Noname manuscript No.
(will be inserted by the editor)

Nonlinear Spectral Analysis via One-homogeneous Functionals - Overview and Future Prospects

Guy Gilboa · Michael Moeller · Martin Burger

Received: date / Accepted: date

Abstract We present in this paper the motivation and theory of nonlinear spectral representations, based on convex regularizing functionals. Some comparisons and analogies are drawn to the fields of signal processing, harmonic analysis and sparse representations. The basic approach, main results and initial applications are shown. A discussion of open problems and future directions concludes this work.

Keywords Nonlinear spectral representations · One-homogeneous functionals · Total variation · Nonlinear eigenvalue problem · Image decomposition.

1 Introduction

Nonlinear variational methods have provided very powerful tools in the design and analysis of image processing and computer vision algorithms in recent decades. In parallel, methods based on harmonic analysis, dictionary learning and sparse-representations, as well as spectral analysis of linear operators (such as the graph-Laplacian) have shown tremendous advances in processing highly complex signals such as natural images, 3D data and speech. Recent studies now suggest variational methods can also be analyzed and understood through a nonlinear generalization of eigenvalue analysis, referred to as nonlinear spectral methods. Most of the current knowledge is focused on one-homogeneous functionals, which will be the focus of this paper.

Electrical Engineering Department, Technion IIT

Department of Computer Science, Technische Universität München

Institute for Computational and Applied Mathematics, University of Münster

Address(es) of author(s) should be given

The motivation and interpretation of classical linear filtering strategies is closely linked to the eigendecomposition of linear operators. In this manuscript we will show that one can define a nonlinear spectral decomposition framework based on the gradient flow with respect to arbitrary convex 1-homogeneous functionals and obtain a remarkable number of analogies to linear filtering techniques. To closely link the proposed nonlinear spectral decomposition framework to the linear one, let us summarize earlier studies concerning the use of nonlinear eigenfunctions in the context of variational methods.

One notion which is very important is the concept of nonlinear eigenfunctions induced by convex functionals. Throughout this paper we will consider convex functionals J either in a function space setting $J : \mathcal{X} \to \mathbb{R}$ for a Banach space $\mathcal{X}$ embedded into L^2 or - in a discrete setting - as a function $J : \mathbb{R}^n \to \mathbb{R}$. We will specify the considered setting at those points where it is crucial. In either setting, we refer to u as an *eigenfunction* of J if it admits the following eigenvalue problem

$$\lambda u \in \partial J(u), \quad \text{and } \|u\|_2 = 1, \tag{1}$$

where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue, and $\partial J(u) = \{p \in \mathcal{X}^* \mid J(v) - J(u) - \langle p, v - u\rangle \geq 0\}$ the subdifferential of J at u.

The analysis of eigenfunctions related to non-quadratic convex functionals was mainly concerned with the total variation (TV) regularization, given by

$$J_{TV}(u) = \sup_{\substack{z \in C_c^\infty, \\ \|z\|_{L^\infty(\Omega)} \leq 1}} \int_\Omega u(x) \operatorname{div} z(x)\, dx,$$

on a domain Ω, which for smooth functions u reduces to $J_{TV}(u) = \int_\Omega |\nabla u(x)|\, dx$.

In the analysis of the variational TV denoising, i.e. the ROF model from [51], Meyer [41] has shown an explicit solution for the case of a disk (an eigenfunction of TV), quantifying explicitly the loss of contrast and advocating the use of $TV - G$ regularization. Within the extensive studies of the TV-flow [1,2,7,52,15,6,29] eigenfunctions of TV (referred to as *calibrable sets*) were analyzed and explicit solutions were given for several cases of eigenfunction spatial settings. In [17] an explicit solution of a disk for the inverse-scale-space flow is presented, showing its instantaneous appearance at a precise time point related to its radius and height.

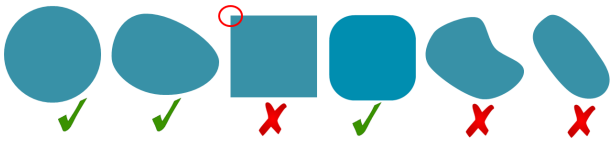


Fig. 1: Example of valid and non-valid shapes as nonlinear eigenfunctions with respect to the TV functional. Smooth enough convex sets, which admit condition (2) (marked with a green check), are solutions of the eigenvalue problem (1). Shapes with either too high curvature, not convex or too elongated (marked with red X) are not valid.

A highly notable contribution in [7] is the precise geometric definition of convex sets which are eigenfunctions. Let χ_C, be a characteristic function of the set $C \subset \mathbb{R}^2$, then it admits (1), with J the TV functional J_{TV}, if

$$\operatorname*{ess\,sup}_{q \in \partial C} \kappa(q) \leq \frac{P(C)}{|C|}, \tag{2}$$

where C is convex, $\partial C \in C^{1,1}$, $P(C)$ is the perimeter of C, $|C|$ is the area of C and κ is the curvature. In this case the eigenvalue is $\lambda = \frac{P(C)}{|C|}$. See Fig. 1 for some examples. Theorem 6 in [7] furthermore classified all possible eigenfunctions arising from the indicator function of a set. Such eigenfunctions must arise from a union of sets C with the above properties, all having the same eigenvalue $\frac{P(C)}{|C|}$ and additionally being sufficiently far apart.

In [44,9,48,46] eigenfunctions related to the total-generalized-variation (TGV) [11] and the infimal convolution total variation (ICTV) functional [20] are analyzed in the case of a bounded domain and their different reconstruction properties on particular eigenfunctions of the TGV are demonstrated theoretically as well as numerically.

Examples of certain eigenfunctions for different extensions of the TV to color images are given in [26].

In [52] Steidl et al. have shown the close relations, and equivalence in a 1D discrete setting, of the Haar wavelets to both TV regularization and TV flow. This was later developed for a 2D setting in [57]. The connection between Haar wavelets and TV methods in 1D was made more precise in [10], who indeed showed that the Haar wavelet basis is an orthogonal basis of eigenfunctions of the total variation (with appropriate definition at the domain boundary) - in this case the Rayleigh principle holds for the whole basis. In the field of morphological signal processing, nonlinear transforms were introduced in [25,40].

2 Spectral Representations

2.1 Scale Space Representation

We will use the scale space evolution, which is straightforward, as the canonical case of spectral representation. Consider a convex (absolutely) one-homogeneous functional J, i.e. a functional for which $J(\lambda u) = |\lambda| J(u)$ holds for all $\lambda \in \mathbb{R}$.

The *scale space* or *gradient flow* is

$$\partial_t u(t) = -p(t), \quad p(t) \in \partial J(u(t)), \; u(0) = f. \tag{3}$$

We refer the reader to [56,3] for an overview over scale space techniques.

A spectral representation based on the gradient flow formulation (3) was the first work towards defining a nonlinear spectral decomposition and has been conducted in [30,31] for the case of J being the TV regularization. In our conference paper [14], we extended this notion to general 1-homogeneous functionals by observing that the solution of the gradient flow can be computed explicitly for any 1-homogeneous J in the case of f being an eigenfunction. For $\lambda f \in \partial J(f)$, the solution to (3) is given by

$$u(t) = \begin{cases} (1 - t\lambda) f & \text{for } t \leq \frac{1}{\lambda}, \\ 0 & \text{else.} \end{cases} \tag{4}$$

Note that in linear spectral transformations such as Fourier or Wavelet based approaches, the input data being an eigenfunction leads to the energy of the spectral representation being concentrated at a single wavelength. To preserve this behavior for nonlinear spectral representations, the wavelength decomposition of the input data f is defined by

$$\phi(t) = t \partial_{tt} u(t). \tag{5}$$

Note that due to the piecewise linear behavior in (4), the wavelength representation of an eigenfunction f becomes $\phi(t) = \delta(t - \frac{1}{\lambda}) f$, where δ denotes a Dirac

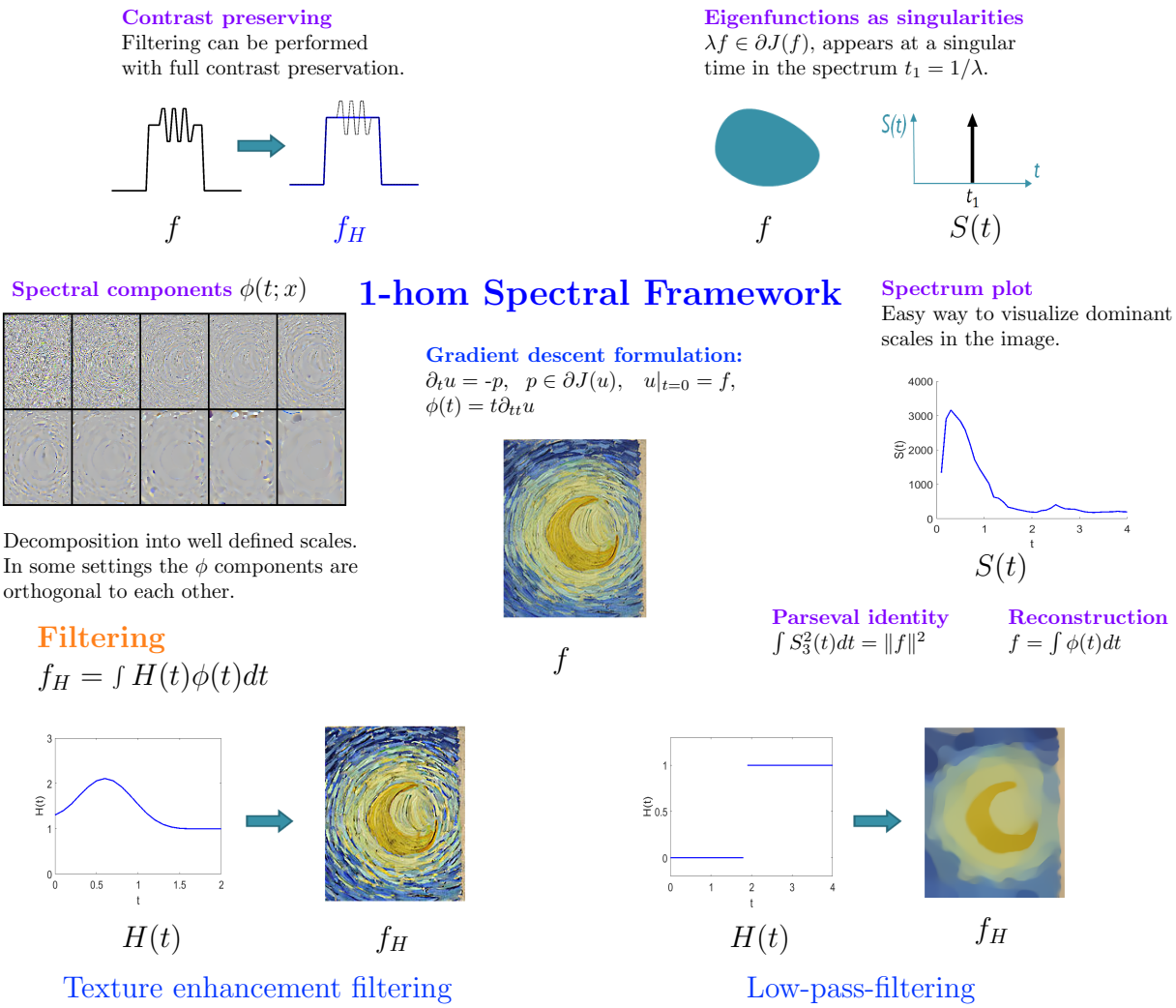


Fig. 2: Overview of the spectral framework.

delta distribution. The name *wavelength* decomposition is natural because for $\lambda f \in \partial J(f)$, $\|f\| = 1$, one readily shows that $\lambda = J(f)$, which means that the eigenvalue λ is corresponding to a generalized frequency. In analogy to the linear case, the inverse relation of a peak in ϕ appearing at $t = \frac{1}{\lambda}$ motivates the interpretation as a wavelength, as discussed in more details in the following section.

For arbitrary input data f one can reconstruct the original image by:

$$f(x) = \int_0^\infty \phi(t;x)dt. \tag{6}$$

Given a *transfer function* $H(t) \in \mathbb{R}$, image filtering can be performed by

$$f^H(x) := \int_0^\infty H(t)\phi(t;x)dt. \tag{7}$$

Figure 2 gives an overview of the spectral decomposition framework including two examples of applying transfer or *filter functions* for texture enhancement and low pass filtering. Some of the terms and notations (such as low-pass-filter and spectrum S) are explained hereafter in Section 3.

We'd like to point out that equation (5) is an informal definition of the wavelength decomposition which is supposed to illustrate the general idea of nonlinear spectral decompositions. We refer the reader to our more theoretical work [16] in which we give sense to (5) for one-homogeneous J in the spatially discrete setting as elements in $(W^{1,1}_{loc}(\mathbb{R}^+, \mathbb{R}^n))^*$. The discrete case also permits to show a sufficiently rapid decrease of $\partial_t u(t)$ for the reconstruction (7) to hold for all absolutely one-homogeneous regularization functions J. For the sake of simplicity and readability we will skip the technical

details of the underlying approach in most parts of this paper.

3 Signal processing analogy

Up until very recently, nonlinear filtering approaches such as (3) or related variational methods have been treated independently of the classical linear point of view of changing the representation of the input data, filtering the resulting representation and inverting the transform. In [30,31] it was proposed to use (3) in the case of J being the total variation to define a TV spectral representation of images that allows to extend the idea of filtering approaches from the linear to the nonlinear case. This was later generalized to one-homogeneous functionals in [14].

Classical Fourier filtering has some very convenient properties for analyzing and processing signals:

1. Processing is performed in the transform (frequency) domain by simple attenuation or amplification of desired frequencies
2. A straightforward way to visualize the frequency activity of a signal is through its spectrum plot. The spectral energy is preserved in the frequency domain, through the well-known Parseval's identity.
3. The spectral decomposition corresponds to the coefficients representing the input signal in a new orthonormal basis.
4. Both, transform and inverse-transform are linear operations.

In the nonlinear setting the first two characteristics are mostly preserved. Orthogonality could so far only be obtained for the particular (discrete) case of $J(u) = \|Vu\|_1$ with VV^* being diagonally dominant, [16]. Further orthogonality statements are still an open issue. Finally, linearity is certainly lost. In essence we obtain a *nonlinear* forward transform and a *linear* inverse transform. Thus following the nonlinear decomposition, filtering can be performed easily.

In addition, we gain edge-preservation and new scale features, which, unlike sines and cosines, are data-driven and are therefore highly adapted to the image. Thus the filtering has far less tendency to create oscillations and artifacts.

Let us first derive the relation between Fourier and the eigenvalue problem (1). For

$$J(u) = \frac{1}{2}\int |\nabla u(x)|^2 dx,$$

we get $-\Delta u \in \partial J(u)$. Thus, with appropriate boundary conditions, the sine waves arising from Fourier analysis are eigenfunctions, in the sense of (1), where, for a frequency ω of the sine waves, we have the relation $\lambda = \omega^2$. Other convex regularizing functionals, such as TV and TGV, can therefore be viewed as natural nonlinear generalizations.

A fundamental concept in linear filtering are *ideal filters* [49]. Such filters either retain or diminish completely frequencies within some range. In a linear time (space) invariant system, a filter is fully determined by its transfer function $H(\omega)$. The filtered response of a signal $f(x)$, with Fourier transform $F(\omega)$, filtered by a filter H is

$$f^H(x) := \mathcal{F}^{-1}\left(H(\omega)\cdot F(\omega)\right),$$

with $\mathcal{F}^{-1}$ the inverse Fourier transform. For example, an ideal low-pass-filter retains all frequencies up to some cutoff frequency ω_c. Its transfer function is thus

$$H(\omega) = \begin{cases} 1 & \text{for } 0 \le \omega \le \omega_c, \\ 0 & \text{else.} \end{cases}$$

Viewing frequencies as eigenvalues in the sense of (1) one can define generalizations of these notions.

3.1 Nonlinear ideal filters

The (ideal) low-pass-filter (LPF) can be defined by Eq. (7) with $H(t) = 1$ for $t \ge t_c$ and 0 otherwise, or

$$LPF_{t_c}(f) := \int_{t_c}^{\infty} \phi(t;x)dt. \tag{8}$$

Its complement, the (ideal) high-pass-filter (HPF), is defined by

$$HPF_{t_c}(f) := \int_{0}^{t_c} \phi(t;x)dt. \tag{9}$$

Similarly, band-(pass/stop)-filters are filters with low and high cut-off scale parameters ($t_1 < t_2$)

$$BPF_{t_1,t_2}(f) := \int_{t_1}^{t_2} \phi(t;x)dt, \tag{10}$$

$$BSF_{t_1,t_2}(f) := \int_{0}^{t_1} \phi(t;x)dt + \int_{t_2}^{\infty} \phi(t;x)dt. \tag{11}$$

For f being a single eigenfunction with eigenvalue $\lambda_0 = \frac{1}{t_0}$ the above definitions coincide with the linear definitions, where the eigenfunction is either completely preserved or completely diminished, depending on the cutoff eigenvalue(s) of the filter.

Again the above point of view is rather illustrative than mathematically precise. As mentioned above, the discrete case allows to identify ϕ with elements of $(W^{1,1}_{loc})^*$ such that only mollified version of the above ideal filters are permissible, see [16] for details.

In Fig. 3 an example of spectral TV processing is shown with the response of the four filters defined above in Eqs. (8) through (11).

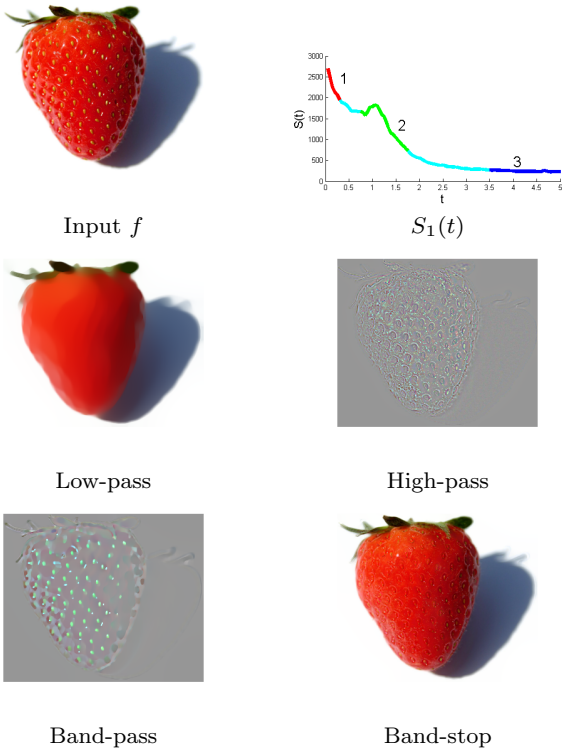


Fig. 3: Total-variation spectral filtering example. The input image (top left) is decomposed into its $\phi(t)$ components, the corresponding spectrum $S_1(t)$, Eq. (12), is on the top right. Integration of the ϕ's over the t domains 1, 2 and 3 (top right) yields high-pass, band-pass and low-pass filters, respectively (chosen manually here). The band-stop filter (bottom right) is the complement integration domain of region 2. Taken from [14].

3.2 Spectral response

As in the linear case, it is very useful to measure in some sense the "activity" at each frequency (scale). This can help identify dominant scales and design better the filtering strategies (either manually or automatically). Moreover, one can obtain a notion of the type of energy which is preserved in the new representation using some analog of Parseval's identity.

In [30,31] a L^1 type spectrum was suggested for the TV spectral framework (without trying to relate to a Parseval rule),

$$S_1(t) := \|\phi(t;x)\|_{L^1(\Omega)} = \int_\Omega |\phi(t;x)|dx. \tag{12}$$

In [14] the following definition was suggested,

$$S_2^2(t) = t^2 \frac{d^2}{dt^2} J(u(t)) = \langle \phi(t), 2tp(t) \rangle. \tag{13}$$

With this definition the following analogue of the Parseval identity was shown:

$$\begin{aligned} \|f\|^2 &= -\int_0^\infty \frac{d}{dt}\|u(t)\|^2 \, dt \\ &= 2\int_0^\infty \langle p(t), u(t) \rangle \, dt \\ &= 2\int_0^\infty J(u(t)) \, dt \\ &= \int_0^\infty S_2^2(t) \, dt. \end{aligned} \tag{14}$$

In [16] a third definition for the spectrum was suggested (which is simpler and admits Parseval),

$$S_3^2(t) = \langle \phi(t), f \rangle. \tag{15}$$

It can be shown that a similar Parseval-type equality holds here

$$\begin{aligned} \|f\|^2 = \langle f, f \rangle = \langle \int_0^\infty \phi(t)dt, f \rangle = \int_0^\infty \langle \phi(t), f \rangle dt \\ = \int_0^\infty S_3^2(t) \, dt. \end{aligned} \tag{16}$$

While (12), (13), and (15) yield the intuition of possible definitions of the spectral decomposition, care has to be taken to obtain mathematically precise definitions. In the discrete case we have $\phi \in (W^{1,1}_{loc}(\mathbb{R}^+, \mathbb{R}^n))^*$, such that S_2^2 and S_3^2 have to be interpreted as elements in $(W^{1,1}_{loc}(\mathbb{R}^+, \mathbb{R}))^*$. In fact, in [16] we showed these two definitions to be equivalent for sufficiently regular subdifferentials of J.

As an overview example, Table 1 summarizes the analogies of the Fourier and total variation based spectral transformation. There remain many open questions of further possible generalizations regarding Fourier-specific results, some of which are listed like the convolution theorem, Fourier duality property (where the transform and inverse transform can be interchanged up to a sign), the existence of phase in Fourier and more.

4 Decomposition into eigenfunctions

Let us consider the discrete case of J being a semi-norm on $\mathbb{R}^n$. As discussed in the section 3, item 3, one of the fundamental interpretations of linear spectral decompositions arises from it being the coefficients for representing the input data in a new basis. This basis is composed of the eigenvectors of the transform. For example, in the case of the cosine transform, the input signal is represented by a linear combination of cosines with increasing frequencies. The corresponding spectral

	TV transform	**Fourier transform**
Transform	Gradient-flow:$^+$ $\phi(t;x) = t\partial_{tt}u$, $u_t \in -\partial J(u)$. $u\|_{t=0} = f$, (more representations in Table 1)	$F(\omega) = \int f(x)e^{-i\omega x}dx$
Inverse transform	$f(x) = \int_0^\infty \phi(t;x)dt + \bar{f}$ $^+$	$f(x) = \frac{1}{(2\pi)^n}\int F(\omega)e^{i\omega x}d\omega$
Eigenfunctions	$\lambda u \in \partial J(u)$ $^+$	$e^{i\omega x}$
Spectrum - amplitude	Three alternatives: $^+$ $S_1(t) = \|\phi(t;x)\|_{L^1}$, $S_2(t)^2 = \langle\phi(t), 2tp(t)\rangle$, $S_3(t)^2 = \langle\phi(t), f\rangle$.	$\|F(\omega)\|$
Translation $f(x-a)$	$\phi(t;x-a)$	$F(\omega)e^{-ia\omega}$
Rotation $f(Rx)$	$\phi(t;Rx)$,	$F(R\omega)$
Contrast change $af(x)$	$\phi(t/a;x)$	$aF(\omega)$
Spatial scaling $f(ax)$	$a\phi(at;ax)$	$\frac{1}{\|a\|}F(\frac{\omega}{a})$
Linearity $f_1(x)+f_2(x)$	Not in general	$F_1(\omega)+F_2(\omega)$
Parseval's identity	$\int \|f(x)\|^2dx = \int S_2(t)^2dt = \int S_3(t)^2dt$ $^+$	$\int \|f(x)\|^2dx = \int \|F(\xi)\|^2d\xi$, $\xi := \frac{\omega}{2\pi}$
Orthogonality	$\langle\phi(t), u(t)\rangle = 0$ $^+$	$\langle e^{i\omega_1 x}, e^{i\omega_2 x}\rangle = 0$, $\omega_1 \neq \omega_2$
Open issues (some)		
Fourier Duality	–	$F(x) \leftrightarrow f(-\xi)$, $\xi := \frac{\omega}{2\pi}$
Convolution	–	$f_1(x) * f_2(x) \leftrightarrow F_1(\omega)F_2(\omega)$
Spectrum - phase	–	$\tan^{-1}(\Im(F(\omega))/\Re(F(\omega)))$

Table 1: Some properties of the TV transform in $\mathbb{R}^n$, compared to the Fourier transform. The symbol $^+$ denotes that the property holds for general 1-homogeneous functionals.

decomposition consists of the coefficients in this representation, which therefore admit an immediate interpretation.

Although the proposed nonlinear spectral decomposition does not immediately correspond to a change of basis anymore, it is interesting to see that the property of representing the input data as a linear combination of (generalized) eigenfunctions can be preserved: In [14] we showed that for the case of $J(u) = \|Vu\|_1$ and V being any orthonormal matrix, the solution of the scale space flow (3) meets

$$Vu(t) = \text{sign}(\zeta)\ \max(|\zeta| - t, 0), \tag{17}$$

where $\zeta = Vf$ are the coefficients for representing f in the orthonormal basis of V. It is interesting to see that the subgradient $p(t)$ in (3) admits a componentwise representation of $Vp(t)$ as

$$(Vp)_i(t) = \begin{cases} \text{sign}(\zeta_i) & \text{if } |\zeta_i| \geq t, \\ 0 & \text{else.} \end{cases} \tag{18}$$

The latter shows that $p(t)$ can be represented by $V^Tq(t)$ for some $q(t)$ which actually meets $q(t) \in \partial\|Vp(t)\|_1$. In a single equation this means $p(t) \in \partial J(p(t))$ and shows that the $p(t)$ arising from the scale space flow are eigenfunctions (up to normalization). Integrating the scale space flow equation (3) from zero to infinity and using that there are only finitely many times at which $p(t)$ changes, one can see that one can indeed represent f as a linear combination of eigenfunctions. Our spectral decomposition $\phi(t)$ then corresponds to the *change* of the eigenfunctions during the (piecewise) dynamics of the flow.

While the case of $J(u) = \|Vu\|_1$ for an orthogonal matrix V is quite specific because it essentially recovers the linear spectral analysis exactly, we have shown in [16] that the property of finitely many times at which $p(t)$ changes can be proved for any polyhedral one-homogeneous regularization. If in addition the subdifferentials are sufficiently regular, e.g. as in the case of $J(u) = \|Vu\|_1$ and VV^* being diagonally dominant, then the property of $p(t)$ being an eigenfunction for all times t remains true. We refer the reader to [16] for more details. The analysis of such properties for more general one-homogeneous regularizations or the extension of these results to a function space setting remain open questions.

5 Explicit TV Eigenfunctions in 1D

Let us consider the function space setting and give an analytic expression of a large set of eigenfunctions of the TV in one dimension with the signal domain being the

entire real line. We will later see that Haar wavelets are a small subset of these, hence eigenfunctions are expected to represent signals more concisely, with much fewer elements.

We give the presentation below in a somewhat informal manner with the goal of rather explaining the main concept of the findings than giving mathematical details.A formal presentation of the theory of TV eigenfunctions can be found e.g. in [7].

The TV functional can be expressed as

$$J_{TV}(u) = \sup_{\substack{z\in C_c^\infty \\ \|z\|_{L^\infty(\Omega)}\leq 1}} \langle u, \operatorname{div} z\rangle. \tag{19}$$

Let z_u be an argument admitting the supremum of (19) then it immediately follows that $\operatorname{div} z_u \in \partial J_{TV}(u)$: in the one homogeneous case we need to show $\langle u, \operatorname{div} z_u\rangle = J_{TV}(u)$ which is given in (19); and in addition that for any v in the space we have $J_{TV}(v) \geq \langle v, p\rangle$, $p \in \partial J(u)$,

$$J_{TV}(v) = \sup_{\|z\|_{L^\infty(\Omega)}\leq 1} \langle v, \operatorname{div} z\rangle \geq \langle v, \operatorname{div} z_u\rangle.$$

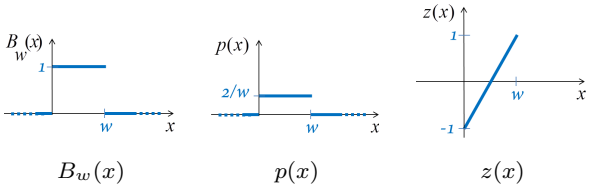
From here on we will refer to z_u simply as z.

To understand better what z stands for, we can check the case of smooth u and perform integration by parts in (19) to have

$$J_{TV}(u) = \langle \nabla u, -z\rangle.$$

Then as z also maximizes $\langle \nabla u, -z\rangle$, we can solve this pointwise, taking into account that $|z(x)| \leq 1$ and that the inner product of a vector is maximized for a vector at the same angle, to have

$$z(x) \begin{cases} = -\frac{\nabla u(x)}{|\nabla u(x)|} & \text{for } \nabla u(x) \neq \mathbf{0}, \\ \in [-1,1] & \nabla u(x) = \mathbf{0}. \end{cases} \tag{20}$$



$B_w(x)$ $p(x)$ $z(x)$

Fig. 4: A classical single peak example of a TV eigenfunction in $\mathbb{R}$, $u(x) = B_w(x)$.

5.1 Single peak

Let us consider the case of one dimensional TV regularization and define the following function of a single unit peak of width w

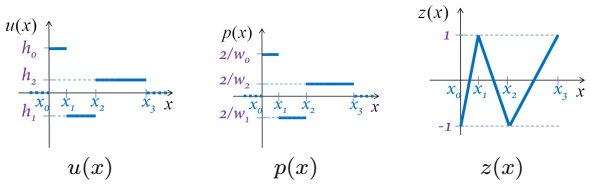
$$B_w(x) = \begin{cases} 1 & \text{for } x \in [0, w), \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Then it is well known (e.g. [41]) that any function of the type $h \cdot B_w(x - x_0)$ is an eigenfunction of TV in $x \in \mathbb{R}$ with eigenvalue $\lambda = \frac{2}{hw}$. Let us illustrate the latter by considering the characterization of the subdifferential (20) and define

$$z(x) = \begin{cases} -1 & \text{for } x \in]-\infty, x_0], \\ \frac{2(x-x_0)}{w} - 1 & \text{for } x \in (x_0, x_0 + w), \\ 1 & \text{for } x \in [x_0 + w, \infty[. \end{cases}$$

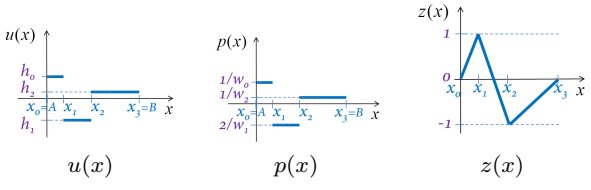
Although this z is clearly not a C_c^∞ function, it was shown in [1,7] that functions $z \in L^\infty(\mathbb{R})$ with $\operatorname{div} z \in L^2(\mathbb{R})$ are sufficient for the characterization of subgradients.

For the above z, on the one hand we have $\partial_x z(x) = \frac{w}{2} B_w(x - x_0)$, and on the other hand z meets equation (20) for $u = \frac{w}{2} B_w(\cdot - x_0)$. Therefore, $\frac{w}{2} B_w(\cdot - x_0) \in \partial J_{TV}(\frac{w}{2} B_w(\cdot - x_0))$, and after normalization with $\|\frac{w}{2} B_w(\cdot - x_0)\|_2 = \frac{w^2}{2}$ we find that $\frac{1}{w} B_w(\cdot - x_0)$ is an eigenfunction with eigenvalue $\lambda = \frac{2}{w}$.



$u(x)$ $p(x)$ $z(x)$

Fig. 5: Illustration of Prop. 1, a TV eigenfunction $u(x)$ in $\mathbb{R}$.



$u(x)$ $p(x)$ $z(x)$

Fig. 6: A TV eigenfunction $u(x)$ in a bounded domain.

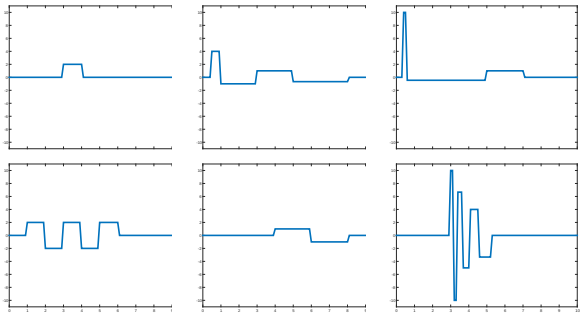


Fig. 7: A few examples of functions meeting $f \in \partial J_{TV}(f)$ in $\mathbb{R}$.(Note that these are piecewise constant functions and the jumps should be understood as discontinuities).

5.2 Set of 1D eigenfunctions

Generalizing this analysis one can construct for any eigenvalue λ an infinite set of piece-wise constant eigenfunctions (with a compact support).

Proposition 1 *Let $-\infty < x_0 < x_1, .. < x_n < \infty$ be a set of $n+1$ points on the real line. Let*

$$u(x) = \sum_{i=0}^{n-1} h_i B_{w_i}(x - x_i), \tag{22}$$

with $B_w(\cdot)$ defined in (21), $w_i = x_{i+1} - x_i$, and

$$h_i = \frac{2(-1)^i}{\lambda w_i}. \tag{23}$$

Then $u(x)$ admits the eigenvalue problem (1) with $J = J_{TV}$.

Proof One can construct the following z in the shape of "zigzag" between -1 and 1 at points x_i,

$$z(x) = (-1)^i \left(\frac{2(x - x_i)}{w_i} - 1 \right), \quad x \in [x_i, x_{i+1}),$$

and $\partial_x z = 0$ otherwise. In a similar manner to the single peak case we get the subgradient element in $\partial J_{TV}(u)$

$$p(x) = \begin{cases} \partial_x z = (-1)^i \frac{2}{w_i}, & x \in [x_i, x_{i+1}) \\ 0, & x \notin [x_0, x_n). \end{cases} \tag{24}$$

This yields $p(x) = \lambda u(x)$.

In Figs. 5, 6, we see examples of u, p and z which construct simple TV eigenfunctions in the unbounded and bounded domains, respectively. In Fig. 7 several different types of eigenfunctions are shown. Compared to wavelets, they also oscillate (with mean zero value) but their variability is significantly larger.

5.2.1 The bounded domain case

The unbounded domain is easier to analyze in some cases, however in practice we can implement only signals with a bounded domain. We therefore give the formulation for $u \in \Omega = [A, B) \subset \mathbb{R}$. Requiring $J_{TV}(1) = \langle p, 1 \rangle = 0$ and $\langle u, p \rangle = \langle -\nabla u, z \rangle$ leads to the boundary condition $z|_{\partial\Omega} = 0$.

Thus, on the boundaries we have $z = 0$ with half the slope of the unbounded case (see Fig. 6), all other derivations are the same. Setting $x_0 = A$, $x_n = B$ we get the solution of (22) with h_i defined slightly differently as

$$h_i = \frac{2a_i(-1)^i}{\lambda w_i}, \tag{25}$$

where $a_i = \frac{1}{2}$ for $i \in \{1, n-1\}$ and $a_i = 1$ otherwise. See a numerical convergence of λp to u in Fig. 12.

Remark: Note that if u is an eigenfunction so is $-u$ so the formulas above are all valid also with the opposite sign.

6 Wavelets and Hard Thresholding

As we have seen in section 4, the gradient flow with respect to regularizations of the form $J(u) = \|Vu\|_1$ has a closed form solution for any orthogonal matrix V. From equation (18) one deduces that

$$\phi(t) = \sum_i \zeta_i \delta(t - |\zeta_i|) v_i,$$

where v_i are the rows of the matrix V, and $\zeta_i = (Vf)_i$. In particular, using definition (15) we obtain

$$(S_3(t))^2 = \sum_i (\zeta_i)^2 \delta(t - |\zeta_i|).$$

Peaks in the spectrum therefore occur ordered by the magnitude of the coefficients $\zeta_i = (Vf)_i$: The smaller $|\zeta_i|$ the earlier it appears in the wavelength representation ϕ. Hence, applying an ideal low pass filter (8) on the spectral representation with cutoff wavelength t_c is exactly the same as hard thresholding by t_c, i.e. setting all coefficients ζ_i with magnitude less than t_c to zero.

6.1 Haar wavelets

Wavelet functions are usually continuous and cannot represent discontinuities very well. A special case are the Haar wavelets which are discontinuous and thus are expected to represent much better discontinuous signals. The relation between Haar wavelets and TV regularization in one dimension has been investigated,

most notably in [52]. In higher dimensions, there is no straightforward analogy, in the case of isotropic TV, as it is not separable, contrary to the wavelet decomposition (thus one gets disk-like shapes as eigenfunctions as oppose to rectangles).

We will show here that even in the one-dimensional case TV eigenfunctions can represent signals in a much more concise manner.

Let $\psi_H(x)$ be a Haar mother wavelet function defined by

$$\psi_H(x) = \begin{cases} 1 & \text{for } x \in [0, \frac{1}{2}), \\ -1 & \text{for } x \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

For any integer pair $n, k \in \mathbb{Z}$, a Haar function $\psi_H^{n,k}(x) \in \mathbb{R}$ is defined by

$$\psi_H^{n,k}(x) = 2^{n/2}\psi_H(2^n x - k). \tag{27}$$

We now draw the straightforward relation between Haar wavelets and TV eigenfunctions.

Proposition 2 *A Haar wavelet function $\psi_H^{n,k}$ is an eigenfunction of TV with eigenvalue $\lambda = 2^{(2+n/2)}$.*

Proof One can express the mother wavelet as

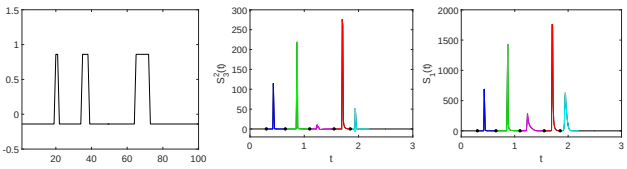
$$\psi_H(x) = B_{\frac{1}{2}}(x) - B_{\frac{1}{2}}(x - \frac{1}{2}),$$

and in general, any Haar function as

$$\psi_H^{n,k}(x) = h^n \left(B_{w^n}(x - x_0^{n,k}) - B_{w^n}(x - x_0^{n,k} - w^n) \right),$$

with $h^n = 2^{(n/2)}$, $w^n = 2^{-(n+1)}$ and $x_0^{n,k} = 2^{-n}k$. Thus, based on Prop. 1, we have that for any $n, k \in \mathbb{Z}$, $\psi_H^{n,k}$ is an eigenfunction of TV with $\lambda = \frac{2}{h^n w^n} = 2^{(2+n/2)}$.

In Figs. 8, 9 and 10 we show the decomposition of a signal f composed of 3 peaks of different width. The TV spectral decomposition shows 5 numerical deltas (corresponding to the 5 elements depicted in Fig. 9). On the other hand, the Haar wavelet decomposition needs 15 elements to represent this signal, thus the representation is less sparse. In the 2D case, Fig. 11, we see the consequence of an inadequate representation, which is less adapted to the data, where Haar thresholding is compared to ideal TV low-pass filtering, depicting blocky artifacts in the Haar case.



f $S_3^2(t)$ $S_1(t)$

Fig. 8: Example for a spectral decomposition: We evaluate the flow of equation (3) for given data f shown on the left. We compute the spectral decomposition (5), and compute a (discretized) version of the power spectral S_3^2 and S_1 from (15) and (12) shown in the middle and right plot respectively. Since there appear five distinct peaks, one can integrate the spectral ϕ components at the respective time intervals (=band-pass filter), visualized in different colors, to obtain the five components shown in figure 9.

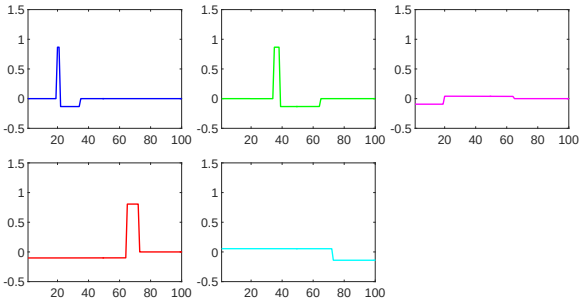


Fig. 9: Decomposing f (Fig. 8 left) into the 5 elements represented by the peaks in the spectrum using spectral TV.

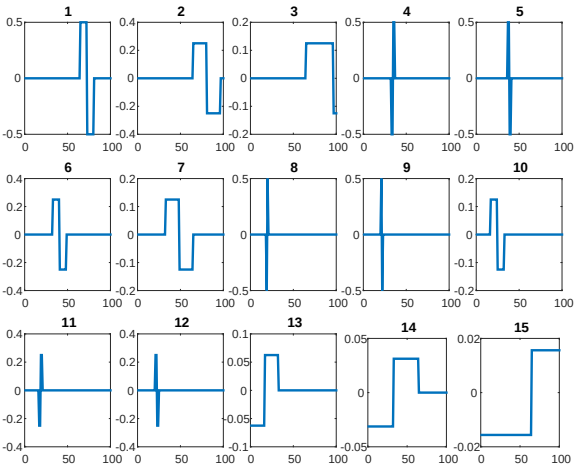
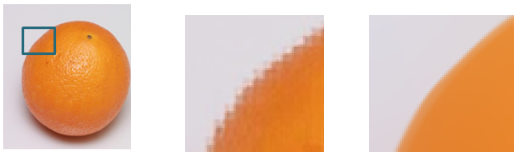


Fig. 10: Decomposing f (Fig. 8 left) with Haar wavelets using 15 elements.



Input f | Haar wavelet hard threshold | Spectral TV ideal LPF

Fig. 11: Comparison of wavelet Haar hard-thresholding to spectral TV hard-thresholding (ideal LPF). Although both representations can handle well discontinuities, the spectral TV representation is better adapted to the image edges and produces less artifacts.

7 Rayleigh Quotients and SVD Decomposition

In the classical Hilbert space case and J being a quadratic form, the Rayleigh quotient is defined by

$$R_M(v) := \frac{v^T M v}{v^T v}, \tag{28}$$

where for the real-valued case M is a symmetric matrix and v is a nonzero vector. It can be shown that, for a given M, the Rayleigh quotient reaches its minimum value at $R_M(v) = \lambda_1$, where λ_1 is the minimal eigenvalue of M and $v = v_1$ is the corresponding eigenvector (and similarly for the maximum).

One can generalize this quotient to functionals, similar as in [10], by

$$R_J(u) := \frac{J(u)^2}{\|u\|^2}, \tag{29}$$

where $\|\cdot\|$ is the L^2 norm. We restrict the problem to the non-degenerate case where $J(u) > 0$, that is, u is not in the null-space of J.

To find a minimizer, an alternative formulation can be used (as in the classical case):

$$\min_u \{J(u)\} \text{ s.t. } \|u\|^2 = 1. \tag{30}$$

Using Lagrange multipliers, the problem can be recast as

$$\min_u \{J(u) - \frac{\lambda}{2}\left(\|u\|^2 - 1\right)\},$$

with the optimality condition

$$0 \in \partial J(u) - \lambda u,$$

which coincides with the eigenvalue problem (1). Note that with the additional constraint of u being in the orthogonal complement of the nullspace of J in the case of J being absolutely one-homogeneous, in which case one obtains the minimal nonzero eigenvalue as a minimum of the generalized Rayleigh quotient (30).

The work by Benning and Burger in [10] considers more general variational reconstruction problems involving a linear operator in the data fidelity term, i.e.,

$$\min_u \frac{1}{2}\|Au - f\|_2^2 + t\ J(u),$$

and generalizes equation (1) to

$$\lambda A^* A u \in \partial J(u), \qquad \|Au\|_2 = 1,$$

in which case u is called a *singular vector*. Particular emphasis is put on the *ground states*

$$u^0 = \arg \min_{\substack{u \in \text{kern}(J)^\perp, \\ \|Au\|_2 = 1}} J(u)$$

for semi-norms J, which were proven to be singular vectors with the smallest possible singular value. Although the existence of a ground state (and hence the existence of singular vectors) is guaranteed for all reasonable J in regularization methods, it was shown that the Rayleigh principle for higher singular values fails. As a consequence, determining or even showing the existence of a basis of singular vectors remains an open problem for general semi-norms J.

In the setting of nonlinear eigenfunctions for one-homogeneous functionals, the Rayleigh principle for the second eigenvalue (given the smallest eigenvalue λ_1 and ground state u_1) leads to the optimization problem

$$\min_u \{J(u)\} \text{ s.t. } \|u\|^2 = 1, \langle u, u_1\rangle = 0. \tag{31}$$

With appropriate Lagrange multipliers λ and μ we obtain the solution as a minimizer of

$$\min_u \{J(u) - \frac{\lambda}{2}\left(\|u\|^2 - 1\right) + \mu\langle u, u_1\rangle\},$$

with the optimality condition

$$\lambda u - \mu u_1 = p \in \partial J(u).$$

We observe that we can only guarantee u to be an eigenvector of J if $\mu = 0$, which is not guaranteed in general. A scalar product with u_1 and the orthogonality constraint yields $\mu = -\langle p, u_1\rangle$, which only needs to vanish if $J(u) = \|u\|$, but not for general one-homogeneous J.

An example of a one-homogeneous functional failing to produce the Rayleigh principle for higher eigenvalues (which can be derived from the results in [10]) is given by $J : \mathbb{R}^2 \to \mathbb{R}$,

$$J(u) = \|Du\|_1, \qquad D = \begin{pmatrix} 1 & -2\epsilon \\ 0 & \frac{1}{\epsilon} \end{pmatrix},$$

for $0 < \epsilon < \frac{1}{2}$. The ground state $u = (u^1, u^2)$ minimizes $|u^1 - 2\epsilon u^2| + \frac{1}{\epsilon}|u^2|$ subject to $\|u\|_2 = 1$. It is easy to see

that $u = \pm(1,0)$ is the unique ground state, since by the normalization and the triangle inequality

$$1 = \|u\|_2 \leq \|u\|_1 \leq |u^1 - 2\epsilon u^2| + (1+2\epsilon)|u^2| \\ \leq |u^1 - 2\epsilon u^2| + \frac{1}{\epsilon}|u^2|,$$

and the last inequality is sharp if and only if $u^2 = 0$. Hence the only candidate v being normalized and orthogonal to u is given by $v = \pm(0,1)$. Without restriction of generality we consider $v = (0,1)$. Now, $Dv = (-2\epsilon, \frac{1}{\epsilon})$ and hence

$$\partial J(v) = \{D^T(-1,1)\} = \{(-1, 2\epsilon + \frac{1}{\epsilon})\},$$

which implies that there cannot be a $\lambda > 0$ with $\lambda v \in \partial J(v)$. A detailed characterization of functionals allowing for the Rayleigh principle for higher eigenvalues is still an open problems as well as the question whether there exists an orthogonal basis of eigenvectors in general. However, in the above example we are not able to construct an orthogonal basis of eigenvectors - this will become more apparent from the characterization of eigenvectors in the next section.

8 Sparse Representation by Eigenfunctions

With our previous definitions a convex functional induces a dictionary by its eigenfunctions. Let us formalize this notion by the following definition:

Definition 1 (Eigenfunction Dictionary) Let $\mathcal{D}_J$ be a dictionary of functions in a Banach space $\mathcal{X}$, with respect to a convex functional J, defined as the set of all eigenfunctions,

$$\mathcal{D}_J := \bigcup_{\lambda \in \mathbb{R}} \{u \in \mathcal{X} \,|\, \lambda u \in \partial J(u),\ \|u\|_2 = 1\}.$$

There are natural questions to be asked such as:

1. What functions can be reconstructed by a linear combination of functions in the dictionary $\mathcal{D}_J$?
2. Is $\mathcal{D}_J$ an overcomplete dictionary?
3. Does $\mathcal{D}_J$ contain a complete orthonormal system ?
4. How many elements are needed to express some type of functions (or how sparse are they in this dictionary)?

We notice that for many functionals, in particular seminorms, there is a natural ambiguity between u and $-u$, which are both elements of $\mathcal{D}_J$. Thus we might either ask for conical combinations or identify u and $-u$ as the same elements of the dictionary.

So far the above questions have not been investigated systematically and there is no general answer. However, from the existing examples we expect the eigenfunctions of a convex one-homogeneous function to contain a basis, and often be extremely overcomplete. Regarding the third aspect, it would be desirable to have a criterion for when input data consisting of a linear combination of eigenfunctions can actually be decomposed into its original atoms.

For some simple cases, we can immediately give an answer to the questions raised above. Firstly, note that in the case of $J(u) = \|u\|_1$ any $u = \frac{g}{\|g\|_2}$ with $g_i \in \{-1, 0, +1\}$ is an eigenfunction (except $g \equiv 0$). Thus, even after removing the sign ambiguity there still exist $3^n/2 - 1$ different eigenfunctions in $\mathbb{R}^n$. Remarkably, the number of eigenfunctions grows exponentially with the dimension of the space.

8.1 Total Variation Dictionaries

In the continuous setting, the result of Proposition 2 stating that any Haar-wavelet function is a TV eigenfunction immediate yields the existence of a basis of eigenfunction in this case, too. (See for example [22] for details on the wavelet reconstruction properties). Moreover, the explicit construction of eigenfunctions according to Proposition 1 along with the freedom to choose the $x_i \in \mathbb{R}$ arbitrarily yields that there are uncountably many eigenfunctions and the dictionary is overcomplete.

Without having the immediate connection to wavelets, the above results extend to the c-dimensional TV regularization:

Corollary 1 *For $J_{TV} : BV(\mathbb{R}^c) \to \mathbb{R}$ being the TV functional, any $f \in L^2(\mathbb{R}^c)$ can be approximated up to a desired error ε by a finite linear (respectively conical) combination of elements from $\mathcal{D}_{J_{TV}}$. Moreover, $\mathcal{D}_{J_{TV}}$ is an overcomplete dictionary (so the linear combination is not unique).*

Proof First of all, we can approximate any L^2-function to arbitrary precision with a linear combination of a finite number of piecewise constant functions. Since the Borel sets are generated by balls, we can further construct an approximation to arbitrary precision with functions piecewise constant on balls, i.e. linear combinations of characteristic functions of balls. Since the latter are eigenfunctions of the total variation, i.e., elements of $\mathcal{D}_J$, the dictionary is complete. Moreover, since there are Cheeger sets not equal to a ball, and their characteristic functions are in $\mathcal{D}_J$, too, the dictionary is overcomplete.

While the properties of eigenfunctions with respect to the total variation regularization are fairly well understood, the question for the properties of a dictionary

of eigenfunctions for general one-homogeneous regularization functionals remains open. In [16] we showed that the regularization $J(u) = \|Du\|_1$ for DD^* being diagonally dominant always yields the existence of a basis of eigenfunctions. We do, however, expect to have a significantly overdetermined dictionary of eigenfunctions. Unfortunately, classical ways of determining eigenfunctions – such as the Rayleigh principle – fail in the setting of generalized eigenfunctions (1) for non-quadratic J as we have seen in the previous section. Nevertheless, for absolutely one-homogeneous functions we can describe the set of eigenvectors more precisely.

8.2 Dictionaries from One-homogeneous Functionals

In the following let us consider one-homogeneous functionals $J : \mathbb{R}^n \to \mathbb{R}$, which by duality can be represented as

$$J(u) = \sup_{q \in K} \langle q, u \rangle \tag{32}$$

for some convex set K. In order to avoid technicalities (related to a possible nullspace of J) we assume that K has nonempty interior. The following result provides a characterization of eigenvectors for nonzero eigenvalues:

Lemma 1 *Let J be defined by (32) and $p \in \partial K$ satisfy*

$$\langle p, p - q \rangle \geq 0 \qquad \forall\ q \in K. \tag{33}$$

Then $u = \frac{p}{\|p\|}$ is an eigenvector with eigenvalue $\lambda = \|p\|$. Vice versa, if u is an eigenvector with eigenvalue $\lambda \neq 0$, then $p = \lambda u$ satisfies (33).

Proof We have $p \in \partial J(u)$ if and only if p is the maximal element in $\langle q, u \rangle$ over all $q \in K$. Hence, u is an eigenvector, i.e. a positive multiple of p if and only if (33) holds.

This result has a straightforward geometric interpretation: if $u = \lambda p$ is an eigenvector corresponding to $\lambda \neq 0$, then a ball centered at zero is tangential to ∂K at p. In other words the hyperplane through p and orthogonal to p is tangential to ∂K, more precisely K lies on one side of this hyperplane.

The above geometric insight allows a more precise understanding of the example from the previous section. For $J(u) = \|Du\|_1$, we have

$$K = \{\ D^T e \mid \|e\|_\infty \leq 1\},$$

i.e. K is a sheared version of the unit square. For the specific form of D in the example, the facets remaining parallel to the second coordinate axis remain unchanged. If the shearing is not too strong ($\epsilon < \frac{1}{2}$) the first tangential touch of a ball centered at zero happens at $p = \pm(1, 0)$, which gives the first eigenvector. A second tangential touch happens at the extremal points of the sheared square, which obviously yields an eigenvalue with nonzero first coordinate, hence not orthogonal to the first eigenvector. The construction of the minimal and maximal eigenvalue from this example can indeed be generalized with a straight-forward proof to the general case:

Proposition 3 *Let J be defined by (32) and p be a solution of*

$$\|p\| \to \min_{p \in \partial K}, \tag{34}$$

then $\lambda = \|p\|$ is the minimal eigenvalue of J and $u = \lambda p$ is a ground state. Let p be a solution of

$$\|p\| \to \max_{p \in \partial K}, \tag{35}$$

then $\lambda = \|p\|$ is the maximal eigenvalue of J and $u = \lambda p$ is a corresponding eigenvector.

9 Cheeger sets and spectral clustering

As mentioned in the introduction, there is a large literature on calibrable sets respectively Cheeger sets, whose characteristic function is an eigenfunction of the total variation functional, compare (2). Particular interest is of course paid to the lowest eigenvalues and their eigenfunctions (ground states), which are related to solutions of the isoperimetric problem, easily seen from the relation $\lambda = \frac{P(C)}{|C|}$. Due to scaling invariance of λ, the shape of C corresponding to the minimal eigenvalue can be determined by minimizing $P(C)$ subject to $|C| = 1$. Hence, the ground states of the total variation functional are just the characteristic functions of balls in the isotropic setting. Note that the normalization of $|C| = 1$ is consistent with the constraint of ground states having a unit L^2 norm. Interesting variants are obtained with different definitions of the total variation via anisotropic vector norms. E.g. if one uses the ℓ^1-norm of the gradient in the definition, then it is a simple exercise to show that the ground state is given by characteristic functions of squares with sides parallel to the coordinate directions. In general, the ground state is the characteristic function of the Wulff shape related to the special vector norm used (cf. [8,28]) and is hence a fingerprint of the used metric.

On bounded domains or in (finite) discrete versions of the total variation the situation differs, a trivial argument shows that the first eigenvalue is simply zero, with a constant given as the ground state. Since this

is not interesting, it was even suggested to redefine the ground state as an eigenfunction of the smallest non-zero eigenvalue (cf. [10]). The more interesting quantity is the second eigenfunction, which is orthogonal to the trivial ground state (i.e. has zero mean). In this case the Rayleigh-principle always works (cf. [10]) and we obtain

$$\lambda_2 = \inf_{u, \int u=0} \frac{J_{TV}(u)}{\|u\|_2}. \tag{36}$$

Depending on the underlying domain on which total variation is considered (and again the precise definition of total variation) one obtains the minimizer as a function positive in one part and negative in another part of the domain, usually with constant values in each. Hence, the computation of the eigenvalue can also be interpreted as some optimal cut through the domain.

The latter idea had a strong impact in data analysis (cf. [53,13,12,39,50]), where the total variation is defined as a discrete functional on a graph, hence one obtains a relation to graph cuts and graph spectral clustering. For data given in form of a weighted graph, with weights w_{ij} on an edge between two vertices i, j in the vertex set $\mathcal{W}$, the total variation is defined as

$$J_{1,w}(u) = \sum_{i,j \in \mathcal{V}} w_{ij} |u_i - u_j| \tag{37}$$

for a function $u : \mathcal{V} \to \mathbb{R}$. Then a graph cut is obtained from a minimizing u in

$$\lambda_2 = \inf_{u, \sum u_i = 0} \frac{J_{1,w}(u)}{\|u\|_2}, \tag{38}$$

i.e. an eigenfunction corresponding to the second eigenvalue. The cut directly yields a clustering into the two classes

$$\mathcal{C}_+ = \{i \in \mathcal{V} \mid u_i > 0\}, \qquad \mathcal{C}_+ = \{i \in \mathcal{V} \mid u_i < 0\}. \tag{39}$$

Note that in this formulation the graph cut technique is completely analogous to the classical spectral clustering technique based on second eigenvalues of the graph Laplacian (cf. [55]). In the setting (38), the spectral clustering based on the graph Laplacian can be incorporated by using the functional

$$J_{2,w}(u) = \sqrt{\sum_{i,j \in \mathcal{V}} w_{ij} |u_i - u_j|^2}. \tag{40}$$

Some improvements in spectral clustering are obtained if the normalization in the eigenvalue problem is not done in the Euclidean norm (or a weighted variant thereof), but some ℓ^p-norm, i.e.,

$$\lambda_2 = \inf_{u, \sum u_i = 0} \frac{J_{r,w}(u)}{\|u\|_p}, \tag{41}$$

for $r = 1$ or generalizations to weighted norms. We refer to [39] for an overview, again the case $p = 1$ has received particular attention. In the case of graph-total-variation, see a recent initial analysis and a method to construct certain type of concentric eigenfunctions in [4]. A thorough theoretical study of eigenvalue problems and spectral representations for such generalized eigenvalue problems is still an open problem, as for any non-Hilbert space norm. The corresponding gradient flow becomes a doubly nonlinear evolution equation of the form

$$0 \in \partial \|\partial_t u\|_p + \partial J(u). \tag{42}$$

The gradient flow for $p = 1$ and total variation in the continuum is related to the L^1-TV model, which appears to have interesting multiscale properties (cf. [58]).

A challenging problem is the numerical computation of eigenvalues and eigenfunctions in the nonlinear setup, which is the case also in the general case beyond the spectral clustering application. Methods reminiscent of the classical inverse power method in eigenvalue computation (cf. e.g. [34]) have been proposed and are used with some success, but it is difficult to obtain and guarantee convergence to the smallest or second eigenvalue in general.

10 Numerical Implementation Issues

Here we give the gradient flow approximation. We would like to approximate $u_t = -p(u)$. We use the Moreau-Yosida approximation [43] for gradient flow. The implicit discrete evolution (which is unconditionally stable in dt) is

$$u(n+1) = u(n) - dtp(u(n+1)).$$

We can write the above expression as

$$u(n+1) - u(n) + dtp(u(n+1)) = 0,$$

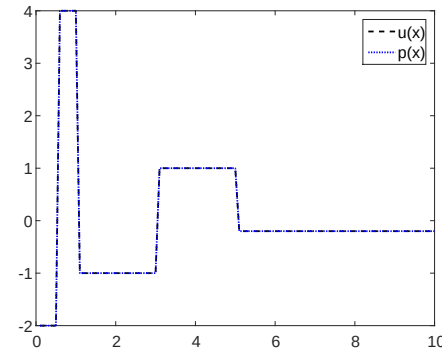
and see it coincides with the Euler-Lagrange of the following minimization:

$$E(u, u(n)) = J(u) + \frac{1}{2dt} \|u - u(n)\|_{l^2}^2, \tag{43}$$

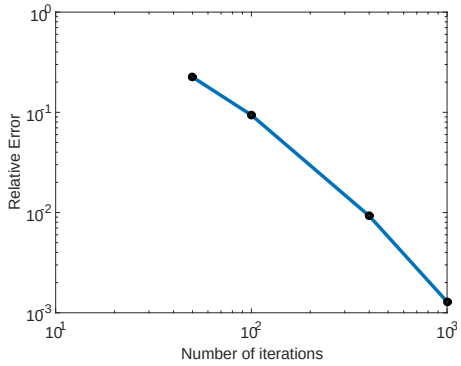
where $u(n+1)$ is the minimizer u of $E(u, u(n))$ and $u(n)$ is fixed. We now have a standard convex variational problem which can be solved using various algorithms (e.g. [24,19,33,47,21]).

To approximate the second time derivative $u_{tt}(t)$ we store in memory 3 consecutive time steps of u and use the standard central scheme:

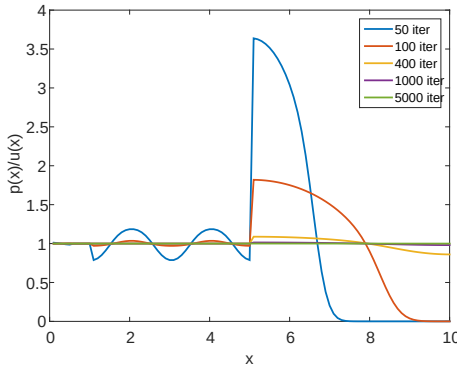
$$D^2 u(n) = \frac{u(n-1) + u(n+1) - 2u(n)}{(dt)^2},$$



$u(x)$ and $p(x)$ at convergence



Relative absolute error to the solution at convergence as a function of iterations.



Ratio $\frac{p^m(x)}{u^m(x)}$ for different number of iterations m.

Fig. 12: Numerical implementation of a 1D TV eigenfunction, Chambolle-Pock scheme [21]. The ratio $\frac{p(x)}{u(x)} = \lambda_1$ should converge to $1/(1-\epsilon)$. It takes about 5,000 iterations for the numerical scheme to fully converge. Note that the error is data-dependent, where larger flat regions converge much slower.

with $n = 1, 2, ..,\ u(0) = f,\ D^2u(0) = 0$. The time t is discretized as $t(n) = ndt$. Therefore:

$$\phi(n) = D^2u(n)t(n) = \frac{n}{dt}\left(u(n-1) + u(n+1) - 2u(n)\right), \tag{44}$$

and for instance the spectrum S_1 is

$$S_1(n) = \sum_{i \in \mathcal{N}} |\phi_i(n)|, \tag{45}$$

where i is a pixel in the image and $\mathcal{N}$ is the image domain.

In Fig. 12 we show a numerical computation for eigenfunctions and visualize the spatial convergence. We do a single time step of a TV flow in the following way. Let f be a TV eigenfunction in 1D in a bounded domain, which admits Eqs. (22), (25), for simplicity we chose $\lambda = 1$. We set $n = 0$, $u(0) = f$ and find the minimizer of (43) for a very small time step $dt = \epsilon$ (in our implementation we had $dt = 0.0001$). Following Eq. (4), we should get that $u(n = 1, x) \equiv u(x) = (1 - dt)f(x)$. Note that the subdifferential of any absolutely one-homogeneous regularization function is zero-homogeneous, i.e. $\partial J(cu) = \partial J(u)$ for all $c > 0$, which means that $f \in \partial J_{TV}(f)$ implies $\frac{1}{1-dt}u(n = 1, x) \in \partial J(u(n = 1, x))$. For the minimization we use the scheme of Chambolle-Pock [21], where we have internal m iterations. Let $u^m(x)$, $p^m(x)$ be the result of the numerical scheme at iteration m. At full convergence, as $m \to \infty$, we have $\lim_{m\to\infty} \frac{p^m(x)}{u^m(x)} = \lambda_1 \approx 1$ *pointwise*, as shown in the green flat line of Fig. 12 (for $m = 5000$ iterations).

It is shown that such schemes are far better for oscillating signals (works well for denoising), but when there are large flat regions the local magnitude of p depends on the support of that region and convergence is considerably slower (can be orders of magnitude). For the 1D case shown here, alternative schemes can be used which are much faster, such as [23,35]. Furthermore, for our purpose of constructing a full solution path, variants of the homotopy method [45] or the adaptive inverse scale space flow [18] would be well suited for the 1D case as they can construct exact solutions without discretizing the underlying equation in time. However, as far as we know, there are no efficient schemes for higher dimensions which deal well with flat regions. Thus alternative ways are desired to solve such cases more efficiently.

11 Extensions to other nonlinear decomposition techniques

As presented in [14] the general framework of nonlinear spectral decompositions via one-homogeneous regularization functionals does not have to be based on the gradient flow equation (3), but may as well be defined via a variational method of the form

$$u(t) = \arg\min_u \frac{1}{2}\|u - f\|_2^2 + tJ(u), \tag{46}$$

	Wavelength	**Frequency**
Gradient flow $\partial_t u(t) \in -\partial J(u(t))$, $u\vert_{t=0} = f$.	$\phi(t) = t\partial_{tt} u(t)$	$\psi(s) = \frac{1}{s^2}\phi(\frac{1}{s})$
Variational minimization $tJ(u) + \frac{1}{2}\|f-u\|_2^2$.	$\phi(t) = t\partial_{tt} u(t)$	$\psi(s) = \frac{1}{s^2}\phi(\frac{1}{s})$
Inverse-scale-space $\partial_s p(s) = f - u(s)$, $p(s) \in \partial J(u(s))$.	$\phi(t) = \frac{1}{t^2}\psi(\frac{1}{t})$	$\psi(s) = \partial_s u(s)$

Table 2: Six spectral representations of one-homogeneous functionals.

or an inverse scale space flow of the form

$$\partial_s p(s) = f - u(s), \qquad p(s) \in \partial J(u(s)). \tag{47}$$

Since the variational method shows exactly the same behavior as the scale space flow for the input data f being an eigenvector of J, the definitions of the spectral decomposition coincides with the one of the scale the flow. The behavior of the inverse scale space flow on the other hand is different in two respects. Firstly, it start at $u(0)$ being the projection of f onto the kernel of J and converges to $u(s) = f$ for sufficiently large s. Secondly, the dynamics for f being an eigenfunction of J is piecewise constant in $u(s)$ such that only a single derivative is needed to obtain a peak in the spectral decomposition. The behavior on eigenfunctions furthermore yields the relation $s = \frac{1}{t}$ when comparing the spectral representation obtained by the variational or scale space method with the one obtained by the gradient flow.

In analogy to the linear spectral analysis it is natural to call the spectral representation of the inverse scale space flow a *frequency representation* opposed to the *wavelength representation* of the variational and scale space methods.

To be able to obtain frequency and wavelength representations for all three types of methods, we make the convention that a filter $H(t)$ acting on the wavelength representation ϕ should be equivalent to the filter $H(1/t)$ acting the frequency representation ψ, i.e.

$$\int_0^\infty \phi(t)H(t)\ dt = \int_0^\infty \psi(t)H(1/t)\ dt.$$

By a change of variables one deduces that

$$\psi(t) = \frac{1}{t^2}\phi(\frac{1}{t}) \text{ respectively } \phi(t) = \frac{1}{t^2}\psi(\frac{1}{t})$$

are the appropriate conversion formulas to switch from the frequency to the wavelength representation and vice versa. Table 2 gives an overview over the three nonlinear decomposition methods and their spectral representation in terms of frequencies and wavelength.

It can be verified that in particular cases such as regularizations of the form $J(u) = \|Vu\|_1$ for an orthonormal matrix V, or for the data f being an eigenfunction of J, the gradient flow, the variational method and the inverse scale space flow yield exactly the same spectral decompositions ϕ and ψ. In [16] we are investigating more general equivalence results and we refer the reader to [27] for a numerical comparison of the above approaches. A precise theory of the differences of the three approaches for a general one-homogeneous J, however, remains an open question.

12 Preliminary applications

12.1 Filter design for denoising

One particular application of the spectral decomposition framework could be the design of filters in denoising applications. It was shown in [31, Theorem 2.5] that the popular denoising strategy of evolving the scale space flow (3) for some time t_1, is equivalent to the particular filter

$$H(t) = \begin{cases} 0 & \text{for } 0 \le t \le t_1 \\ \frac{t-t_1}{t} & \text{for } t_1 \le t \le \infty \end{cases}$$

in the framework of equation (7). The latter naturally raises the question if this particular choice of filter is optimal for practical applications. While in case of a perfect separation of signal and noise an ideal low pass filter (8) may allow a perfect reconstruction (as illustrated on synthetic data in figure 13), the spectral decomposition of natural noisy images will often contain noise as well as texture in high frequency components.

In [42] a first approach to learning optimal denoising filters on a training data set of natural images demonstrated promising results. In particular, it was shown that optimal filters neither had the shape of ideal low pass filters nor of the filter arising from evolving the gradient flow.

In future research projects different regularization terms for separating noise and signals in a spectral framework will be investigated. Moreover, the proposed

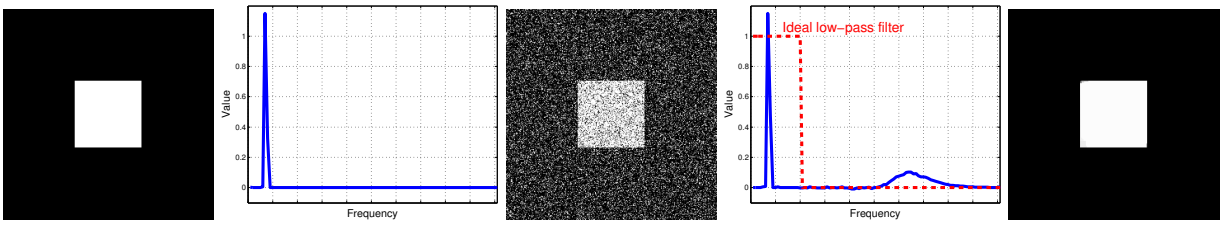


Fig. 13: Example for perfect separation of noise and signal via anisotropic TV regularization in the framework of nonlinear spectral decompositions using the inverse scale space flow. From left to right: Clean image, corresponding spectrum of the clean image, noisy image, spectrum of the noisy image with an ideal low pass filter to separate noise and signal, reconstructed image after applying the ideal low pass filter. We used the definition S_3 in this illustration of the spectrum.

spectral framework allows to filter more general types of noise, i.e. filters are not limited to considering only the highest frequencies as noise. Finally, the complete absence of high frequency components may lead to an unnatural appearance of the images, such that the inclusion of some (damped) high frequency components may improve the visual quality despite possibly containing noise. Figure 14 supports this conjecture by a preliminary example.

12.2 Texture processing

One can view the spectral representation as an extension to infinite dimensions of multiscale approaches for texture decomposition, such as [54,32]. In this sense $\phi(t)$ of (5) is an infinitesimal textural element (which goes in a continuous manner from "texture" to "structure", depending on t). A rather straightforward procedure is therefore to analyze the spectrum of an image and either manually or automatically select integration bands that correspond to meaningful textural parts in the image. This was done in [36], where a multiscale orientation descriptor, based on Gabor filters, was constructed. This yields for each pixel a multi-valued orientation field, which is more informative for analyzing complex textures. See Fig. 15 as an example of such a decomposition of the Barbara image.

In [37,38] a unique way of using the spectral representation was suggested for texture processing. It was shown that textures with gradually varying pattern-size, pattern-contrast or illumination can be represented by surfaces in the three dimensional TV transform domain. A spatially varying texture decomposition amounts to estimating a surface which represents significant maximal response of ϕ, within a time range $[t_1, t_2]$,

$$\max_{t\in[t_1,t_2]} \{\phi(t;x)\} > \epsilon$$

for each spatial coordinate x. In Fig. 16 a wall is shown with gradually varying texture scale and its scale representation in the spectral domain. A decomposition of structures with gradually varying contrast is shown in Fig. 17 where the band-spectral decomposition (spatially varying scale separation) is compared to the best TV-G separation of [5].

13 Discussion and Conclusion

In this paper we presented the rationale for analyzing one-homogeneous variational problems through a spectral approach. It was shown that solutions of the generalized nonlinear eigenvalue problem (1) are a fundamental part of this analysis. The quadratic (Dirichlet) energy yields Fourier frequencies as solutions, and thus eigenfunctions of one-homogeneous functionals can be viewed as new non-linear extensions of classical frequencies.

Analogies to the Fourier case, to wavelets and to dictionary representations were drawn. However, the theory is only beginning to be formed and there are many open theoretical problems, a few examples are:

1. More exact relations of the one-homogeneous spectral representations to basis and frames representations.
2. Understanding the difference between the gradient-flow, variational and inverse-scale-space representations. As we have shown in [16], a regularization $J(u) = \|Du\|_1$ with DD^* being diagonally dominant yields the equivalence between the spectral decomposition of all three methods. Under a (weaker) regularity assumption on the subdifferentials, we obtained the equivalence between the gradient flow and the variational approach. We do, however, expect the equivalence of all three methods to hold under more general assumptions. It would be of particular interest to classify the difference in cases where the representations are not equivalent.
3. Consistency of the decomposition: If we filter the spectral decomposition of some input data f, e.g.

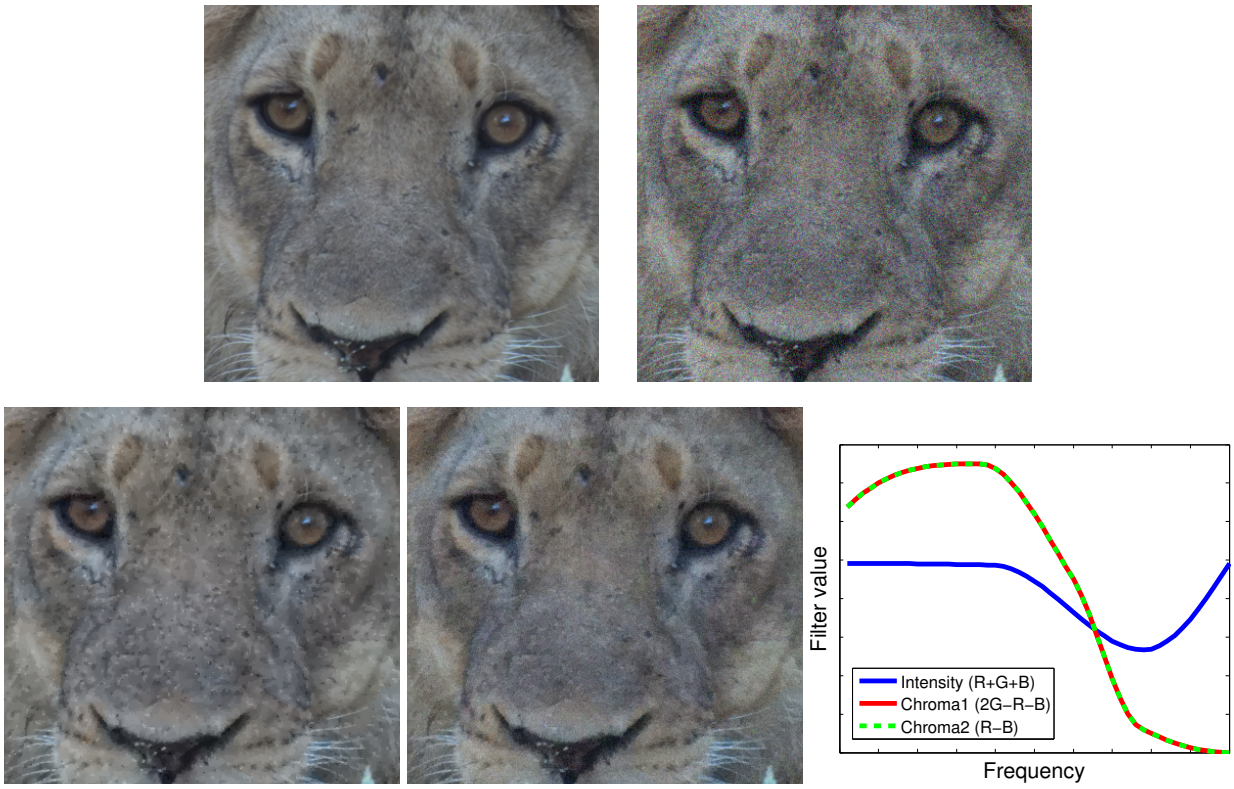


Fig. 14: Example for denoising natural images: The original image (upper left) contains a lot of texture. If one tries to denoise the upper right image with an ideal low pass filter (using TV in a color space that decouples intensities and chromaticities), we obtain the lower left image. Since some of the high frequencies are completely suppressed, the image looks unnatural. A designed filter as shown in the lower right can produce more visually pleasing results (lower middle image). This example was produced using the ISS framework such that the filters are shown in frequency representation. Note that the suppression of color artifacts seems to be significantly more important than the suppression of oscillations.

by an ideal low pass filter at frequency T, and apply the spectral decomposition again, will we see any frequencies greater than T?

4. Are the regularizations investigated in [16] the most general class of functions for which one can guarantee orthogonality of the spectral representation ϕ?
5. Spectral analysis of random noise: Can we expect to classify the decay of coefficients from high to low frequencies, for instance for Gaussian random noise? Does this lead to implications about the optimal choice of filters for suppressing noise?
6. Can the theory be extended from the one-homogeneous case to the general convex one?

In addition, there are many practical aspects, such as

1. Learning the regularization on a training set with the goal to separate certain features.
2. Numerical issues, computing ϕ in a stable manner (as it involves a second derivative in time), also can one design better schemes than the ones given here, which are less local and converge fast for any eigenfunction?
3. Additional applications where the new representations can help in better design of variational algorithms.

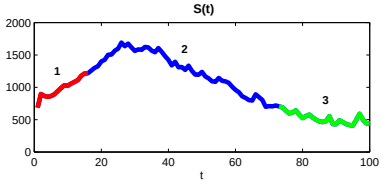
Some initial applications related to filter design and to texture processing were shown. This direction seems as a promising line of research which can aid in better understanding of variational processing and that has a strong potential to provide alternative improved algorithms.

References

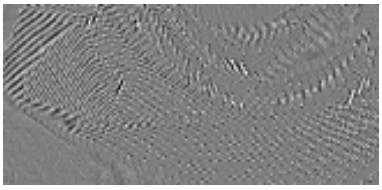
1. F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. Differential and Integral Equations, 14(3):321–360, 2001.



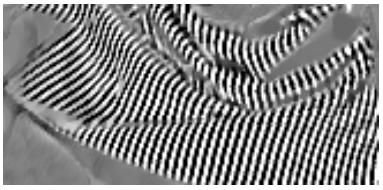
(a) Input



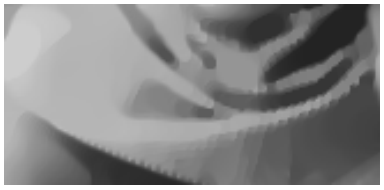
(b) Spectrum



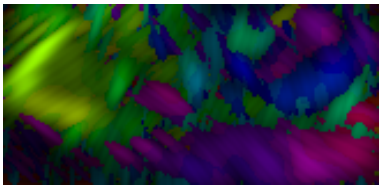
(c) Scale 1



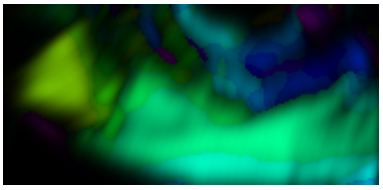
(d) Scale 2



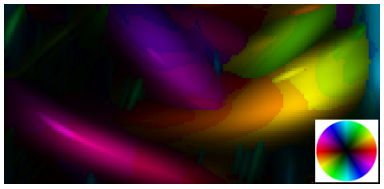
(e) Scale 3



(f) Orientation 1

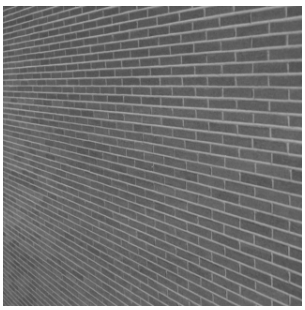


(g) Orientation 2

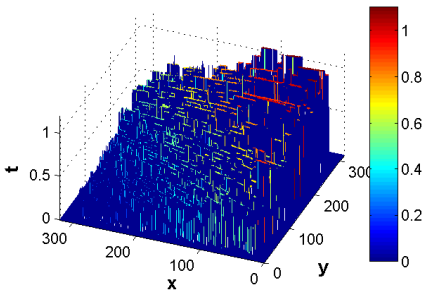


(h) Orientation 3

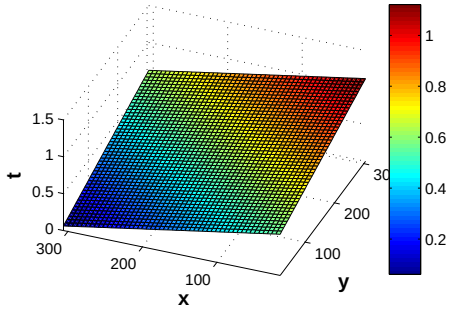
Fig. 15: Multiscale orientation analysis based on spectral decomposition. (a) Barbara image, (b) TV spectrum of the image with separated scales marked in different colors. (c-e) Multiscale decomposition, (f-h) Corresponding orientation maps.



(a) Input image



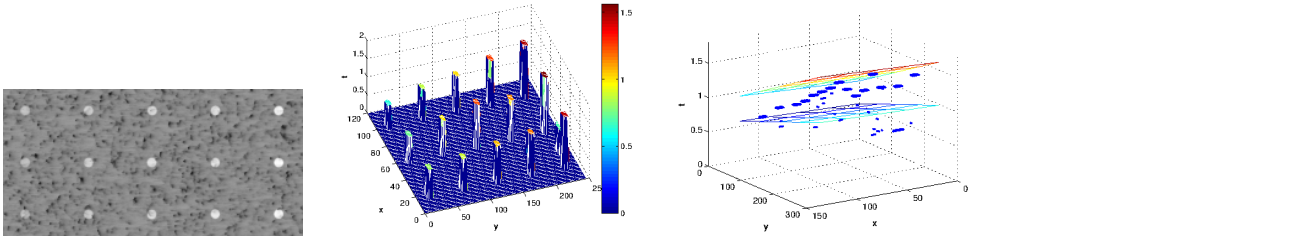
(b) Maximal ϕ



(c) Texture surface

Fig. 16: Spectral analysis of a wall image, from [38]. (a) Input image of a brick wall (b) $S(x) = \max_t \phi(t;x)$. (c) Approximation of $S(x)$ by a plain.

2. F. Andreu, V. Caselles, JI Dıaz, and JM Mazón. Some qualitative properties for the total variation flow. Journal of Functional Analysis, 188(2):516–547, 2002.
3. G. Aubert and P. Kornprobst. Mathematical Problems in Image Processing, volume 147 of Applied Mathematical Sciences. Springer-Verlag, 2002.
4. J.-F. Aujol, G. Gilboa, and N. Papadakis. Fundamentals of non-local total variation spectral theory. In Scale Space and Variational Methods in Computer Vision, pages 66–77. Springer, 2015.
5. J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. JMIV, 22(1), January 2005.
6. S. Bartels, R.H. Nochetto, J. Abner, and A.J. Salgado. Discrete total variation flows without regularization. arXiv preprint arXiv:1212.1137, 2012.



(a) Input image (b) Maximal ϕ (c) Separation band

(d) Spectral band separation [37], layer 1 (e) Spectral band separation, layer 2 (f) TV-G [5], layer 1 (g) TV-G, layer 2

Fig. 17: Decomposition by a separation band in the spectral domain. (a) Input image, (b) The maximal ϕ response, (c) The separation band, (d-e) spectral decomposition, (f-g) TV-G decomposition. Taken from [37].

7. G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in R^N. Journal of Differential Equations, 184(2):475–525, 2002.
8. Marino Belloni, Vincenzo Ferone, and Bernd Kawohl. Isoperimetric inequalities, wulff shape and related questions for strongly nonlinear elliptic operators. Zeitschrift für angewandte Mathematik und Physik ZAMP, 54(5):771–783, 2003.
9. M. Benning, C. Brune, M. Burger, and J. Müller. Higher-order tv methods:enhancement via bregman iteration. J Sci Comput, 54:269–310, 2013.
10. M. Benning and M. Burger. Ground states and singular vectors of convex variational regularization methods. Methods and Applications of Analysis, 20(4):295–334, 2013.
11. K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. SIAM J. Imaging Sciences, 3(3):492–526, 2010.
12. Xavier Bresson, Thomas Laurent, David Uminsky, and James V Brecht. Convergence and energy landscape for cheeger cut clustering. In Advances in Neural Information Processing Systems, pages 1385–1393, 2012.
13. Xavier Bresson and Arthur D Szlam. Total variation, cheeger cuts. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pages 1039–1046, 2010.
14. M. Burger, L. Eckardt, G. Gilboa, and M. Moeller. Spectral representations of one-homogeneous functionals. In Scale Space and Variational Methods in Computer Vision, pages 16–27. Springer, 2015.
15. M. Burger, K. Frick, S. Osher, and O. Scherzer. Inverse total variation flow. Multiscale Modeling & Simulation, 6(2):366–395, 2007.
16. M. Burger, G. Gilboa, M. Moeller, L. Eckardt, and D. Cremers. Spectral decompositions using one-homogeneous functionals, 2015. Submitted. Online at http://arxiv.org/pdf/1601.02912v1.pdf.
17. M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. Comm. in Math. Sci., 4(1):179–212, 2006.
18. M. Burger, M. Moeller, M. Benning, and S. Osher. An adaptive inverse scale space method for compressed sensing. Mathematics of Computation, 82:269–299, 2013.
19. A. Chambolle. An algorithm for total variation minimization and applications. JMIV, 20:89–97, 2004.
20. A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. Numerische Mathematik, 76(3):167–188, 1997.
21. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1):120–145, 2011.
22. Charles K Chui. An introduction to wavelets, volume 1. Academic press, 1992.
23. L. Condat. A direct algorithm for 1-d total variation denoising. IEEE Signal Processing Letters, 20(11):1054–1057, 2013.
24. J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part i: Fast and exact optimization. Journal of Mathematical Imaging and Vision, 2006.
25. L. Dorst and R. Van den Boomgaard. Morphological signal processing and the slope transform. Signal Processing, 38(1):79–98, 1994.
26. J. Duran, M. Moeller, C. Sbert, and D. Cremers. Collaborative total variation: A general framework for vectorial tv models. Submitted. Preprint at http://arxiv.org/abs/1508.01308.
27. L. Eckardt. Spektralzerlegung von bildern mit tv-methoden, 2014. Bachelor thesis, University of Münster.
28. Selim Esedoglu and Stanley J Osher. Decomposition of images by the anisotropic rudin-osher-fatemi model. Communications on pure and applied mathematics, 57(12):1609–1626, 2004.
29. Y. Giga and R.V. Kohn. Scale-invariant extinction time estimates for some singular diffusion equations. Hokkaido University Preprint Series in Mathematics, (963), 2010.
30. G. Gilboa. A spectral approach to total variation. In A. Kuijper et al. (Eds.): SSVM 2013, volume 7893 of Lecture Notes in Computer Science, pages 36–47. Springer, 2013.

31. G. Gilboa. A total variation spectral framework for scale and texture analysis. SIAM J. Imaging Sciences, 7(4):1937–1961, 2014.
32. J. Gilles. Multiscale texture separation. Multiscale Modeling & Simulation, 10(4):1409–1427, 2012.
33. T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. SIAM Journal on Imaging Sciences, 2(2):323–343, 2009.
34. Matthias Hein and Thomas Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse pca. In Advances in Neural Information Processing Systems, pages 847–855, 2010.
35. W. Hinterberger, M. Hintermüller, K. Kunisch, M. von Oehsen, and O. Scherzer. Tube methods for bv regularization. Journal of Mathematical Imaging and Vision, 19:219–235, 2003.
36. D. Horesh and G. Gilboa. Multiscale texture orientation analysis using spectral total-variation decomposition. In Scale Space and Variational Methods in Computer Vision, pages 486–497. Springer, 2015.
37. D. Horesh and G. Gilboa. Separation surfaces in the spectral TV domain for texture decomposition, 2015. Submitted. Preprint at http://arxiv.org/abs/1511.04687.
38. Dikla Horesh. Separation surfaces in the spectral TV domain for texture decomposition, 2015. Master thesis, Technion.
39. Leonardo Jost, Simon Setzer, and Matthias Hein. Nonlinear eigenproblems in data analysis: Balanced graph cuts and the ratiodca-prox. In Extraction of Quantifiable Information from Complex Systems, pages 263–279. Springer, 2014.
40. U. Köthe. Local appropriate scale in morphological scale-space. In ECCV'96, pages 219–228. Springer, 1996.
41. Y. Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The 15th Dean Jacquelines B. Lewis Memorial Lectures.
42. M. Moeller, J. Diebold, G. Gilboa, and D. Cremers. Learning nonlinear spectral filters for color image reconstruction. In ICCV 2015, 2015. To appear.
43. J-J. Moreau. Proximité et dualité dans un espace hilbertien. Bulletin de la Société mathématique de France, 93:273–299, 1965.
44. J. Müller. Advanced image reconstruction and denoising: Bregmanized (higher order) total variation and application in pet, 2013. Ph.D. Thesis, Univ. Münster.
45. M.R. Osborne, B. Presnell, and B.A. Turlach. A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis, 20(3):389–403, 2000.
46. K. Papafitsoros and K. Bredies. A study of the one dimensional total generalised variation regularisation problem. arXiv preprint arXiv:1309.5900, 2013.
47. T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the mumford-shah functional. In Computer Vision, 2009 IEEE 12th International Conference on, pages 1133–1140. IEEE, 2009.
48. C. Pöschl and O. Scherzer. Exact solutions of one-dimensional tgv. arXiv preprint arXiv:1309.7152, 2013.
49. L.R. Rabiner and B. Gold. Theory and application of digital signal processing. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1975. 777 p., 1, 1975.
50. Syama Sundar Rangapuram and Matthias Hein. Constrained 1-spectral clustering. arXiv preprint arXiv:1505.06485, 2015.
51. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D, 60:259–268, 1992.
52. G. Steidl, J. Weickert, T. Brox, P. Mrzek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. SIAM Journal on Numerical Analysis, 42(2):686–713, 2004.
53. Arthur Szlam and Xavier Bresson. A total variation-based graph clustering algorithm for cheeger ratio cuts. UCLA CAM Report, pages 09–68, 2009.
54. E. Tadmor, S. Nezzar, and L. Vese. A multiscale image representation using hierarchical (BV,L2) decompositions. SIAM Multiscale Modeling and Simulation, 2(4):554–579, 2004.
55. Ulrike Von Luxburg. A tutorial on spectral clustering. Statistics and computing, 17(4):395–416, 2007.
56. J. Weickert. Anisotropic Diffusion in Image Processing. Teubner-Verlag, Stuttgart, Germany, 1998.
57. M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Applied and Computational Harmonic Analysis, 24(2):195–224, 2008.
58. Wotao Yin, Donald Goldfarb, and Stanley Osher. The total variation regularized l^1 model for multiscale decomposition. Multiscale Modeling & Simulation, 6(1):190–211, 2007.