde{v}) \leq (|a| + |b|)(\alpha J(\tilde{u}) + (1 - \alpha)J(\tilde{v})),$$

by using the one-homogeneity and convexity of J. Since $J(\tilde{u}) = J(u) = 0$ and $J(\tilde{v}) = J(v) = 0$ hold due to the one-homogeneity of J, we conclude $J(au + bv) = 0$ as well. Thus, $au + bv \in \ker(J)$ holds true. $\square$

EXAMPLE 2. Considering $J = \text{TV}$ again, we easily see that $\ker(\text{TV})$ equals the set of all constant functions, because the estimation of the kernel can simply be reduced to estimating $\ker(\nabla)$.
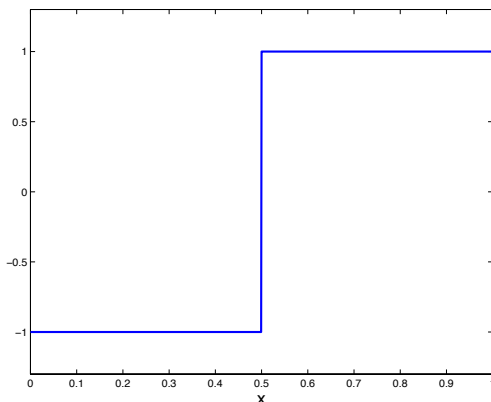

FIG. 1. *The function u^a as defined in (3.5), for $a = 1/2$. This function is a ground state of $K = I$, $J = \text{TV}$, according to Definition 2.*

DEFINITION 2 (Ground State). *Under the above assumptions on J and K, a ground state u_0 is defined as an element*

$$(3.3) \qquad u_0 \in \underset{\substack{u \in \ker(J)^\perp \\ \|Ku\|_{\mathcal{H}} = 1}}{\arg\min} \{J(u)\} .$$

Moreover, if u_0 exists we call

$$(3.4) \qquad \lambda_0 = J(u_0)$$

the smallest singular value.

Under standard assumptions on variational regularization methods, the ground state indeed exists, which can be verified by usual arguments (cf. [74]):

THEOREM 1. *Let d be a metric on $\mathcal{U}$, let K be continuous from this metric topology to the strong topology of $\mathcal{H}$, and let J be lower semicontinuous with respect to this metric topology. Moreover, let the sublevel sets of $u \mapsto \|Ku\|_{\mathcal{H}} + J(u)$ be compact in the metric topology. Then there exists at least one ground state $u_0 \in \mathcal{U}$.*

We want to give a brief example of ground states in case of $J(u) = \text{TV}(u)$.

EXAMPLE 3. For $\Omega = [0,1] \subset \mathbb{R}$, $K = I$ and $J(u) = \text{TV}(u)$ we want to consider

$$(3.5) \qquad u^a(x) = \begin{cases} \frac{\sqrt{(1-a)a}}{1-a} & x \geq a \\ -\frac{\sqrt{(1-a)a}}{a} & x < a \end{cases} ,$$

for $a \in (0, 1)$. We easily see that u^a is orthogonal to the kernel of TV, which consists of all constant functions due to Example 2, since $\int_0^1 u^a(x) \, dx = 0$ holds. Moreover, u^a guarantees the normalization constraint $\|u^a\|_{L^2([0,1])} = 1$ for every $a \in]0, 1[$. However, the TV-value $\mathrm{TV}(u^a) = \frac{1}{\sqrt{(1-a)a}}$ is a strictly convex function in a with unique minimum at $a = 1/2$. Thus, $u_0 = u^{\frac{1}{2}}$, which is visualized in Figure 1, is a ground state if we can prove that there does not exist a function $\tilde{u}$ with $\langle \tilde{u}, 1 \rangle_{L^2([0,1])} = 0$, $\|\tilde{u}\|_{L^2([0,1])} = 1$ and $\mathrm{TV}(\tilde{u}) < 2$.

LEMMA 2.   *There exists no function $\tilde{u} \in C([0,1])$ with $\|\tilde{u}\|_{L^2([0,1])} = 1$ and $\langle 1, \tilde{u} \rangle_{L^2([0,1])} = 0$ such that $\mathrm{TV}(\tilde{u}) < 2$ holds.*

Proof. It is easy to see that for the monotonic rearrangement of an arbitrary function $\tilde{u} \in C([0,1])$, which we want to denote by

$$\tilde{u}^*(x) = \sup\{t \mid |\{u > t\}| > |x|\},$$

we have (cf. e.g [2])

$$\mathrm{TV}(\tilde{u}^*) \leq \mathrm{TV}(\tilde{u}).$$

Thus, in the following we are going to consider monotonically increasing functions $\tilde{u}^* \in C([0,1])$ with $\|\tilde{u}^*\|_{L^2([0,1])} = 1$ and $\langle 1, \tilde{u}^* \rangle_{L^2([0,1])} = 0$ only, without loss of generality. Now we want to prove $\mathrm{TV}(\tilde{u}^*) \geq 2$ by contradiction and therefore subdivide the proof into two parts. First of all we are going to prove the inequalities

$$(3.6) \qquad (\tilde{u}^*(1))^2 - 1 \geq \left((\tilde{u}^*(1))^2 - (\tilde{u}^*(0))^2\right) y$$

and

$$(3.7) \qquad 1 \leq \mathrm{TV}(\tilde{u}^*) \left(-\int_0^y \tilde{u}^*(x)\right),$$

with y denoting a root of $\tilde{u}^*$. Subsequently we are going to use (3.6) and (3.7) to conclude the contradiction.

For a monotonically increasing function $\tilde{u}^*$ we can rewrite the normalization constraint $\|\tilde{u}^*\|_{L^2([0,1])} = 1$ to

$$
\begin{aligned}
1 = \|\tilde{u}^*\|_{L^2([0,1])}^2 &= \int_0^y (\tilde{u}^*(x))^2 \, dx + \int_y^1 (\tilde{u}^*(x))^2 \, dx \\
&\leq y \, (\tilde{u}^*(0))^2 + (1-y) \, (\tilde{u}^*(1))^2 ,
\end{aligned}
$$

with y denoting a root of $\tilde{u}^*$. Rearranging immediately yields (3.6).

Moreover, according to the second mean value theorem of integration there exists a $\xi \in]0, 1[$ such that we can rewrite the normalization constraint to

$$
\begin{aligned}
1 &= \int_0^1 \tilde{u}^*(x)\tilde{u}^*(x) \, dx \\
&= \tilde{u}^*(0) \int_0^\xi \tilde{u}^*(x) \, dx + \tilde{u}^*(1) \int_\xi^1 \tilde{u}^*(x) \, dx \\
&= \underbrace{(\tilde{u}^*(1) - \tilde{u}^*(0))}_{=\mathrm{TV}(\tilde{u}^*)} \int_\xi^1 \tilde{u}^*(x) \, dx ,
\end{aligned}
$$

where the last equality holds due to $\langle 1, \tilde{u}^* \rangle_{L^2([0,1])} = 0$. Since the value of $\int_\xi^1 \tilde{u}^*(x) \, dx$ gets maximal for $\xi = y$ and since we know $\int_y^1 \tilde{u}^*(x) \, dx = -\int_0^y \tilde{u}^*(x) \, dx$ we obtain (3.7).

Finally, we are now able to proof the lemma's statement via contradiction. We assume $\tilde{u}^*$ to satisfy $\mathrm{TV}(\tilde{u}^*) < 2$. Then, the normalization constraint $\|\tilde{u}^*\|_{L^2([0,1])} = 1$ however implies either $\tilde{u}^*(1) \geq 1$ or $\tilde{u}^*(0) \leq -1$. Without loss of generality we assume $\tilde{u}^*(1) \geq 1$. Thus, there exists a constant $c \geq 0$ such that $\tilde{u}^*(1) = 1 + c$ holds. Due to $\mathrm{TV}(\tilde{u}^*) < 2$ this automatically implies $\tilde{u}^*(0) \geq -1 + c$ and $(\tilde{u}^*(0))^2 \leq (-1 + c)^2$. Applying (3.6) therefore yields

$$(1+c)^2 - 1 \geq \left((1+c)^2 - (1-c)^2 \right) y \,,$$

which can be rewritten to $y \leq c/4 + \frac{1}{2}$. We are therefore able to estimate (3.7) to obtain

$$\begin{aligned}
1 &\leq \mathrm{TV}(\tilde{u}^*) \left(-\int_0^y \tilde{u}^*(x) \, dx \right) \\
&\leq \mathrm{TV}(\tilde{u}^*) y \left(-\tilde{u}^*(0) \right) \leq \mathrm{TV}(\tilde{u}^*) y (1-c) \\
&\leq \mathrm{TV}(\tilde{u}^*) \left(\frac{c}{4} + \frac{1}{2} \right) (1-c) \\
&\leq \mathrm{TV}(\tilde{u}^*) \left(\frac{1}{2} - \frac{c^2}{4} - \frac{c}{4} \right) \leq \frac{1}{2} \mathrm{TV}(\tilde{u}^*) \,,
\end{aligned}$$

which yields $2 \leq \mathrm{TV}(\tilde{u}^*)$ and therefore is a contradiction to the assumption $\mathrm{TV}(\tilde{u}^*) < 2$. $\square$

Note that u_0 of the previous example is not a unique ground state, since $-u_0$ yields the same minimum. However, this is also true for ground states of quadratic variational schemes and therefore no surprise. Nevertheless, the following example shows that in the general case there exist ground states that differ substantially, and not just only in terms of their sign.

EXAMPLE 4. Let $\mathcal{U} = \ell^1(\mathbb{R}^N)$ and $\mathcal{H} = \ell^2(\mathbb{R}^M)$. Consider the regularization energy $J(u) = \|u\|_{\ell^1}$, and as an operator a matrix K with columns being normalized with respect to the ℓ^2-norm, i.e. $(Ke_j)^T \cdot (Ke_j) = 1$, with e_j denoting the j-th unit vector. Then every vector $e_i = (\delta_{ji})_{j=1,\dots,N}$, with δ_{ij} denoting the Kronecker delta, is a ground state.

In some cases it is useful to have an alternative definition of ground states of variational regularizations like (2.2):

DEFINITION 3 (Ground State II). *A ground state u_0 is defined as*

$$(3.8) \qquad u_0 \in \arg\max_{\substack{u \in \ker(J)^\perp \\ J(u) \leq 1}} \{ \|Ku\|_{\mathcal{H}} \} \,.$$

If u_0 exists, we shall call

$$(3.9) \qquad \lambda_0 = \frac{1}{\|Ku_0\|_{\mathcal{H}}}$$

the smallest singular value.

At first glance, Definition 3 appears to be very different from Definition 2; however, for one-homogeneous functionals J both definitions are equivalent up to normalization of the singular vector and yield the same singular value, as we will see with the following Lemma.

LEMMA 1. *Let J be a proper and one-homogeneous functional. Then, Definition 2 and Definition 3 are equivalent up to multiplication of u_0 by λ_0.*

Proof. $\Rightarrow$: Let u_0 be a ground state that satisfies Definition 2. Then, for $\tilde{u} := u_0/J(u_0) = u_0/\lambda_0$ we obtain $J(\tilde{u}) = 1$ and $\|K\tilde{u}\|_{\mathcal{H}} = 1/J(u_0) = 1/\lambda_0$. Thus, in order to satisfy Definition 3, $\tilde{u}$ is supposed to maximize $\|Ku\|_{\mathcal{H}}$, i.e. $\|K\tilde{u}\|_{\mathcal{H}} \geq \|Kv\|_{\mathcal{H}}$ for all v with $J(v) \leq 1$. We prove this statement by contradiction and assume that there exists a function v with $\|Kv\|_{\mathcal{H}} > 1/\lambda_0$ and $J(v) \leq 1$. However, if such a function exists, we can define $\tilde{v} := v/\|Kv\|_{\mathcal{H}}$ to obtain a function that satisfies $\|K\tilde{v}\|_{\mathcal{H}} = 1$ and $J(\tilde{v}) \leq 1/\|Kv\|_{\mathcal{H}} < \lambda_0$, which is a contradiction to u_0 being a ground state in the sense of Definition 2.

$\Leftarrow$: Now let u_0 be a ground state in terms of Definition 3, i.e. $J(u_0) \leq 1$ such that $\|Ku_0\|_{\mathcal{H}} =: 1/\lambda_0$ is maximized. Then we can define $\tilde{u} = \lambda_0 u_0$, which satisfies $\|K\tilde{u}\|_{\mathcal{H}} = 1$ and $J(\tilde{u}) = \lambda_0$. In analogy to the first part of the proof, we prove by contradiction that $\tilde{u}$ already has to be a ground state in terms of Definition 2. We therefore assume that there exists a function v such that $J(v) < \lambda_0$ and $\|Kv\|_{\mathcal{H}} = 1$ holds. If such a function exists, than $\tilde{v} := v/J(v)$ exists as well. However, for $\tilde{v}$ we observe $J(\tilde{v}) = 1$ and $\|K\tilde{v}\|_{\mathcal{H}} > 1/\lambda_0$, which is a contradiction to u_0 being a ground state in terms of Definition 3. $\square$

3.2. Singular vectors. In analogy to singular vectors of linear operators we want to extend the concept of singular vectors to variational frameworks of the form (2.2). The motivation is to consider the formal optimality condition for the ground state, which is obtained by considering stationary points of the Lagrange functional (with parameter $\lambda \in \mathbb{R}$)

$$(3.10) \qquad L(u; \lambda) = J(u) - \frac{\lambda}{2}(\|Ku\|_{\mathcal{H}}^2 - 1),$$

which is given by

$$\lambda K^* K u \in \partial J(u),$$

where ∂J denotes the subdifferential.
Hence, we define a singular vector as follows:

DEFINITION 4 (Singular Vector). *Let J be convex with non-empty subdifferential ∂J at every $u \in \mathrm{dom}(J)$. Then, every function u_λ with $\|Ku_\lambda\|_{\mathcal{H}} = 1$ satisfying*

$$(3.11) \qquad \lambda K^* K u_\lambda \in \partial J(u_\lambda)$$

is called singular vector of J with corresponding singular value λ. The subgradient $p_\lambda = \lambda K^* K u_\lambda$ is called dual singular vector in the following.

We mention that taking a dual product with u_λ yields the singular value relation

$$\lambda = \frac{\langle p_\lambda, u_\lambda \rangle_{\mathcal{U}}}{\|Ku_\lambda\|_{\mathcal{H}}^2} = \langle p_\lambda, u_\lambda \rangle_{\mathcal{U}}.$$

In the case of J being one-homogeneous we even have

$$\lambda = J(u_\lambda)\,,$$

which also implies $\lambda \geq \lambda_0$, for any singular value λ.

For smooth J one can prove that the ground state is a singular vector by analyzing the Lagrange functional above (cf. [42]). In the one-homogeneous case we give an alternative proof:

PROPOSITION 1. *Let J be one-homogeneous and let u_0 be the ground state with $\lambda_0 = J(u_0)$. Then λ_0 is a singular value and u_0 is a singular vector.*

Proof. For J being one-homogeneous, $p_0 \in \partial J(u_0)$ is equivalent to $\langle p_0, u_0\rangle_\mathcal{U} = J(u_0)$ and

$$\langle p_0, u\rangle_\mathcal{U} \leq J(u), \qquad \forall\, u \in \mathcal{U}\,,$$

due to (2.5). We verify this property for $p_0 = \lambda_0 K^* K u_0$. First of all we obtain

$$\langle p_0, u_0\rangle_\mathcal{U} = \lambda_0 \langle K u_0, K u_0\rangle_\mathcal{H} = J(u_0)$$

by the definition of λ_0 and the normalization of u_0. Moreover, for arbitrary $u \in \mathcal{U}$ with $Ku \neq 0$ we define $v = u/\|Ku\|_\mathcal{H}$ and find

$$\langle p_0, u\rangle_\mathcal{U} = \|Ku\|_\mathcal{H} \lambda_0 \langle K u_0, Kv\rangle_\mathcal{H} \leq \|Ku\|_\mathcal{H} \lambda_0.$$

Since v is normalized, we have by the definition of the ground state

$$\lambda_0 = J(u_0) \leq J(v) = \frac{J(u)}{\|Ku\|_\mathcal{H}}$$

and thus, $\langle p_0, u\rangle_\mathcal{U} \leq J(u)$. If $Ku = 0$, then

$$\langle p_0, u\rangle_\mathcal{U} = \lambda_0 \langle K u_0, K u\rangle_\mathcal{H} = 0 \leq J(u),$$

thus $p_0 \in \partial J(u_0)$. □

Higher singular values and singular vectors are difficult to characterize as we shall see from examples below. Also orthogonality of singular vectors corresponding to different singular values is lost. Consider $\lambda K^* K u_\lambda = p_\lambda$ and $\mu K^* K u_\mu = p_\mu$ for singular vectors u_λ and u_μ, then we only have

$$\frac{1}{\lambda}\langle p_\lambda, u_\mu\rangle = \frac{1}{\mu}\langle p_\mu, u_\lambda\rangle\,.$$

We want to give two examples:

EXAMPLE 5. Let us investigate the functional $J(u) = \frac{1}{2}\|\nabla u\|_{L^2(\Omega;\mathbb{R}^n)}^2$, in order to demonstrate that the definition of singular vectors is consistent with the definition of singular vectors of linear operators. Since J is Fréchet-differentiable, its subdifferential consists of its Fréchet-derivative only, i.e. $\partial J(u) = \{-\Delta u\}$. Considering $K = I$ for simplicity, (3.11) reads as the classical Eigenfunction problem of the Laplace operator, i.e.
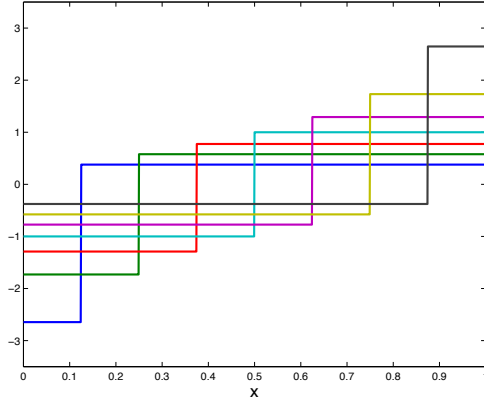
$$-\lambda u_\lambda = \Delta u_\lambda\,,$$

FIG. 2. *The functions u^a plotted for the values $a \in \{1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8\}$. All of these functions are singular values according to Definition 4, but only $u^{1/2}$ is a ground state.*

or equivalently as the singular vector decomposition problem of the gradient operator ∇. Due to the compactness of the inverse Laplacian, the spectrum is discrete, i.e. there are countably many different singular values.

The structure of the spectrum, by which we formally denote the set of singular values, changes if we consider more degenerate cases as the total variation frequently used in inverse problems and imaging:

EXAMPLE 6. Let us now consider K being the embedding operator from $\mathrm{BV}(\Omega)$ to $L^2(\Omega)$ on the unit interval $\Omega = [0, 1]$. For $a \in (0, 1)$ the function u^a defined by (3.5) is a singular vector of TV, with singular value $\lambda = 1/\sqrt{(1-a)a}$, as we shall see in the following: We can characterize the subdifferential of TV as

(3.12)
$$\partial \mathrm{TV}(u) = \left\{ \mathrm{div}\varphi \ \Big| \ \|\varphi\|_{L^\infty(\Omega;\mathbb{R}^n)} \le 1, \ \varphi \cdot n|_{\partial\Omega} = 0, \langle \mathrm{div}\varphi, u\rangle_{L^2(\Omega)} = \mathrm{TV}(u) \right\},$$

for $u \in \mathrm{BV}(\Omega)$. We see that the distributional derivative of the continuous function $q^a : [0, 1] \to [-1, 0]$ defined as

$$q^a(x) := \begin{cases} \frac{x-1}{1-a} & x \ge a \\ -\frac{x}{a} & x < a \end{cases},$$

for $a \in]0, 1[$, is an element of $\partial \mathrm{TV}(u^a)$, since we have $\|q^a\|_{L^\infty([0,1])} = 1$, $q^a(0) = q^a(1) = 0$ and $\langle (q^a)', u^a\rangle_{L^2([0,1])} = \mathrm{TV}(u^a) = \lambda$ (here the derivative has to be considered in a distributional way). Moreover, the distributional derivative of q^a satisfies the singular vector relation $\lambda u^a = (q^a)' \in \partial \mathrm{TV}(u^a)$, for $\lambda = 1/\sqrt{(1-a)a}$. Hence, u^a is a singular vector of TV. The singular vector is visualized for various choices of a in Figure 2 Now we see that there exists a continuous spectrum $[2, \infty)$ although in spatial dimension one the embedding operator is compact.

We are going to consider various further examples in Section 8.

3.3. Failure of the Rayleigh principle. Above we have defined the ground state by a Rayleigh principle, i.e. minimizing J with respect to normalization and orthogonality to the kernel. Again in the linear case, further singular vectors can be obtained by similar Rayleigh principles, e.g. by minimizing $J(u)$ subject to $u \in \ker(J)^\perp$, $\|u\|_K = 1$ and $\langle u, u_0 \rangle_K = 0$. We have already seen from Example 6 that in the nonlinear case we will not be able to compute all singular values and singular vectors this way. However, it seems at least interesting whether we can obtain an orthonormal basis of singular vectors (with orthogonality defined in the K-scalar product, assuming for the moment that K has trivial nullspace). As we shall demonstrate in the following, this is not possible in general.

Let us first discuss formally why the Rayleigh principle can fail to yield singular vectors in the nonlinear case. Assume for this sake we have a system of orthonormal singular vectors u_j, $j = 0, \ldots, n$, i.e.

$$\lambda_j K^* K u_{\lambda_j} = p_{\lambda_j} \in \partial J(u_{\lambda_j}), \qquad \langle u_{\lambda_j}, u_{\lambda_k} \rangle_K = \delta_{jk}.$$

Then we define

$$(3.13) \qquad u_{\lambda_{n+1}} \in \arg\min_{u \in \mathcal{C}_n} J(u),$$

with the constraint set

$$(3.14) \qquad \mathcal{C}_n := \left\{ u \in \mathcal{U} \;\middle|\; \|u\|_K = 1, \langle u_{\lambda_j}, u \rangle_K = 0, j = 0, \ldots, n \right\}.$$

Note that the existence of $u_{\lambda_{n+1}}$ follows under the same conditions as the existence of a ground state if in addition $\mathcal{C}_n$ is nonempty (which is always the case in infinite dimensions). Again we may set up the Lagrange functional

$$(3.15) \qquad L = J(u) - \frac{\lambda}{2} \left(\|u\|_K^2 - 1 \right) - \sum_{j=0}^{n} \mu_j \langle u_{\lambda_j}, u \rangle_K$$

and thus obtain the optimality condition as the variation with respect to u via

$$(3.16) \qquad \lambda K^* K u_{\lambda_{n+1}} + \sum_{j=0}^{n} \mu_j K^* K u_{\lambda_j} = p_{\lambda_{n+1}} \in \partial J(u_{\lambda_{n+1}}).$$

Now $u_{\lambda_{n+1}}$ is a singular vector if and only if $\mu_j = 0$ for $j = 1, \ldots, n$. To understand the latter we take a product with u_{λ_k}, $k \in \{0, \ldots, n\}$, and obtain

$$\mu_k = \langle p_{\lambda_{n+1}}, u_{\lambda_k} \rangle_{\mathcal{U}},$$

due to the orthonormality. Now there is no particular reason why $\langle p_{\lambda_{n+1}}, u_{\lambda_k} \rangle_{\mathcal{U}} = 0$ should hold, since we cannot use the usual argument as in the linear case, namely a simple eigenvalue equation for u_{λ_k}.

We finish this section with a simple example that explicitly shows the failure of the Rayleigh principle:

EXAMPLE 7. Let $\mathcal{U} = \ell^1(\mathbb{R}^2)$ and $\mathcal{H} = \ell^2(\mathbb{R}^2)$, $J(u) = \|u\|_1$, and choose K such that

$$(3.17) \qquad K^* K = \begin{pmatrix} 1 & 2\epsilon \\ 2\epsilon & \epsilon \end{pmatrix}$$

for $0 < \epsilon < \frac{1}{4}$. Then it is straight-forward to see that the ground state is given by $u_0 = \pm(1, 0)^T$ with $\lambda_0 = 1$ and $\|Ku_0\|_2 = 1$. All vectors of the form $u_1 = c(-2\epsilon, 1)$, $c \in \mathbb{R}$, are orthogonal to u_0. Without restriction of generality we can use the one with $c = 1$ and verify that it is not a singular vector. We find

$$\lambda_1 K^* K u_1 = \lambda_1 \begin{pmatrix} 0 \\ \epsilon - 4\epsilon^2 \end{pmatrix} \notin \partial \|u_1\|_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Though the second entry is positive due to $0 < \epsilon < \frac{1}{4}$, the first entry remains zero for any choice of λ_1, thus implying that u_1 is not a singular vector.

From the last example we see that it is not possible to construct an orthonormal basis of singular vectors starting from the ground state in general. However, in certain cases this may still be possible, as we shall see also in a surprising example in the next section.

4. Scales and scale estimates. In the linear singular value decomposition, i.e. $\lambda K^* K u = \frac{u}{\|u\|_K}$ in our setup, the singular vectors and singular values carry information about scale. The usual definition of scale respectively frequency is related to the value of λ. In analogy to the eigenvalues of the Laplace operator mentioned above, small λ means small frequency respectively large scale, and the scale information is carried by the singular vectors. For increasing λ the frequency is increasing, respectively the scale is decreasing. It turns out that the interpretation of scale is even more striking in some nonlinear regularization cases, in particular for the ROF denoising model discussed in Example 6. Here, intuitively the ground state is the largest scale, since it includes two plateaus of size $\frac{1}{2}$. For other values of a we have $\lambda = \frac{1}{a(1-a)}$, i.e. decreasing scale the farther a is away from $\frac{1}{2}$. This is quite intuitive as well, since the singular value has plateaus of length a or $1 - a$ and the size of the smaller plateau is decreasing in scale as λ is increasing. As we shall see in the following, we can use an appropriate selection of singular vectors of the ROF problem to obtain a standard multiscale representation.

4.1. Scales in total variation and the Haar wavelet. The probably most frequently studied and best understood multiscale decomposition is the one of signals in the Haar wavelet basis, given on the unit interval by $\psi_{0,0}(x) = \chi(x)$ and
(4.1)
$$\psi_{j,k}(x) = 2^{(j-1)/2}(\chi(2^j x - 2k) - \chi(2^j x - 2k + 1)), \qquad k = 0, \ldots, 2^{j-1} - 1, \; j = 1, 2, \ldots$$

with the scale function χ being the characteristic function of the unit interval, i.e.

$$(4.2) \qquad \chi(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{else} \end{cases}.$$

It is a kind of folklore that the Haar wavelet decomposition is closely related to total variation methods in spatial dimension one (respectively to anisotropic total variation in higher dimension), and rough connections between ROF denoising and filtering with Haar wavelets have been established (cf. [24, 44, 63]). Here we shall provide a more explicit connection between the Haar wavelet basis and singular vectors of the ROF functional. For this sake we use a slightly nonstandard definition of the total variation in the form

$$(4.3) \qquad \mathrm{TV}_*(u) := \sup_{\substack{\varphi \in C^\infty(\Omega; \mathbb{R}^n) \\ \|\varphi\|_{L^\infty(\Omega; \mathbb{R}^n)} \le 1}} \int_\Omega u \, \mathrm{div}\varphi \, dx.$$

Note that in contrast to (2.3) we do not choose test functions φ with compact support, which yields an additional boundary term. For functions $u \in W^{1,1}([0,1]) \cap C([0,1])$ it is straight-forward to see that

$$(4.4) \qquad \mathrm{TV}_*(u) = \int_0^1 |u'(x)| \ dx + |u(1)| + |u(0)|.$$

This in particular eliminates the kernel of the total variation, so that the ground state is indeed a constant $u_0 \equiv 1$ with $\lambda_0 = \mathrm{TV}_*(u_0) = 2$. Then our main result is the following:

THEOREM 2. *Let $K : BV([0,1]) \to L^2([0,1])$ be the embedding operator and let $J = TV_*$. Then the Haar wavelet basis is an orthonormal set of singular vectors for K and J, i.e.*

$$(4.5) \qquad \lambda_{j,k} \psi_{j,k} \in \partial TV_*(\psi_{j,k})$$

with singular value $\lambda_{j,k} = 2^{(j+3)/2}$ for $j \geq 1$ and $\lambda_0 = 2$. In particular u_0 is a ground state.

Proof. We first show that $u_0 = \psi_{0,0}$ is a ground state. Take an arbitrary continuously differentiable function u satisfying the normalization condition $\int_0^1 u^2 \ dx = 1$. Then there exists $x_0 \in [0,1]$ such that $|u(x_0)| \geq 1$. Thus, by the triangle inequality

$$\mathrm{TV}_*(u) \geq |u(1) - u(x_0)| + |u(x_0) - u(0)| + |u(1)| + |u(0)| \geq 2|u(x_0)| \geq 2$$

holds. Since $C^1([0,1])$ is dense in $BV([0,1])$ we conclude that the infimum over TV_* over all functions of bounded variation with normalized L^2-norm is 2. Since $\mathrm{TV}_*(u_0) = 2$ and $\|u_0\|_{L^2([0,1])} = 1$, it is a ground state.

We further need to verify that $p_{j,k} = \lambda_{j,k} \psi_{j,k}$ is a subgradient of the one-homogeneous functional TV_* at $\psi_{j,k}$, for $j \geq 1$. For this sake we first compute $\mathrm{TV}_*(\psi_{j,k}) = 2^{(j-1)/2} 4 = 2^{(j+3)/2}$ and immediately see that

$$\langle p_{j,k}, \psi_{j,k} \rangle = \lambda_{j,k} \int_0^1 |\psi_{j,k}|^2 \ dx = 2^{(j+3)/2}.$$

The remaining step is to prove that

$$\langle p_{j,k}, u \rangle \leq \mathrm{TV}_*(u)$$

for arbitrary $u \in BV([0,1])$. For this sake we consider the primitive q satisfying $q' = p_{j,k}$ and $q(0) = -1$. We observe that indeed q attains its maxima and minima on the jump set of $p_{j,k}$ and they equal $+1$ respectively -1. Hence, $\|q\|_\infty = 1$ and we find

$$\langle p_{j,k}, u \rangle = \int_0^1 q'(x) u(x) \ dx \leq \sup_{\substack{\varphi \in W^{1,2}([0,1];\mathbb{R}) \\ \|\varphi\|_{L^\infty([0,1])} \leq 1}} \int_\Omega u \ \varphi' \ dx.$$

Finally, by a density argument we conclude that this supremum equals the one over C^∞, hence

$$\langle p_{j,k}, u \rangle \leq \sup_{\substack{\varphi \in C^\infty([0,1];\mathbb{R}) \\ \|\varphi\|_{L^\infty([0,1])} \leq 1}} \int_\Omega u \ \varphi' \ dx = \mathrm{TV}_*(u).$$

$\square$

4.2. Scale estimates. In the following we want to demonstrate how singular vectors can be used to derive sharp error estimates on the components at different scales without any additional prior knowledge on the solution, as e.g. the assumption of specific source conditions (cf. [22, 43]). We will make the connection to scale explicit by again considering the case of the ROF-model and the corresponding singular vectors.

Let us assume we are given a singular vector u_λ with singular value λ and dual singular vector p_λ, i.e.

$$\lambda K^* K u_\lambda = p_\lambda \in \partial J(u_\lambda)$$

for $\|Ku\|_{\mathcal{H}} = 1$. Then we can estimate solutions of (2.2) for input data given in terms of $f = K\tilde{u} + \eta$, with $\tilde{u} \in \text{dom}(K) \cap \text{dom}(J)$ and $\eta \in \mathcal{H}$, with respect to this particular singular vector.

THEOREM 3. *For input data $f = K\tilde{u} + \eta$, $\tilde{u} \in \text{dom}(K)$ and $\eta \in \mathcal{H}$, the solution $\hat{u}$ of (2.2) satisfies the estimate*

$$\left| \left\langle \frac{1}{\lambda} p_\lambda, \hat{u} - \tilde{u} \right\rangle_{\mathcal{U}} \right| \leq |\langle \eta, K u_\lambda \rangle_{\mathcal{H}}| + \alpha \, |\langle \hat{p}, u_\lambda \rangle_{\mathcal{U}}| \, .$$

Proof. The estimate is simply a consequence of taking a dual product of the optimality condition of (2.2) with u_λ. The optimality condition reads as

$$K^* K(\hat{u} - \tilde{u}) + \alpha \hat{p} = K^* \eta \, ,$$

for $\hat{p} \in \partial J(\hat{u})$. Taking the dual product with u_λ yields

$$\langle K^* K u_\lambda, \hat{u} - \tilde{u} \rangle_{\mathcal{U}} = \langle K^* \eta - \alpha \hat{p}, u_\lambda \rangle_{\mathcal{U}} \, ,$$

which is equivalent to

$$\left\langle \frac{1}{\lambda} p_\lambda, \hat{u} - \tilde{u} \right\rangle_{\mathcal{U}} = \langle K^* \eta - \alpha \hat{p}, u_\lambda \rangle_{\mathcal{U}} \, .$$

The estimate for the absolute value of the right-hand side simply follows from the triangular inequality. □

REMARK 1. In case of J being one-homogeneous we have $|\langle \hat{p}, u_\lambda \rangle_{\mathcal{U}}| \leq J(u_\lambda) = \lambda$ and the estimate therefore can be used to deduce

$$\left| \left\langle \frac{1}{\lambda} p_\lambda, \hat{u} - \tilde{u} \right\rangle_{\mathcal{U}} \right| \leq \alpha\lambda + |\langle \eta, K u_\lambda \rangle_{\mathcal{H}}| \, .$$

The two terms in the error estimate can be easily interpreted. The term $\alpha\lambda$ is a worst-case estimate for the bias at the scale of the used singular value, while the second term describes the impact of noise on a specific scale. Note that the scalar product of the noise with Ku_λ implies actually some averaging of the noise, which usually reduces the noise variance stronger on larger scales than on smaller ones. This can be analyzed in particular for statistical noise models such as additive Gaussian noise.

In the following we want to demonstrate how Theorem 3 can be used to derive estimates for the difference of the reconstruction $\hat{u}$ and the input data $\tilde{u}$ on different scales.

EXAMPLE 8. With this first example we want to find a worst-case estimate for a reconstruction $\hat{u}$ of (2.4), with respect to the input data $f = \tilde{u} + \eta$ on the scale $[0, a] \subset [0, 1]$, for $0 < a < 1$. We want to point out that the ROF model is mean-value preserving, i.e. $\int_0^1 (\hat{u} - \tilde{u}) \, dx = 0$. Let us investigate the integral equation

$$\int_0^a (\hat{u} - \tilde{u}) \, dx = c_0 \int_0^1 u_0 \, (\hat{u} - \tilde{u}) \, dx + c_1 \int_0^1 u_1 \, (\hat{u} - \tilde{u}) \, dx \,,$$

$$= \int_0^1 (c_0 u_0 + c_1 u_1) \, (\hat{u} - \tilde{u}) \, dx \,,$$

$$(4.6) \qquad = \int_0^a (c_0 u_0 + c_1 u_1) \, (\hat{u} - \tilde{u}) \, dx + \int_a^1 (c_0 u_0 + c_1 u_1) \, (\hat{u} - \tilde{u}) \, dx \,,$$

with the function u_0 defined as $u_0(x) \equiv 1$, and with $u_1(x) := u^a(x)$ being the singular vector of Example 6. Then, (4.6) reads as

(4.7)
$$\int_0^a (\hat{u} - \tilde{u}) \, dx = \left(c_0 - \sqrt{\frac{1 - a}{a}} c_1 \right) \int_0^a (\hat{u} - \tilde{u}) \, dx + \left(c_0 + \sqrt{\frac{a}{1 - a}} c_1 \right) \int_a^1 (\hat{u} - \tilde{u}) \, dx \,.$$

In order to satisfy (4.7), the coefficients c_0 and c_1 have to be chosen such that they solve the linear system of equations

(4.8)
$$\begin{pmatrix} 1 & -\sqrt{\frac{1-a}{a}} \\ 1 & \sqrt{\frac{a}{1-a}} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \,.$$

It is easy to see that $c_0 = a$ and $c_1 = -\sqrt{a(1 - a)}$ solve (4.8). Thus, we obtain the following scale estimate

$$\left| \int_0^a (\hat{u} - \tilde{u}) \, dx \right| = \left| a \underbrace{\int_0^1 \hat{u} - \tilde{u} \, dx}_{=0} - \sqrt{a(1 - a)} \int_0^1 u_1 \, (\hat{u} - \tilde{u}) \, dx \right| \,,$$

$$= \sqrt{a(1 - a)} \left| \int_0^1 u_1 \, (\hat{u} - \tilde{u}) \, dx \right| \,,$$

$$\leq \sqrt{a(1 - a)} \left(\frac{\alpha}{\sqrt{a(1 - a)}} + \left| \int_0^1 \eta \, u_1 \, dx \right| \right) \,,$$

$$= \left(\alpha + \sqrt{a(1 - a)} \left| \int_0^1 \eta \, u_1 \, dx \right| \right) \,,$$

since $1/\sqrt{a(1 - a)}$ is the singular value of u_1. Note that for $\eta = 0$ we simply obtain

$$\left| \int_0^a (\hat{u} - \tilde{u}) \, dx \right| \leq \alpha \,.$$

This estimate is obviously not sharp if $a \to 1$, since $\int_0^1 \hat{u} - \tilde{u} \, dx = 0$, but gives a worst case estimate if a is far away from the boundary as in case of $a = 1/2$. If we consider e.g. $\tilde{u}(x) = x - 1/2$, the estimate guarantees that the area visualized in Figure 3 is
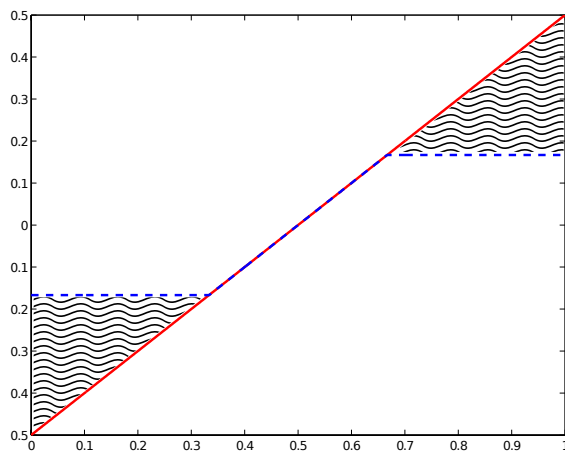
FIG. 3. *The function $\tilde{u}(x) = x - 1/2$ (solid red line) and the solution of* (2.4) *for input data given in terms of $f = \tilde{u}$ and $\alpha = 1/18$ (dashed blue line). The areas indicated by wavy lines equal $-1/18$ and $1/18$, respectively.*

equal or smaller than α. Indeed the area equals α, which becomes clear by computing the exact solution of (2.4) for $f(x) = \tilde{u}(x) = x - 1/2$. For $0 < \alpha < 1/8$ the solution satisfies

$$\hat{u}(x) = \begin{cases} \sqrt{2\alpha} - \frac{1}{2} & x \in \left[0, \sqrt{2\alpha}\right[\\ x - \frac{1}{2} & x \in \left[\sqrt{2\alpha}, 1 - \sqrt{2\alpha}\right] \\ \frac{1}{2} - \sqrt{2\alpha} & x \in \left]1 - \sqrt{2\alpha}, 1\right] \end{cases} ,$$

and thus, $\int_0^{1/2}(\hat{u} - \tilde{u})\ dx = -\int_{1/2}^1 (\hat{u} - \tilde{u})\ dx = \alpha$ holds true.

EXAMPLE 9. As a second example we want to focus on a scale estimate on the scale $[a, b] \subset [0, 1]$, for $0 < a < b < 1$. Again we study (2.4) and, in analogy to the previous example, consider

(4.9) $$\int_a^b \hat{u} - \tilde{u}\ dx = \int_0^1 (c_0 u_0 + c_1 u_1 + c_2 u_2)\ (\hat{u} - \tilde{u})\ dx\,,$$

with $u_0(x) \equiv 1$, $u_1(x) := u^a(x)$ and $u_2(x) := -u^b(x)$, for u^a and u^b being singular

vectors as defined in Example 6. Thus, we can rewrite (4.9) to

$$
\int_a^b (\hat{u} - \tilde{u})\ dx = \int_0^a (c_0 u_0 + c_1 u_1 + c_2 u_2)(\hat{u} - \tilde{u})\ dx +
$$

$$
\int_a^b (c_0 u_0 + c_1 u_1 + c_2 u_2)(\hat{u} - \tilde{u})\ dx +
$$

$$
\int_b^1 (c_0 u_0 + c_1 u_1 + c_2 u_2)(\hat{u} - \tilde{u})\ dx
$$

$$
= \left(c_0 - \sqrt{\frac{1-a}{a}} c_1 + \sqrt{\frac{1-b}{b}} c_2 \right) \int_0^a (\hat{u} - \tilde{u})\ dx +
$$

$$
\left(c_0 + \sqrt{\frac{a}{1-a}} c_1 + \sqrt{\frac{1-b}{b}} c_2 \right) \int_a^b (\hat{u} - \tilde{u})\ dx +
$$

$$
\left(c_0 + \sqrt{\frac{a}{1-a}} c_1 - \sqrt{\frac{b}{1-b}} c_2 \right) \int_b^1 (\hat{u} - \tilde{u})\ dx\,.
$$

Similar to the previous example we therefore have to make sure that c_0, c_1 and c_2 satisfy

(4.10)
$$
\begin{pmatrix} 1 & -\sqrt{\frac{1-a}{a}} & \sqrt{\frac{1-b}{b}} \\ 1 & \sqrt{\frac{a}{1-a}} & \sqrt{\frac{1-b}{b}} \\ 1 & \sqrt{\frac{a}{1-a}} & -\sqrt{\frac{b}{1-b}} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.
$$

Solving (4.10) for c_0, c_1 and c_2 yields $c_0 = b - a$, $c_1 = \sqrt{a(1-a)}$ and $c_2 = \sqrt{b(1-b)}$. Consequently, Theorem 3 allows us to compute the estimate

$$
\left| \int_a^b (\hat{u} - \tilde{u})\ dx \right| \leq (b-a) \underbrace{\left| \int_0^1 \hat{u} - \tilde{u}\ dx \right|}_{=0} + \sqrt{a(1-a)} \left| \int_0^1 u_1 (\hat{u} - \tilde{u})\ dx \right|
$$

$$
+ \sqrt{b(1-b)} \left| \int_0^1 u_2 (\hat{u} - \tilde{u})\ dx \right|
$$

$$
\leq 2\alpha + \sqrt{a(1-a)} \left| \int_0^1 \eta\ u_1\ dx \right| + \sqrt{b(1-b)} \left| \int_0^1 \eta\ u_2\ dx \right|.
$$

In case of clean data, i.e. $\eta = 0$, we see that

$$
\left| \int_a^b (\hat{u} - \tilde{u})\ dx \right| \leq 2\alpha
$$

holds. It is remarkable that the worst-case estimate on the scale $[a, b]$ does not depend on the boundary values a and b. However, if we want to estimate the mean value, then an additional factor $\frac{1}{b-a}$ comes up, which increases with decreasing size of the interval.

5. Exact reconstruction of singular vectors. In this section we discuss exact solutions of variational regularization methods, namely the recovery of singular vectors. We will relate this results to properties of Bregman distances. First of all, we want to recall a criterion that has been shown by Meyer in [52] in order to derive trivial ground states for the ROF-model. These considerations can be generalized by the use of Bregman distances as it can be seen by the following theorem.

THEOREM 4. *Let f be such that*

(5.1)
$$\frac{1}{\alpha} K^* f \in \partial J(0)$$

is satisfied, then a minimizer of (2.2) *is given by* $\hat{u} \equiv 0$. *Vice versa,* $\hat{u} \equiv 0$ *is a minimizer of* (2.2) *only if* (5.1) *holds.*

Proof. We can rewrite (2.2) to

$$\hat{u} \in \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku\|_{\mathcal{H}}^2 + \alpha \left(J(u) - \left\langle \frac{1}{\alpha} K^* f, u \right\rangle_{\mathcal{U}} \right) + \frac{1}{2} \|f\|_{\mathcal{H}}^2 \right\}$$

Since (5.1) is satisfied, we can define $q := (K^* f)/\alpha$ such that

$$D_J^q(u, 0) = J(u) - J(0) - \langle q, u \rangle_{\mathcal{U}}$$

is a non-negative Bregman distance. Hence, ignoring the constant part $1/2\|f\|_{\mathcal{H}}$ we have

$$\hat{u} \in \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku\|_{\mathcal{H}}^2 + \alpha D_J^q(u, 0) \right\}$$

for which the obvious minimizer is given via $\hat{u} = 0$, since both terms are non-negative and vanish for $u = 0$. It is straightforward to see the opposite condition from the optimality condition for $\hat{u} = 0$. □

REMARK 2. Note that if (5.1) is satisfied for a specific $\tilde{\alpha}$, then (5.1) is automatically guaranteed for every $\alpha \geq \tilde{\alpha}$, since $(K^* f)/\tilde{\alpha} \in \partial J(0)$ implies

$$J(v) \geq \left\langle \frac{1}{\tilde{\alpha}} K^* f, v \right\rangle_{\mathcal{U}},$$

for all $v \in \mathrm{dom}(J)$. If we multiply both sides of the inequality with $\tilde{\alpha}$ we obtain

$$\tilde{\alpha} J(v) \geq \langle K^* f, v \rangle_{\mathcal{U}},$$

since $\tilde{\alpha}$ is positive. Due to the positivity of J we even have

$$\alpha J(v) \geq \tilde{\alpha} J(v)$$

for all $v \in \mathrm{dom}(J)$ and $\alpha \geq \tilde{\alpha}$, and hence, (5.1) is guaranteed for all $\alpha \geq \tilde{\alpha}$.

Theorem 4 yields an explicit condition on the regularization parameter α to enforce the solution of (2.2) to be zero. Furthermore, according to the following Lemma for singular vectors u_λ of one-homogeneous functionals there even have to exist parameters α such that (5.1) is fulfilled for $f = K u_\lambda$.

LEMMA 2. *Let J be one-homogeneous. If $u \in \mathrm{dom}(J) \cap \mathrm{dom}(K)$ is a function such that (5.1) does not hold for any $\alpha \in \mathbb{R}_{>0}$ with data $f = Ku$, i.e.*

$$\frac{1}{\alpha} K^* K u \notin \partial J(0) \quad \forall \alpha \in \mathbb{R}_{>0} \ ,$$

then, u is not a singular vector with singular value $\lambda \neq 0$.

Proof. We want to prove the statement by contradiction. We therefore assume that on the one hand, u is a singular vector with singular value λ, i.e. $\lambda K^* K u = p \in \partial J(u)$. Taking a duality product of this relation with u yields the equality

$$(5.2) \qquad\qquad\qquad \lambda \|Ku\|_{\mathcal{H}}^2 = J(u) \, ,$$

due to the one-homogeneity of J. Moreover, from the definition of the subdifferential, the singular value property yields

$$(5.3) \qquad\qquad J(v) \geq J(u) + \lambda \left\langle K^* K u, v - u \right\rangle_{\mathcal{U}} \quad \forall v \in \mathrm{dom}(J) \, .$$

On the other hand, we know due to $(K^* K u)/\alpha \notin \partial J(0)$ for all $\alpha \in \mathbb{R}_{>0}$ that there has to exist a function $v \in \mathrm{dom}(J)$ with

$$(5.4) \qquad\qquad\qquad \left\langle Ku, Kv \right\rangle_{\mathcal{H}} > \alpha J(v) \, .$$

If we insert (5.3) into (5.4), for the particular choice of v we therefore obtain

$$\left\langle Ku, Kv \right\rangle_{\mathcal{H}} > \alpha \left(J(u) + \lambda \left\langle Ku, Kv \right\rangle_{\mathcal{H}} - \lambda \|Ku\|_{\mathcal{H}}^2 \right)$$
$$(5.5) \qquad \Leftrightarrow (1 - \lambda\alpha) \left\langle Ku, Kv \right\rangle_{\mathcal{H}} > \alpha \left(J(u) - \lambda \|Ku\|_{\mathcal{H}}^2 \right) \, .$$

Equation (5.5) is supposed to be true for every $\alpha \in \mathbb{R}_{>0}$, especially for the particular choice $\alpha = 1/\lambda$. In this case, (5.5) reads as

$$\lambda \|Ku\|_{\mathcal{H}}^2 > J(u) \, ,$$

for $\lambda > 0$, and therefore is a contradiction to (5.2). ☐

The equivalent reverse statement of Lemma 2 is that for every data f created by a singular vector u_λ, i.e. $f = Ku_\lambda$, there exists a parameter $\tilde{\alpha}$ such that (5.1) is valid for $\alpha \geq \tilde{\alpha}$. Moreover, condition (5.1) guarantees that the data f needs to satisfy certain properties in order to vanish for a large regularization parameter α, e.g. f does need to have zero mean for $K = I$ in the case of TV-regularization.

REMARK 3. Note that, however, it is possible for a particular function f that there does not exist a parameter α such that (5.1) is fulfilled, although $f = Ku_0$ is given in terms of a trivial ground state (and therefore in terms of a singular vector) with singular value $\lambda = 0$. A simple example would be that $f \not\equiv 0$ is a constant function in case of $K = I$ and $J = \mathrm{TV}$.

5.1. Clean data. In case of clean data $f = \gamma K u_\lambda$, $\gamma > 0$ and u_λ being a non-trivial singular vector, we are interested in finding a solution of (2.2) that can be expressed in terms of this singular vector, i.e. $\hat{u} = c\tilde{u}$ for a positive constant c. We want to call such a function *almost exact solution*.

The following theorem gives us the conditions on α needed for recovering a multiple of u_λ.

THEOREM 5. *Let J be one-homogeneous. Furthermore, let u_λ be a singular vector with corresponding singular value λ. Then, if the data f is given as $f = \gamma K u_\lambda$ for a positive constant γ, a solution of (2.2) is $\hat{u} = cu_\lambda$ for*

$$c = \gamma - \alpha\lambda\,,$$

if $\gamma > \alpha\lambda$ is satisfied.

Proof. Again, we rewrite (2.2) in terms of a Bregman distance. Inserting $f = \gamma K u_\lambda$ yields

$$\hat{u} \in \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - \gamma K u_\lambda\|_{\mathcal{H}}^2 + \alpha J(u) \right\}$$

$$\in \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - cK u_\lambda\|_{\mathcal{H}}^2 + \alpha J(u) + \alpha J(cu_\lambda) - \frac{\gamma - c}{\lambda} \langle \lambda K^* K u_\lambda, u \rangle_{\mathcal{U}} \right.$$

$$\left. + \frac{1}{2} \left(\langle \gamma K u_\lambda, \gamma K u_\lambda \rangle_{\mathcal{H}} + \langle cK u_\lambda, cK u_\lambda \rangle_{\mathcal{H}} \right) - \alpha J(cu_\lambda) \right\}\,.$$

By ignoring the constant part, for $\gamma > \alpha\lambda$ and $c = \gamma - \alpha\lambda > 0$ we therefore obtain

$$\hat{u} = \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - cK u_\lambda\|_{\mathcal{H}}^2 + \alpha D_J^q(u, cu_\lambda) \right\}\,,$$

with

$$q = \lambda K^* K u_\lambda \in \partial J(u_\lambda) \overset{J \text{ one-homogeneous}}{=} \partial J(cu_\lambda)\,.$$

An obvious minimizer is $\hat{u} = cu_\lambda$. $\square$

Note that the above result does not yield that the singular value is the unique minimizer, except for K having trivial nullspace. To see this, let us consider model (2.2) with $J(u) = \|u\|_{\ell^1}$ and K being the matrix

$$K := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\,.$$

It is obvious that K is normalized with respect to the ℓ^2-norm, but neither is injective nor surjective. Due to Example 4 both e_1 and e_2 are singular vectors. However, both yield the same output $f = (1/\sqrt{2}, 1/\sqrt{2})^T$ and therefore both $\hat{u} = (1 - \alpha)e_1$ and $\hat{u} = (1 - \alpha)e_2$ satisfy Theorem 5.

We also mention that the main line of Theorem 5 also holds for p-homogeneous functionals, but with different constants c. In the following we turn our attention to noisy data, where the one-homogeneity is much more essential, e.g. exact reconstruction for a wide class of noise realizations cannot hold for quadratic regularizations.

5.2. Noisy data. The multivaluedness of the subdifferential ∂J allows to obtain almost exact solutions even in the presence of noisy data, i.e. $f = \gamma K u_\lambda + n$, though the case of noisy data is slightly more complicated to prove. If the most significant features of u_λ with respect to the regularization energy J are left unaffected by the

noise, then the following theorem guarantees almost exact recovery of the singular vector u_λ.

THEOREM 6. *Let J be one-homogeneous. Furthermore, let u_λ be a singular vector with corresponding singular value λ. The data f is assumed to be corrupted by noise n, i.e. $f = \gamma K u_\lambda + n$ for a positive constant γ, such that there exist positive constants μ and η with*

$$(5.6) \qquad \mu K^* K u_\lambda + \eta K^* n \in \partial J(u_\lambda) \, .$$

Then, a solution of (2.2) is given by $\hat{u} = c u_\lambda$ for

$$c = \gamma - \alpha\lambda + \frac{\lambda - \mu}{\eta} \, ,$$

if γ satisfies the SNR-condition

$$(5.7) \qquad \gamma > \frac{\mu}{\eta} \, ,$$

and if $\alpha \in [1/\eta, \gamma/\lambda + 1/\eta[$ holds.

Proof. Similar to the proof of Theorem 5 we rewrite (2.2) to

$$
\begin{aligned}
\hat{u} &= \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - \gamma K u_\lambda - n\|_{\mathcal{H}}^2 + \alpha J(u) \right\} \\
&= \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - c K u_\lambda\|_{\mathcal{H}}^2 + \alpha J(u) - \alpha \left\langle \frac{\gamma - c}{\alpha} K^* K u_\lambda + \frac{1}{\alpha} K^* n, u \right\rangle_{\mathcal{U}} \right\} \\
&= \underset{u \in \mathrm{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - c K u_\lambda\|_{\mathcal{H}}^2 + \alpha D_J^q(u, c u_\lambda) \right\} ,
\end{aligned}
$$

with obvious minimizer $\hat{u} = c u_\lambda$, if we neglect the constant parts and if we can manage to choose c such that

$$\frac{\gamma - c}{\alpha} K^* K u_\lambda + \frac{1}{\alpha} K^* n \in \partial J(u_\lambda) = \partial J(c u_\lambda) \, .$$

Note that since $\partial J(u_\lambda)$ is a convex set not only $\lambda K^* K u_\lambda$ and (5.6) are elements of $\partial J(u_\lambda)$, but also any convex combination, i.e.

$$((1 - \beta) \lambda + \beta\mu) K^* K u_\lambda + \beta\eta K^* n \in \partial J(u_\lambda) \, ,$$

for each $\beta \in [0, 1]$.

Hence, we need to choose $c > 0$ and $\beta \in [0, 1]$ such that $1/\alpha = \beta\eta$ and $(\gamma - c)/\alpha = (1 - \beta)\lambda + \beta\mu$. Therefore, solutions for β and c are

$$\beta = \frac{1}{\alpha\eta}$$

and

$$c = \gamma - \alpha\lambda + \frac{\lambda - \mu}{\eta} \, .$$

In order to satisfy $\beta \leq 1$ and $c > 0$, α has to be chosen such that α is bounded via

$$\frac{1}{\eta} \leq \alpha < \frac{\gamma}{\lambda} + \frac{1}{\eta} + \frac{\mu}{\lambda\eta} \, .$$

This condition can only be satisfied, if $\gamma > \mu/\eta$ holds. □

At a first glance (5.6) seems unreasonably restrictive, e.g. for smooth functionals J it can only hold if the noise is generated by the singular vector itself. This however differs completely in the situation of singular regularization functionals with large subdifferentials. To see this, let us consider the setup of Example 4 with a ground state e_i, together with the reasonable assumptions that no pair of columns in K is linearly independent. For

$$p = \lambda K^* K e_i \in \partial \|e_i\|_{\ell^1(\mathbb{R}^N)}$$

we have

$$p_i = p \cdot e_i = \lambda e_i \cdot K^* K e_i = \lambda = 1$$

and hence for $j \neq i$

$$|p \cdot e_j| = |e_j \cdot K^* K e_i| = |K_j \cdot K_i| < 1,$$

where we denote by K_j the j-th column of K. Now let $n \in \mathbb{R}^M$ be the noise vector and $v = K^* n$. In order to satisfy (5.6) we need

$$1 = p \cdot e_i = \mu e_i \cdot K^* K e_i + \eta e_i \cdot v = \mu + \eta v_i,$$

and

$$1 \geq |p \cdot e_j| = |\mu K_j \cdot K_i + \eta v_j|.$$

The first condition is satisfied if $\mu = 1 - \eta v_i$. Since then

$$|\mu K_j \cdot K_i + \eta v_j| = |K_j \cdot K_i| + \mathcal{O}(\eta)$$

for η small and $|K_j \cdot K_i| < 1$, we indeed conclude that there always exists η such that (5.6) is satisfied. Note that in this case $\lambda = 1$ and hence $c = \gamma - \alpha + v_i$.

6. Bias of variational methods. We have seen in the previous section that there remains a small bias in the exact reconstruction of singular vectors, which is incorporated in the fact that $c < \gamma$, with difference depending on the actual singular value. For $f = \gamma K u_\lambda$, this difference yields a residual

$$\|Ku - f\|_{\mathcal{H}} = \alpha\lambda\|Ku_\lambda\|_{\mathcal{H}} = \alpha\lambda,$$

i.e. a bias in the solution, which is minimal for the ground state λ_0. In this short section we will show that indeed this bias prevails for arbitrary data and the residual is bounded below by $\alpha\lambda_0$, which again confirms the extremal role of the ground state.

THEOREM 7. *Let J be one-homogeneous, $f \in \mathcal{H}$ be arbitrary, $\alpha > 0$, and let*

$$u^\alpha \in \underset{u \in \mathcal{U}}{\arg\min} \left(\frac{1}{2}\|Ku - f\|_{\mathcal{H}}^2 + \alpha J(u) \right).$$

Then

(6.1) $\|Ku^\alpha\|_\mathcal{H} \leq \max\{\|f\|_\mathcal{H} - \alpha\lambda_0, 0\},$

where λ_0 is the smallest singular value. As a direct consequence, if $\|f\|_\mathcal{H} \geq \alpha\lambda_0$, then

(6.2) $\|Ku^\alpha - f\|_\mathcal{H} \geq \alpha\lambda_0,$

which is sharp if f is a multiple of Ku_0 with the ground state u_0.

Proof. If $Ku^\alpha = 0$, then the estimate is obviously satisfied. Thus, we restrict our attention to the case $Ku^\alpha \neq 0$ and define $v := \frac{u^\alpha}{\|Ku^\alpha\|}$. From the dual product of the optimality condition with v we see that for a subgradient $p^\alpha \in \partial J(u^\alpha) = \partial J(v)$ that

$$\langle K^*Ku^\alpha, v\rangle_\mathcal{U} + \alpha\langle p^\alpha, v\rangle_\mathcal{U} = \langle K^*f, v\rangle_\mathcal{U}.$$

Due to the one-homogeneity of J we conclude $\langle p^\alpha, v\rangle_\mathcal{U} = J(v)$ and hence,

$$\|Ku^\alpha\|_\mathcal{H} + \alpha J(v) = \langle f, Kv\rangle_\mathcal{H}.$$

By the definition of the ground state we conclude $J(v) \geq \lambda_0$ and by further estimating the right-hand side via the Cauchy-Schwartz inequality $\langle f, Ku_\alpha\rangle_\mathcal{H} \leq \|f\|_\mathcal{H}\|Ku_\alpha\|_\mathcal{H}$ we have

$$\|Ku^\alpha\|_\mathcal{H} + \alpha\lambda_0 \leq \|f\|_\mathcal{H},$$

which implies the assertion. □

The bias in the residual can to some extent also be translated to the regularization functional, as the following result shows:

THEOREM 8. *Let J be one-homogeneous, $f = K\tilde{u}$ for $\tilde{u} \in \mathcal{U}$ with $J(\tilde{u}) < \infty$, such that $\|f\|_\mathcal{H} \geq \alpha\lambda_0$. Moreover let $\alpha > 0$ and*

$$u^\alpha \in \arg\min_{u\in\mathcal{U}} \left(\frac{1}{2}\|Ku - f\|_\mathcal{H}^2 + \alpha J(u)\right).$$

Then

(6.3) $J(u^\alpha) \leq J(\tilde{u}) - \frac{\alpha}{2}\lambda_0^2,$

with λ_0 denoting the smallest singular value.

Proof. By the definition of u^α as a minimizer and (6.2) we conclude

$$\frac{1}{2}\alpha^2\lambda_0^2 + \alpha J(u^\alpha) \leq \frac{1}{2}\|Ku^\alpha - f\|_\mathcal{H}^2 + \alpha J(u^\alpha) \leq \alpha J(\tilde{u}),$$

which yields the assertion after dividing by α. □

7. Unbiased recovery and inverse scale space. In Section 5 we have seen that in standard variational methods of the form (2.2) one can recover singular vector (almost) exactly with a loss of contrast. In this chapter we want to extend this topic to the question of exact recovery without a loss of contrast, in the absence and presence of noise. For this sake we are going to investigate the concept of the inverse scale

space flows, which have displayed superior properties to solutions of (2.2) in several numerical tests (cf. [19, 60]).

Inverse scale space methods can be derived asymptotically from the Bregman iteration

$$(7.1) \qquad u^{k+1} \in \underset{u \in \text{dom}(J)}{\arg\min} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{H}}^2 + \alpha(J(u) - \langle p^k, u \rangle_{\mathcal{U}}) \right\},$$

for which the subgradient $p^k \in \partial J(u^k)$ satisfies $p^0 \equiv 0$ and

$$(7.2) \qquad p^k = p^{k-1} + \frac{1}{\alpha} K^*(f - Ku^k).$$

In the limit $\alpha \to \infty$ one can interpret $\Delta t = \frac{1}{\alpha}$ as a time step tending to zero. Thus, we obtain the inverse scale space flow

$$(7.3) \qquad \partial_t p(t) = K^*(f - Ku(t)), \qquad p(t) \in \partial J(u(t)).$$

We refer to [19, 10, 18, 20, 21, 23, 53] for detailed discussions of the inverse scale space method and its analysis.

We want to mention that analogous results on exact respectively unbiased reconstruction can be obtained for the Bregman iteration, clearly with some dependence on the value of α, further details can be found in [10].

7.1. Clean data. Similar to Section 5.1, we are going to consider data $f = \gamma Ku_\lambda$, with u_λ being a singular vector. For this setup we are able to derive the following result:

THEOREM 9. *Let J be one-homogeneous and let u_λ be a singular vector with corresponding singular value λ. Then, if the data f are given by $f = \gamma Ku_\lambda$ for a positive constant γ, a solution of the inverse scale space flow (7.3) is given by*

$$(7.4) \qquad u(t) = \begin{cases} 0 & \text{if } t < t_* \\ \gamma u_\lambda & \text{if } t \geq t_* \end{cases}$$

for $t_* = \lambda/\gamma$.

Proof. First of all we see with Theorem 4 that for $t < t_*$ we have

$$p(t) = tK^*f = t\gamma K^*Ku_\lambda = t\frac{\gamma}{\lambda}p_\lambda \in \partial J(0).$$

Since $\partial_t p = K^*f$ and $p(0) = 0$, $u(t) = 0$ is a solution of (7.3).

For time $t \geq t_*$ a continuous extension of p is given by the constant $p(t) = p(t_*)$ and $u(t) = u(t_*)$. Due to the one-homogeneity, with $t_* = \lambda/\gamma$ we obtain

$$p(t_*) \in \partial J(u_\lambda) = \partial J(\gamma u_\lambda), \qquad K^*(f - Ku(t_*)) = 0.$$

Thus, $\partial_t p = 0$ yields indeed a solution of (7.3) for $t \geq t_*$. □

7.2. Noisy data. As in Section 5 the case of noisy data is a bit more complicated. In order to recover a singular vector exactly despite the contamination of the data f with noise, we basically need the signal ratio μ as introduced in Theorem 6 to equal the singular value λ. We also mention that the stopping time is the regularization parameter in inverse scale space methods, thus we can only expect the exact

reconstruction to happen in a time interval (t_*, t_{**}). More precisely we obtain the following result:

THEOREM 10. *Let J be one-homogeneous and let u_λ be a singular vector with corresponding singular value λ. The data f is assumed to be corrupted by noise n, i.e. $f = \gamma K u_\lambda + n$ for a positive constant γ, such that there exist positive constants μ and η that satisfy (5.6) and (5.7). Then, a solution of the inverse scale space flow (7.3) is given by*

$$(7.5) \qquad u(t) = \begin{cases} 0 & \text{if } t < t_* \\ c u_\lambda & \text{if } t_* \leq t < t_{**} \end{cases}$$

for

$$(7.6) \qquad c = \gamma + \frac{\lambda - \mu}{\eta} \,.$$

and

$$t_* = \frac{\lambda \eta}{\lambda + \gamma \eta - \mu} < t_{**} = \eta.$$

Proof. With a similar argumentation as in the proof of Theorem 9 we obtain $u(t) = 0$ for $t < t_*$ and

$$p(t_*) = t_* \gamma K^* K u_\lambda + t_* K^* n$$

as the corresponding subgradient to the first non-zero u for a critical time t_*. Analogous to the proof of Theorem 6 we can treat the relation above as a convex combination of $\lambda K^* K$ and (5.6) for any $\beta \in [0, 1]$, and determine $\beta = \frac{\lambda}{\lambda + \gamma \eta - \mu}$ and subsequently t_* as above. Moreover, we see that $u(t_*) = c u_\lambda$ is a feasible solution with subgradient $p(t_*)$, which we extrapolate as constant for further times up to some time t_{**}. Then, from $p(t_*) = t_* K^* f$ and $\partial_t p(t) = K^*(f - c K u_\lambda)$ we conclude

$$\begin{aligned} p(t) &= t K^* f - (t - t_*) c K^* K u_\lambda \\ &= t K^*(\gamma K u_\lambda + n) - (t - t_*) c K^* K u_\lambda \\ &= t K^* n + (\gamma t - c(t - t_*)) K^* K u_\lambda \,. \end{aligned}$$

To obtain $p(t) \in \partial J(u(t_*))$ we again compare convex combinations of (5.6) and $\lambda K^* K u_\lambda$ with parameter $\beta \in [0, 1]$. We need to choose $\beta = \frac{t}{\eta}$, which is only possible for $t < t_{**} = \eta$. Further we obtain that $\mu = \gamma t - c(t - t_*)$ needs to hold. This identity can be verified with the above formulas for t_* and c. □

8. Further examples. In the following we shall discuss several further examples to illustrate the use of nonlinear singular values:

8.1. Hilbert space norms. The obvious first starting point is to consider regularizations with Hilbert space norms, i.e.

$$(8.1) \qquad J(u) = \|u\|_{\mathcal{U}} \,.$$

Note that we focus on the one-homogeneous case here, i.e. we do not use the squared Hilbert space norm as in the standard formulation of Tikhonov regularization. However, it is easy to check that there is a one-to-one relation between the singular values

of the problems with squared and non-squared norms, such that they are equivalent. It is straight-forward to see that the singular values are determined from

$$(8.2) \qquad \lambda K^* K u = \frac{u}{\|u\|_{\mathcal{U}}},$$

and with our normalization we see $\lambda = \|u\|_{\mathcal{U}}$. Thus if u_n is a singular vector in the classical definition

$$(8.3) \qquad K^* K u_n = \sigma_n^2 u_n,$$

it is also a singular vector in the new definition (at least after appropriate normalization). The linear singular values σ_n are related to the novel values λ_n via

$$(8.4) \qquad \sigma_n^2 = \frac{1}{\lambda_n \|u_n\|_{\mathcal{U}}} = \frac{1}{\lambda_n^2},$$

which is consistent with our original definition as singular values being related to the regularization functional rather than to K.

The fact that singular values yield exact solutions of the variational problem is not new and is directly inferred as a special case of the standard theory (cf. [36]). However, it is surprising how the behavior of the inverse scale space method changes when rescaling from the squared norm to the one-homogeneous case. In the case of $J(u) = \frac{1}{2}\|u\|_{\mathcal{U}}^2$ the inverse scale space method is equivalent to Showalter's method (cf. [62])

$$\partial_t u = K^*(f - Ku),$$

and it is well-known that singular vectors follow an exponential dynamic, i.e. if $f = Ku_\lambda$ for a singular vector u_λ, then

$$u(t) = (1 - e^{-t/\lambda})u_\lambda.$$

The behavior changes completely in the one-homogeneous case as we can conclude from the results in the previous section, since the solution remains zero in finite time and then jumps exactly to $u(t) = u_\lambda$ at the critical time.

8.2. Total variation. We have already used the ROF model for denoising at several instances as a simple illustrative example, in particular in spatial dimension one. In this section we want to extend the considerations of the one dimensional ROF model to data that is corrupted by noise. Moreover, we want to highlight the connection between singular vectors and characteristic functions of so-called calibrable sets as introduced in [5, 4], with respect to the isotropic total variation functional in higher spatial dimension. In [27] the theory of calibrable sets has also been extended to more general (and in particular anisotropic) regularization functionals, which we disregard here for the sake of simplicity. Instead, we introduce an analytical solution of the anisotropic total variation regularization in terms of the singular vector definition, similar to the previous examples in spatial dimension one.

Unbiased recovery in practice. In this section we briefly want to illustrate that in practice the violation of the assumptions of the Theorems 9 and 10 can indeed yield undesired artifacts in the reconstruction. We therefore want to investigate the piece-wise constant function $u_4 : [0,1] \to \{-1, 1\}$

$$(8.5) \qquad u_4(x) := \begin{cases} 1 & x \in \left[\frac{1}{4}, \frac{3}{4}\right] \\ -1 & \text{else} \end{cases}.$$

It is straightforward to see that u_4 is a singular vector of TV with singular value $\lambda = 4$, and with the corresponding dual singular vector p_4 satisfying the relation $p_4 = q_4'$ (in terms of a weak derivative) for q_4 defined as

$$(8.6) \qquad q_4(x) := 4 \begin{cases} -x & x \in \left[0, \frac{1}{4}\right[\\ x - \frac{1}{2} & x \in \left[\frac{1}{4}, \frac{3}{4}\right] \\ 1 - x & x \in \left]\frac{3}{4}, 1\right] \end{cases}.$$

Both functions are visualized in Figure 4. Now assume we want to compute the
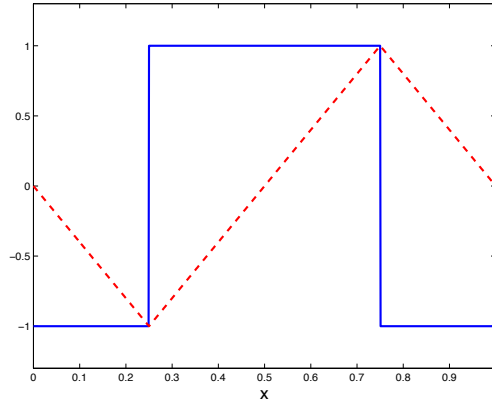
FIG. 4. *The functions u_4 (solid blue line) and q_4 (dashed red line) as defined in* (8.5) *and* (8.6), *respectively.*

minimizer of (2.4) for our data given in terms of $f(x) = u_4(x) + n(x)$. Here, n represents a noise function which we assume to have mean zero (i.e. $\int_0^1 n(x)\, dx = 0$), and to fulfill $N(0) = N(1)$, with N denoting the primitive of n (i.e. $N'(x) = n(x)$). In order to apply Theorem 6 we need to guarantee the existence of constants $\mu > 0$ and $\eta \geq 1/\alpha$ such that (5.6) is satisfied. We therefore make the attempt to define

$$q(x) := \frac{\mu}{4} q_4(x) + \eta N(x),$$

for which we obtain $q(0) = q(1) = 0$, due to the definition of n. Moreover, we discover

$$\langle q', u_4 \rangle_{L^2([0,1])} = \mu + 2\eta \int_{\frac{1}{4}}^{\frac{3}{4}} n(x)\, dx$$
$$= 4 \left(\frac{\mu}{4} + \frac{\eta}{2} \left(N\left(\frac{3}{4}\right) - N\left(\frac{1}{4}\right) \right) \right),$$

which equals $\mathrm{TV}(u_4) = 4$ if μ satisfies

$$\mu = 4 - 2\eta \left(N\left(\frac{3}{4}\right) - N\left(\frac{1}{4}\right) \right).$$

Assume $N(3/4) \geq N(1/4)$, then we even obtain

$$(8.7) \qquad \mu \leq 4 - \frac{2}{\alpha} \left(N\left(\frac{3}{4}\right) - N\left(\frac{1}{4}\right) \right) \leq 4 = \lambda,$$

due to $\eta \geq 1/\alpha$. Note that in order to obtain $\mu > 0$ we need to ensure

$$(8.8) \qquad \alpha > \frac{1}{2} \left(N \left(\frac{3}{4} \right) - N \left(\frac{1}{4} \right) \right),$$

otherwise Theorem 6 cannot be applied. Assuming to choose $\eta = 1/\alpha$, the loss of contrast modifies to

$$c = 1 - 4\alpha + 2 \left(N \left(\frac{3}{4} \right) - N \left(\frac{1}{4} \right) \right)$$

in case that (8.8) does hold.

(a) $\alpha = \frac{19\pi+2}{152\pi}$

(b) $\alpha = \frac{1}{75\pi}$

(c) $\alpha = \frac{1}{76\pi}$

(d) Closeup, $\alpha = \frac{1}{75\pi}$

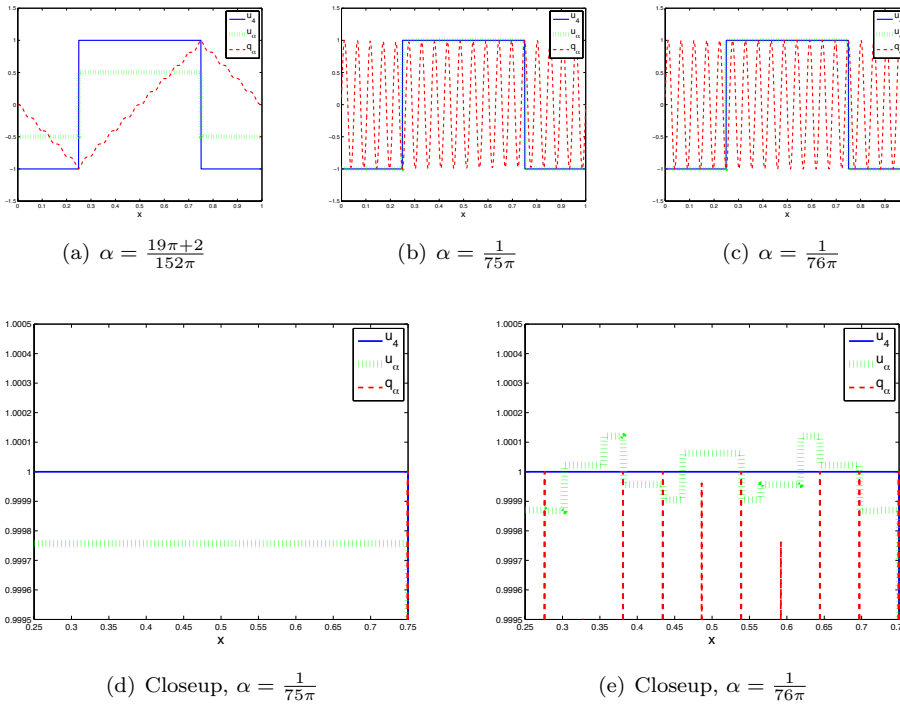(e) Closeup, $\alpha = \frac{1}{76\pi}$

FIG. 5. *Computational ROF-reconstructions for input datum $f = u_4 + n$, with u_4 and n being defined as in (8.5) and (8.9), respectively. It is remarkable to see that as soon as α is chosen such that Theorem 6 cannot be applied, the numerical computations fail to compute a multiple of u_4, indicating the sharpness of the Theorem.*

Let us consider a specific example now. We decide to choose the periodic function

$$(8.9) \qquad n(x) := A \cos (38\pi x)$$

to be our noise function, with amplitude A and frequency 38. Note that this noise function satisfies the mean zero property as well as $N(0) = N(1)$, for $N(x) = (A \sin (38\pi x))/(38\pi)$. Moreover, we compute $N(3/4) = A/(38\pi)$ and $N(1/4) = -A/(38\pi)$, so that (8.7) now reads as $\mu \leq 4 (1 - A/(38\alpha\pi))$, and according to (8.8), α should be chosen to satisfy

$$\alpha > \frac{A}{38\pi}.$$

In Figure 5 you can see several computational solutions of (2.4) for the specified input datum $f = u_4 + n$, for $A = 1/2$ and numerous α-values. The computations nicely indicate that as soon as the assumptions of Theorem 6 are violated, artifacts are introduced in the computational reconstruction.

In the unbiased case of Theorem 10 we may conclude from the considerations above that we obtain $u(t) = u_4$ as the solution of the Inverse Scale Space Flow (7.3), for $4 \leq t < (38\pi)/A$ and $A < (19\pi)/2$.

The ROF model and calibrable sets. In [52] Y. Meyer has basically proved that the characteristic function of a circle is a singular vector of isotropic total variation on $\mathbb{R}^N$. In [5, 4] the class of characteristic functions that correspond to singular vectors in the terminology of this paper has been extended to calibrable sets. In the following we want to recall properties of calibrable sets and show why they correspond to singular vectors of total variation.

DEFINITION 5. *Let $C \subset \mathbb{R}^2$ be a bounded, convex and connected set with its boundary ∂C being of class $C^{1,1}$. Then C is called* calibrable *if there exists a vector field $\xi \in L^\infty(C; \mathbb{R}^2)$ with $\|\xi\|_\infty \leq 1$ such that*

(8.10)
$$-div\,\xi = const = \lambda_C := \frac{P(C)}{|C|} \ in \ C$$
$$\xi \cdot \nu = -1 \ a. \ e. \ on \ \partial C$$

holds, for ν denoting the outer unit normal to ∂C, $P(C)$ denoting the perimeter and with $|C|$ representing the volume of C.

REMARK 4. Note that the perimeter simply equals the isotropic total variation $\mathrm{TV}(\chi_C)$ of the corresponding characteristic function χ_C.

THEOREM 11. *In [4] it has been proved that for calibrable sets the following conditions are satisfied.*
- *C is the solution of the problem*

$$\min_{X \subset C} P(X) - \lambda|X|\,.$$

- *The inequality*

$$\operatorname*{ess\,sup}_{x \in \partial C} \kappa(x) \leq \lambda$$

holds, with $\kappa(x)$ denoting the curvature of ∂C at point x.

In [9] it has already been proved that for calibrable sets C the solution of the isotropic total variation with an characteristic function χ_C as the input datum satisfies $\hat{u}(x) = (1 - \lambda\alpha)\chi_C(x)$. Thus, Theorem 5 implies that the input datum already has to be a singular vector with singular value λ.

We finally mention that Agueh and Carlier [3] have investigated a related class of ground state problems for total variation with constraints of the form $\int_\Omega G(|u|)\,dx = 1$. Such may be interesting for denoising with noise models different from additive Gaussian, where hardly any examples of exact solutions exist, except for single ones in [10].

The anisotropic ROF model in two dimensions. Analytical solutions of the isotropic ROF model have widely been studied and discussed in literature. However, analytical solutions of the anisotropic ROF model have not attracted a similar attention, although many of them are easier to describe, as we are going to see in the following.
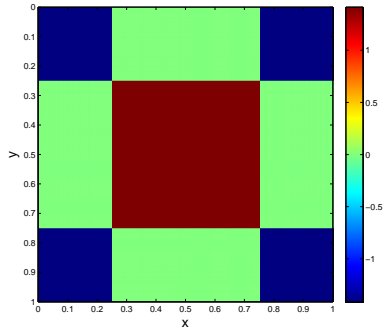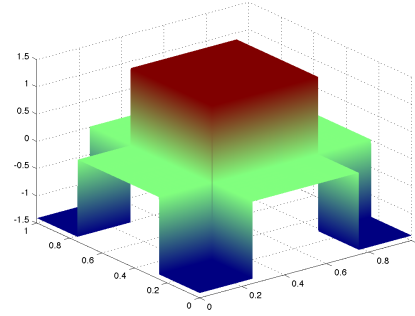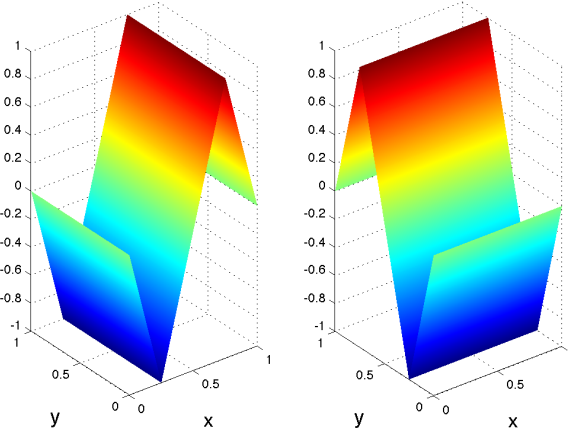

(a) On-top view of $u_{\sqrt{32}}$ (b) 3D view of $u_{\sqrt{32}}$

(c) 3D views of $q^x_{\sqrt{32}}$ and $q^y_{\sqrt{32}}$

FIG. 6. *On-top and three-dimensional view of the singular vector $u_{\sqrt{32}}$ and its dual variables $p^x_{\sqrt{32}}$ and $p^y_{\sqrt{32}}$.*

Recall (8.5) and its dual singular vector $p_4 = q_4'$, with q_4 being defined in (8.6). Let us define the two-dimensional functions $q^x_{\sqrt{32}} : [0,1]^2 \to [-1,1]$ and $q^y_{\sqrt{32}} : [0,1]^2 \to [-1,1]$ with $q^x_{\sqrt{32}}(x,y) := q_4(x)$ and $q^y_{\sqrt{32}} := q_4(y)$. Then it is easy to see that the

weak divergence of $q_{\sqrt{32}} = (q_{\sqrt{32}}^x, q_{\sqrt{32}}^y)$ divided by $\sqrt{32}$ yields the function

(8.11)

$$u_{\sqrt{32}}(x,y) = \sqrt{2} \begin{cases} 1 & (x,y) \in \left[\frac{1}{4}, \frac{3}{4}\right] \\ 0 & \left(\left(\left|x - \frac{1}{2}\right| \le \frac{1}{4}\right) \wedge \left(\left|y - \frac{1}{2}\right| > \frac{1}{4}\right)\right) \vee \left(\left(\left|x - \frac{1}{2}\right| > \frac{1}{4}\right) \wedge \left(\left|y - \frac{1}{2}\right| \le \frac{1}{4}\right)\right) \\ -1 & \text{else} \end{cases}.$$

Thus, with the same considerations as before we are able to prove that $u_{\sqrt{32}}$ is a singular vector of TV with singular value $\lambda = \sqrt{32}$, since we observe

- $q_{\sqrt{32}}^x n_x = 0$ and $q_{\sqrt{32}}^y n_y = 0$, with n_x and n_y denoting the outer unit normals of $q_{\sqrt{32}}^x$ and $q_{\sqrt{32}}^y$ in x- and in y-direction, respectively
- $\|q\|_{L^\infty([0,1]^2;\mathbb{R}^2)} = \max\left(\left|q_{\sqrt{32}}^x\right|, \left|q_{\sqrt{32}}^y\right|\right) = 1$
- $\left\langle \operatorname{div} q_{\sqrt{32}}, u_{\sqrt{32}} \right\rangle_{L^2([0,1]^2;\mathbb{R}^2)} = \mathrm{TV}(u_{\sqrt{32}})$

The singular vector $u_{\sqrt{32}}$ and the dual vectors $q_{\sqrt{32}}^x$ and $q_{\sqrt{32}}^y$ are visualized in Figure 6.

Denoising vector fields. Another generalization of the one-dimensional ROF model to multiple dimensions has been discussed recently in [15], namely the denoising of vector fields $f \in L^2(\Omega; \mathbb{R}^n)$ via minimizing

(8.12) $$u = \arg\min_{u \in L^2(\Omega;\mathbb{R}^n)} \left\{ \frac{1}{2} \int_\Omega (f - u)^2 \, dx + \int_\Omega |\operatorname{div} u| \, dx \right\}.$$

As in the case of total variation, the L^1-norm of the divergence has to be generalized to a weak form

(8.13) $$J(u) = \sup_{\substack{\varphi \in C_0^\infty(\Omega) \\ \|\varphi\|_{L^\infty(\Omega)} \le 1}} \int_\Omega u \cdot \nabla\varphi \, dx.$$

Concerning our investigation of ground states and singular values, this model yields an example with a huge set of trivial ground states. Any function $u \in L^2(\Omega; \mathbb{R}^n)$ such that $\nabla \cdot u = 0$ is obviously a ground state. In order to compute a nontrivial ground state we obtain the condition

$$\int_\Omega u \cdot v \, dx = 0 \qquad \text{for all } v \text{ with } \nabla \cdot v = 0,$$

and that u has to be a gradient field, i.e.

$$\lambda u = \nabla q.$$

The scalar q is obtained from the minimization of

$$\int_\Omega |\Delta q| \, dx \quad \text{subject to} \quad \int_\Omega \|\nabla q\|_{\ell^2(\mathbb{R}^n)}^2 \, dx = 1.$$

8.3. Support pursuit. While sparsity regularization with discrete ℓ^1-functionals has been studied extensively in the last decade, the continuum analog was investigated only recently. At a first glance it seems that $L^1(\Omega)$ would be the straight-forward extension in terms of function spaces, but similar to the case of total variation the lack of a weak-star topology in $L^1(\Omega)$ prevents the applicability and often

the existence of minimizers. Again the solution is to extend the variational approach to a slightly larger space, which is a dual space. In this case this dual space is the space of Radon measures $\mathcal{M}(\Omega)$, which is the dual space of $C_0(\Omega)$. The appropriate regularization functional as introduced in [14] is the zero-order total variation of a measure $\mu \in \mathcal{M}(\Omega)$, i.e.

$$(8.14) \qquad J(\mu) = \sup_{\substack{\varphi \in C_0(\Omega) \\ \|\varphi\|_\infty \leq 1}} \int_\Omega \varphi \, d\mu.$$

The setup used in [14] is to choose $K = L^*$, with $L : \mathcal{H} \to C_0(\Omega)$ being a bounded linear operator. This allows to avoid working with the complicated dual space of $\mathcal{M}(\Omega)$ for some arguments. We note that in this setup regular singular vectors can rather be obtained from

$$(8.15) \qquad \lambda LL^* \mu \in \partial J(\mu) \,.$$

In [30] this problem was analyzed in a compressed sensing setting, when K consists of a finite number of forward projections (in a polynomial basis), so that L can be written down explicitly.

Bounded operators. Here we will consider two related cases including most practical examples we are aware of (except the Radon integral transform): First of all we analyze integral operators $K_\infty : \mathcal{M}(\Omega) \to L^2(\Sigma)$ of the form

$$(8.16) \qquad (K_\infty u)(x) = \int_\Omega k(x, y) \, d\mu(y)$$

with a continuous kernel k. Then we shall turn to projection operators with M measurements $K_M : \mathcal{M}(\Omega) \to \mathbb{R}^M$ of the form

$$(8.17) \qquad (K_M u) = \left(\int_\Omega k_j(x) \, d\mu(x) \right)_{j=1,\dots,M},$$

assuming again the k_j to be continuous. Notice that K_M can also be thought of as a semidiscretization of the operator K_∞, e.g. by collocation methods, so it is natural to compare at least ground states for those operators.

Our aim is to verify a natural extension of the ℓ^1-case, where the ground state is a vector with a single nonzero entry. The natural analogue in the space of Radon measures is a concentrated measure δ_x for $x \in \Omega$ (we use the notation δ_x due to the relation to the Dirac delta distribution), with

$$(8.18) \qquad \int_\Omega \varphi(y) \, d\delta_x = \varphi(x) \qquad \forall \varphi \in C_0(\Omega) \,.$$

Indeed we can show that ground states are of this form and give a reasonably simple condition on their location.

THEOREM 12. *Let K_∞ and K_M be as above. Then*
- *A ground state μ_0^∞ of K_∞ is given by $\mu_0^\infty = c\delta_z$, with z satisfying*

$$\int_\Omega k(x, z)^2 \, dx \geq \int_\Omega k(x, y)^2 \, dx \qquad \forall \, y \in \Omega \,,$$

and with c fulfilling

$$c = \lambda_0^\infty = \frac{1}{\sqrt{\int_\Omega k(x,z)^2 \ dx}} \ .$$

- *A ground state μ_0^M of K_M is given by $\mu_0^M = c\delta_z$, with z satisfying*

$$\sum_j k_j(z)^2 \geq \sum_j k_j(y)^2 \qquad \forall \ y \in \Omega \, ,$$

so that c fulfills

$$c = \lambda_0^\infty = \frac{1}{\sqrt{\sum_j k_j(z)^2}} \ .$$

Proof. We prove a slightly more general version: for an operator $K : \mathcal{M}(\Omega) \to \mathcal{H}$, $\mu_0 = c\delta_z$ is a ground state if

$$f(y) := \|K\delta_y\|_\mathcal{H}$$

attains a maximum at z. This yields the statements of the theorem as special cases. Define $c = \lambda_0 = \frac{1}{\|K\delta_z\|_\mathcal{H}}$. Then we see that

$$\langle \lambda K^* K \mu_0, \mu_0 \rangle_{\mathcal{M}(\Omega)} = c = J(\mu_0).$$

Moreover, for any $\mu \in \mathcal{M}(\Omega)$ we have

$$\langle \lambda K^* K \mu_0, \mu \rangle_{\mathcal{M}(\Omega)} = \lambda \langle K\delta_z, K\mu \rangle_\mathcal{H} \leq \lambda \|K\mu\|_\mathcal{H} \, .$$

We thus verify that $\lambda^2 \|K\mu\|_\mathcal{H}^2 \leq J(\mu)^2$, or equivalently

$$\|K\mu\|_\mathcal{H}^2 \leq J(\mu)^2 \|K\delta_z\|_\mathcal{H}^2$$

holds. First of all we discover

$$\|K\mu\|_\mathcal{H}^2 = \langle \mu, K^* K \mu \rangle_{\mathcal{M}(\Omega)} \leq J(\mu) \|K^* K \mu\|_\infty \, .$$

For $y \in \Omega$ we further estimate

$$|(K^* K \mu)(y)| = |\langle K\delta_y, K\mu \rangle_\mathcal{H}| \leq \|K\delta_y\|_\mathcal{H} \ \|K\mu\|_\mathcal{H} \, ,$$

and thus obtain

$$\|K^* K \mu\|_\infty = \sup_y |(K^* K \mu)(y)| \leq \sup_y f(y) \|K\mu\|_\mathcal{H} \, ,$$

which finally yields the assertion. $\square$

An Unbounded operator. As a simple sketch of the ROF model in three dimensions (when embedding of BV into L^2 fails), we study the case of K^*K being the inverse Laplacian, i.e. $Ku = v$, with $v \in H_0^1(\Omega)$ solving $-\Delta v = u$ if $u \in \mathcal{M}(\Omega) \cap H^{-1}(\Omega)$, and for Ω denoting the unit ball in $\mathbb{R}^3$. Since a Dirac delta distribution is not an element of $H^{-1}(\Omega)$ in spatial dimension three, we cannot obtain such measures as singular vectors, hence the latter can at most be concentrated on manifolds with higher dimension.

The equation for the singular vector becomes

$$\lambda\mu = -\Delta p, \qquad p \in \partial J(\mu)$$

and we can look for radially symmetric solutions $\mu = M(r)$, $p = P(r)$. Hence

$$\lambda r^2 M = -\partial_r(r^2\partial_r P)$$

with the additional condition $P(1) = 0$. Now a canoncial measure concentrated on a codimension one manifold is the one on a sphere with radius $R \in (0,1)$, which corresponds to M being a concentrated measure in $r = R$, i.e. $M = c\delta_R$ with c to be determined from the normalization condition

$$\int_0^1 P(r)^2 r^2 \ dr = 1\,.$$

Then P can be computed as

(8.19)
$$P(r) = \begin{cases} \lambda c \frac{R^2}{r}(1-r) & \text{if } r \geq R \\ \lambda c R(1-R) & \text{if } r < R \end{cases}.$$

Now P attains a maximum at $r = R$, and we need to choose λ such that $P(R) = 1$ holds, which yields $\lambda = \frac{1}{cR(1-R)}$. Hence, we conclude that a measure concentrated on a sphere of radius R is a singular vector, the smallest singular value in this class is obtained for $R = \frac{1}{2}$. Note that $\lambda \to \infty$ for $R \to 1$ or $R \to 0$, which confirms that neither a concentrated measure in the origin nor a measure concentrated on $\partial\Omega$ is a singular vector.

8.4. Sparsity and variants. As mentioned above sparsity-enforcing regularizations played an important role in image analysis and inverse problems in the last years. The usual setup in ℓ^1-Regularization is $\mathcal{U} = \ell^1(\mathbb{R}^n)$ and $\mathcal{H} = \ell^2(\mathbb{R}^m)$, the forward operator can thus be identified with a matrix $K \in \mathbb{R}^{m \times n}$. The proof of Theorem 12 immediately implies that a ground state is given by γe_i, where $i \in \{1, \ldots, n\}$ is the index of a row with maximal Euclidean norm, i.e. $\|Ke_i\|_2 \geq \|Ke_j\|_2$ for any $j \in \{1, \ldots, n\}$. In the following we discuss two other relevant examples related to sparsity and their ground states, respectively singular vectors.

Low rank. In order to compute matrix-valued solutions of low rank, the nuclear norm has been considered in various papers recently, respectively shown to be an exact relaxation of minimal rank problems in some cases (cf. [37, 55]). In this case $\mathcal{U} = \mathbb{R}^{m \times n}$ and $\mathcal{H} = \mathbb{R}^M$ with

(8.20)
$$J(u) = \sum_{j=1}^{\min\{m,n\}} \sigma_j(u)\,,$$

for which σ_j are the (classical) singular values of the matrix u, and with $M \ll (m \times n)$ denoting the number of known entries of u.

It seems natural that a ground state is of rank one. To see this, we take an arbitrary matrix u with singular value decomposition

$$u = \sum_{j=1}^{\min\{m,n\}} \sigma_j U_j V_j^T \, .$$

Then we have

$$\|Ku\|_{\mathcal{H}} = \left\| \sum_{j=1}^{\min\{m,n\}} \sigma_j KU_j V_j^T \right\|_{\mathcal{H}} \leq \sum_{j=1}^{\min\{m,n\}} \sigma_j \|KU_j V_j^T\|_{\mathcal{H}} \leq \max \|KU_j V_j^T\|_{\mathcal{H}} J(u).$$

Equality is obtained if u is a rank-one matrix. Thus, we conclude that the ground state is of rank one and obviously it is a multiple of UV^T, where U and V maximize $\|KUV^T\|_{\mathcal{H}}$ among all orthogonal matrices.

Joint sparsity. In some applications it is more reasonable that few groups of variables have nonzero entries instead of just a few single variables being nonzero. This is modeled by so-called joint sparsity or group lasso approaches (cf. [67, 72, 51]), the most prominent example being

$$(8.21) \qquad J(u) = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{m_i} u_{ij}^2},$$

in $\mathcal{U} = \mathbb{R}^N$ and $\mathcal{H} = \mathbb{R}^M$, where $N = \sum_{i=1}^{n} m_i$. Since the goal is to obtain group sparsity, we expect solutions such that $u_{i\cdot}$ vanishes for most indices i. In particular one expects the ground state such that $u_{i\cdot}$ is a nonzero vector for only a single index i.

In order to characterize the ground state we introduce the matrices $A_i \in \mathbb{R}^{M \times m_i}$ representing the linear operator K restricted to elements u supported in the index set $\{i\} \times \{1, \ldots, m_i\}$. We shall also use the notation $U_i = (u_{ij})_{j=1,\ldots,m_i} \in \mathbb{R}^n$. With those we can write

$$J(u) = \sum_{i=1}^{n} \|U_i\|_{\ell^2(\mathbb{R}^{m_i})}, \quad \|Ku\|_{\ell^2(\mathbb{R}^M)} = \left\| \sum_{i=1}^{n} A_i U_i \right\|_{\ell^2(\mathbb{R}^M)} .$$

By the triangle inequality
(8.22)

$$\|Ku\|_{\ell^2(\mathbb{R}^M)} \leq \sum_{i=1}^{n} \|A_i U_i\|_{\ell^2(\mathbb{R}^M)} \leq \sum_{i=1}^{n} \sigma_i^{\max} \|U_i\|_{\ell^2(\mathbb{R}^{m_i})} \leq J(u) \max_{i \in \{1,\ldots,n\}} \sigma_i^{\max},$$

for which $\sigma_i^{\max}$ is the largest singular value of A_i. Equality is obtained if U_k is the singular vector corresponding to singular value $\sigma_k^{\max}$ with

$$\sigma_k^{\max} \geq \sigma_i^{\max} \qquad \forall i \neq k$$

and all other U_i equal zero. Thus, there are ground states with only one row different from zero, which perfectly corresponds to the motivation of group sparsity. The pattern of the nonzero row is a classical singular vector of the restricted matrix A_i.

8.5. Infimal-convolution regularization. Due to deficiencies of standard regularization functionals, constructions like infimal convolution of multiple regularization functionals have been considered recently (cf. [28, 7, 8, 65, 61, 13]). The infimal convolution (inf-convolution) of two convex functionals J_1 and J_2 is defined as

$$(8.23) \qquad J(u) := \inf_{v \in \mathcal{U}} (J_1(v) + J_2(u - v)),$$

and appears to be a good way to combine the advantages of different regularization functionals. Ideally one would hope that the inf-convolution of J_1 and J_2 can lead to exact reconstruction of all solutions that are reconstructed exactly with J_1 or J_2. Since the related variational problem can be reformulated as

$$(8.24) \qquad (v, w) = \arg\min_{v,w} \left\{ \frac{1}{2} \| K(v + w) - f \|^2 + \alpha (J_1(v) + J_2(w)) \right\},$$

we can directly consider singular vectors in the product space for $u = (v, w)$, which are characterized by

$$(8.25) \qquad \lambda K^* K (v + w) = p_1 = p_2, \quad p_1 \in \partial J_1(v), \; p_2 \in \partial J_2(w).$$

In general we cannot expect that singular vectors of J_1 or J_2 are again singular vectors of the inf-convolution. The simplest case would be a singular vector v

$$\lambda K^* K v = p_1 \in \partial J_1(v), \qquad w = 0.$$

Then we need that $p_1 \in \partial J_2(0)$, which is difficult to achieve for general combinations. However, the construction works at least for the ground state of one-homogeneous functionals J_1 and J_2. Let v_0 be the ground state of J_1 and w_0 be the one of J_2. Moreover we assume that $J_1(v_0) \le J_2(w_0)$. Then we can estimate

$$J_1(v) + J_2(w) \ge J_1(v_0)\|Kv\| + J_2(w_0)\|Kw\| \ge J_1(v_0)\|K(v + w)\|.$$

Equality is achieved if $w = 0$ and $v = v_0$, hence the ground state of J_1 is also a ground state of J. Note that for $J_2(w_0) > J_1(v_0)$ we may conclude that w_0 is not a ground state, potentially not even a singular vector. Since such inequalities depend on the scaling of J_1 and J_2 this suggests that one should use a scaling such that the smallest singular values are equal.

9. Conclusions and open problems. In this paper we have generalized the notion of singular values and singular vectors to nonlinear regularization methods in Banach spaces and demonstrated their usefulness in the analysis. In particular we have derived results on the bias of variational methods and scale estimates, which have a particular geometric interpretation in the case of total variation denoising. Moreover, we have shown that singular vectors are the solutions that can be reconstructed exactly (up to a multiplicative constant) by variational regularization techniques.

A major open problem is to obtain a constructive approach for computing singular values and singular vectors, or at least ground states, of arbitrary problems either analytically or numerically. A computational approach for similar problems was already discussed in [40], as well as for similar problems with quadratic constraints in [6, 48, 50, 69, 70]. Our computational experiments indicate that such approaches can indeed compute singular vectors, however they do not converge robustly to the ground state and it is difficult to control to which singular vector the method will converge.

REFERENCES

[1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] M. AGUEH, *A variational approach to uniqueness of ground states for certain quasilinear PDEs*, Adv. Calc. Var., 3 (2010), pp. 233–261.

[3] M. AGUEH AND G. CARLIER, *A class of total variation minimization problems on the whole space*, Nonlinear Analysis: Theor. Meth. Appl., 70 (2009), pp. 2356–2365.

[4] F. ALTER, V. CASELLES, AND A. CHAMBOLLE, *A characterization of convex calibrable sets in $\mathbb{R}^n$*, Mathematische Annalen, 332 (2005), pp. 329–366.

[5] F. ALTER, V. CASELLES, AND A. CHAMBOLLE, *Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow*, Interfaces and Free Boundaries, 7 (2005), pp. 29–53.

[6] T. ARIAS, A. EDELMAN, AND S. SMITH, *The geometry of algorithms with orthogonality constraints*, SIAM J Matrix Anal. Appl., 20 (1998), pp. 303–353.

[7] J. F. AUJOL, G. AUBERT, L. BLANC-FÉRAUD, AND A. CHAMBOLLE, *Image decomposition into a bounded variation component and an oscillating component*, J. Math. Imaging Vision, 22 (2005), pp. 71–88.

[8] J. F. AUJOL AND A. CHAMBOLLE, *Dual norms and image decomposition models*, J. Computer Vision, 63 (2005), pp. 85–104.

[9] G. BELLETTINI, V. CASELLES, AND M. NOVAGA, *The total variation flow in $\mathbb{R}^n$*, J. Diff. Equations, 184 (2002), pp. 475–525.

[10] M. BENNING, *Singular Regularization of Inverse Problems: Bregman Distances and their Applications to Variational Frameworks with Singular Regularization Energies*, PhD thesis, WWU Münster, 2011.

[11] M. BENNING AND M. BURGER, *Error estimates for general fidelities*, Electronic Transactions on Numerical Analysis, 38 (2011), pp. 44–68.

[12] H. BERESTYCKI AND P. L. LIONS, *Existence of stationary states in nonlinear scalar field equations*, in: C.Bardos, D.Bessis, eds., Bifurcation phenomena in mathematical physics and related topics, Dordrecht, Boston, London: 269–292, 1976.

[13] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 492–526.

[14] K. BREDIES AND H. K. PIKKARAINEN, *Inverse problems in spaces of measures*, SFB-Report No. 2010-037, University Graz, November 2010.

[15] A. BRIANI, A. CHAMBOLLE, M. NOVAGA, AND G. ORLANDI, *On the gradient flow of a one-homogeneous functional*, Confluentes Mathematici, 2012, to appear.

[16] F. BROWDER, *Variational methods for nonlinear elliptic eigenvalue problems*, Bull. Amer. Math. Soc., 71 (1965), pp. 176-183.

[17] T. BUEHLER AND M. HEIN, *Spectral Clustering based on the graph p-Laplacian*, In: L. Bottou, M.Littman, eds., Proceedings of the 26th International Conference on Machine Learning (ICML 2009), 81–88, Omnipress, 2009.

[18] M. BURGER, K. FRICK, S. OSHER, AND O. SCHERZER, *Inverse total variation flow*, SIAM Multiscale Model. Simul., 6 (2007), pp. 366–395.

[19] M. BURGER, G. GILBOA, S. OSHER, AND J. XU, *Nonlinear inverse scale space methods*, Comm. Math. Sci., 4 (2006), pp. 179–212.

[20] M. BURGER AND M. MÖLLER, *Multiscale Methods for polyhedral regularizations*, CAM Report 11-74, UCLA, 2011, and submitted.

[21] M. BURGER, M. MÖLLER, M. BENNING, AND S. OSHER, *An adaptive inverse scale space method for compressed sensing*, Math. Comp., 82:281 (2012), pp. 269–299.

[22] M. BURGER AND S. OSHER, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[23] M. BURGER, E. RESMERITA, AND L. HE, *Error estimation for bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135.

[24] J. F.CAI, B. DONG, S. OSHER, AND Z. SHEN, *Image restoration: total variation; wavelet frames; and beyond*, J. Amer. Math. Soc., 25 (2012), pp. 1033–1089.

[25] E. J. CANDES AND T. TAO, *Decoding by linear programming*, IEEE Transactions on Information Theory, 51 (2005), pp. 4203–4215.

[26] E. J. Candes and T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Transactions on Information Theory, 52 (2006), pp. 5406–5425.

[27] V. Caselles, A. Chambolle, S. Moll, and M. Novaga, *A characterization of convex calibrable sets in $\mathbb{R}^n$ with respect to anisotropic norms*, Annales IHP - Analyse Nonlineaire, 25 (2008).

[28] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer.e Math., 76 (1997), pp. 167–188.

[29] B. Dacorogna, W. Gangbo, and N. Suba, *Sur une generalisation de l'inegalite de Wirtinger (A generalization of Wirtinger's inequality)*, Ann. Inst. H. Poincare Anal. Non Lineaire, 9 (1992), pp. 29–50.

[30] Y. De Castro and F. Gamboa, *Exact reconstruction using Beurling minimal extrapolation*, J. Math. Anal. Appl., 2012, to appear.

[31] J. Dolbeault, M. J. Esteban, and M. Loss, *Symmetry of extremals of functional inequalities via spectral estimates for linear operators*, J. Math. Phys., 53 (2012), 095204.

[32] D. L. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.

[33] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via l-minimization*, PNAS, 100 (2003), pp. 2197–2202.

[34] P. P. B. Eggermont, *Maximum entropy regularization for Fredholm integral equations of the first kind*, SIAM J. Math. Anal., 24 (1993), pp. 1557–1576.

[35] I. Ekeland and R. Temam, *Convex analysis and variational problems*, Corrected Reprint Edition. SIAM, Philadelphia, 1999.

[36] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.

[37] M. Fazel, *Matrix Rank Minimization with Applications*, PhD thesis, Stanford University, 2002.

[38] G. Franzina and P. D. Lamberti, *Existence and uniqueness for a p-laplacian nonlinear eigenvalue problem*, Electron. J. Differential Equations, 26 (2010), pp. 1-10.

[39] K. Fujimoto, *What are singular values of nonlinear operators?*, 43rd IEEE Conference on Decision and Control, 2 (2004), pp. 1623–1628.

[40] M. Hein and T. Buehler, *An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA*, Advances in Neural Information Processing Systems 23 (NIPS 2010), 847–855, 2010.

[41] A. Henrot, *Extremum Problems for Eigenvalues of Elliptic Operators*, Frontiers in Mathematics, Birkhäuser, Basel, 2006.

[42] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Mathematical Modelling: Theory and Applications Volume 23. Springer, Netherlands, 2009.

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

[44] U. S. Kamilov, E. Bostan, and M. Unser, *Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising*, IEEE Signal Process. Letters, 19 (2012), pp. :187–190.

[45] B. Kawohl, *Symmetry results for functions yielding best constants in Sobolev-type inequalities*, Discrete Contin. Dyn. Syst., 6 (2000), pp. 683–690.

[46] B. Kawohl and V. Fridman, *Isoperimetric estimates for the first eigenvalue of the p-Laplace operator and the Cheeger constant*, Math. Univ. Carolin., 44 (2003), pp. 659–667.

[47] B. Kawohl, M. Lucia, and S. Prashanth, *Simplicity of the principal eigenvalue for indefinite quasilinear problems*, Adv. Differ. Equ., 12 (2007), pp. 407–434.

[48] R. Lai and S. Osher, *A splitting method for orthogonality constrained problems*, CAM-Report 12-39, UCLA, 2012, and submitted.

[49] D. A. Lorenz, *Convergence rates and source conditions for tikhonov regularization with sparsity constraints*, J. Inverse Ill-Posed Problems, 16 (2008), pp. 463–478.

[50] J. H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Process, 50 (2002), pp. 635–650.

[51] L. Meier, S. van de Geer, and P. Bühlmann, *The group lasso for logistic regression*, J. Royal Stat. Soc. B, 70 (2008), pp. 53–71.

[52] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations*, The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures. American Mathematical Society, Providence, 2001.

[53] M. Möller, *Multiscale Methods for (Generalized) Sparse Recovery and Applications in High Dimensional Imaging*, PhD thesis, WWU Münster, 2011.

[54] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, SIAM Multiscale Model. Simul., 4 (2005), pp. 460–489.

[55] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization*, SIAM Review, 52 (2010), pp. 471–501.

[56] E. Resmerita, *Regularization of ill-posed problems in banach spaces: Convergence rates*, Inverse Problems, 21 (2005), pp. 1303–1314.

[57] E. Resmerita and O. Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems, 22 (2006), pp. 801–814.

[58] R. T. Rockafellar, *Convex Analysis (Princeton Mathematical Series)*, Princeton University Press, 1997.

[59] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[60] O. Scherzer and C. Groetsch, *Inverse scale space theory for inverse problems*, in M. Kerckhove, editor, Scale-space and morphology in computer vision, Proc. 3rd Int. Conf. Scale-space, 317–325, Berlin, 2001. Springer.

[61] S. Setzer, G. Steidl, and T. Teuber, *Infimal convolution regularizations with discrete ℓ^1-type functionals*, Comm. Math. Sci., 9 (2011), pp. 797–872.

[62] D. Showalter, *Representation and computation of the pseudoinverse*, Proc. Amer. Math. Soc., 18 (1967), pp. 584–586.

[63] G. Steidl, J. Weickert, T. Brox, P. Mrazek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs*, SIAM J. Numer. Analysis, 42 (2004), pp. 686–713.

[64] G. W.Stewart, *On the early history of the singular value decomposition*, SIAM Review, 35 (1993), pp. 551–566.

[65] J. L. Starck, M. Elad, and D. L. Donoho, *Image decomposition: separation from texture from piecewise smooth content*, IEEE Trans. Image Proc., 14 (2005), pp. 1570–1582.

[66] A. Szlam and X. Bresson, *Total variation and Cheeger cuts*, Proc. of the 27th International Conference on Machine Learning (ICML), 1039–1046, Omnipress, 2010.

[67] M. B. Wakin, M. F. Duarte, S. Sarvotham, D.Baron, and R. G. Baraniuk, *Recovery of jointly sparse signals from few random projections*, Workshop on Neural Information Processing Systems (NIPS) 2005, Vancouver, 1435–1442, 2005.

[68] M. I. Weinstein, *Nonlinear Schrödinger equation and sharp interpolation estimates*, Comm. Math. Phys., 87 (1983), pp. 567–576.

[69] Z. Wen and W. Yin, *A feasible method for optimization with orthogonality constraints*, CAM-Report 10–77, UCLA, 2010.

[70] I. Yamada and T. Ezaki, *An orthogonal matrix optimization by dual cayley parametrization technique*, in: 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), 2003.

[71] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for ℓ^1-minimization with applications to compressed sensing*, SIAM J. Imaging Sci., 1 (2008), pp. 143–168.

[72] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. Royal Stat. Soc. B, 68 (2007), pp. 49–67.

[73] F. Yang and Z. Wei, *Generalized Euler identity for subdifferentials of homogeneous functions and applications*, Mathematical Analysis and Applications, 337 (2008), pp. 516-523.

[74] E. Zeidler, *Nonlinear Functional Analysis and its Applications III: Variational Methods and Optimization*, Springer, 1984.