A Signal Processing Approach for Generalized 1-D Total Variation

Sovit Kumar Pradhan [1], Swadhinata Bhat [2], Bijaya Kumar Mohapatra [3], Chittaranjan Biswal [4], Ritu Tripathy [5]

DILIP KUMAR NAYAK[6]

[1, 2, 3, 4, 5] Gandhi Institute for Education & Technology, Baniatangi, Khordha, Odisha

[6]NM Institute of Engineering & Technology,Bhubaneswar,Odisha

sovitkumarpradhan@giet.edu.in, swadhinatabhat@giet.edu.in, bijayakumarmohapatra@giet.edu.in,
chittaranjanbiswal@giet.edu.in, ritutripathy@giet.edu.in

Abstract—Total Variation (TV) is a powerful method that brings great utility to edge-preserving regularization. Although widely used in image processing, its applicability to one-dimensional signal processing is limited because piecewise-constant signals form a rather restricted model in many applications. Here we generalize classical TV to one dimension by extending the differential operator found in the regularization term to any linear differential operator. This allows for the presence of significant linear systems and the flexibility to adapt the approach to different types of drive signals. Examples: lacy, piecewise constants, etc. Conventional television remains a special case of these framework conditions. We show the feasibility of this method by considering a linear system of interest and different types of driving signals.

Index Terms—Differential operators, linear systems, regulariza- tion, sparsity, total variation.

I. INTRODUCTION

TIKHONOV regularization is one traditional way in en- gineering and statistics to deal with ill-posed problems; i.e., the solution should provide a balance between the data term (consistency with the measurements) and the regulariza- tion term (based on prior information). Generalizations of this scheme employ the energy of the solution's image under a linear operator instead of the energy of the solution [1]. Such regular- ization strategies typically lead to smooth solutions when em- ploying ℓ_2-norm. Alternatively, ℓ_1-norm regularizations have attracted a lot of attention for their sparsity-promoting property; i.e., they favor sparse representations with many zeros or very small coefficients. The ℓ_1-norm is the relaxed convex counter- part of the ideal sparsity-promoting, but nonconvex, ℓ_0 spar- sity count (the number of nonzero components) [2]. A particular Identifier case of ℓ_1-norm based regularization, total variation (TV), has been widely applied in image processing [3], including applica- tions like image denoising, restoration and deconvolution [4], [5]. In 2-D, TV is defined as the ℓ_1-norm of the magnitude of the gradient. In this respect, TV regularization preserves edge information and hence offers superior approximation quality for most real-world data compared with (smooth) ℓ_2-based regular- ization. In a typical denoising problem, the aim is to recover an object , given its noisy measurements . If the object is known to be sparse in a given representation, this information can be used to come up with a regularization strategy. In this regard, promoting "sparsity" of the analysis or synthesis coeffi- cients leads to two different regularization strategies [6]. More precisely, if denotes the synthesis operator of a frame, "syn- thesis prior" denoising requires the solution to satisfy

$$\dot{\mathbf{x}} = \mathbf{T}\dot{\mathbf{c}}, \text{ where } \ddot{\mathbf{c}} = \arg\min_{\mathbf{c}} \|\mathbf{y} - \mathbf{T}\mathbf{c}\|_2^2 + \|\mathbf{c}\|_p^n \quad (1)$$

where for ℓ_1-constrained fitting regression (see, for ex- ample, [7]–[11]). On the other hand, the "analysis prior" formu- lation looks for the solution of

$$\dot{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \|\mathbf{F}\mathbf{x}\|_p^n \quad (2)$$

where denotes the analysis operator of the frame. For or- thonormal bases, the two approaches yield the same result; for redundant frames, even tight ones, they do not. In this setting, Tikhonov and TV regularized formulations of the denoising problem can be regarded as generalizations (since the associ- ated operators are not necessarily the analysis operators of a frame anymore) of the analysis prior formulation, which rely on the ℓ_2 and the ℓ_1-norm, respectively.

TV exploits the combination of a derivative operator with the ℓ_1-norm, which makes this regularization an ideal option when dealing with piecewise-constant signals. However, in many 1-D applications much more complex signals are encountered. Piecewise-polynomial signals can be dealt with by TV with built-in higher order derivatives [12]–[16]. For example, TV with the second-order derivative is optimal for piecewise-linear signals. The benefit of adding higher order derivatives is that they reduce the staircase effect of the first order derivative in TV and bring flexibility to the reproduced signal. This exten- sion has been used to retain smooth transitions while keeping sharp edges [16], for texture extraction [17], and has recently been reintroduced for MRI reconstruction [18]. The early appli- cation of TV regularization has been hampered by rather high computational complexity, but the recent advances in the field have led to fast algorithms for the ℓ_1 minimization problem. Today, the algorithms at hand range from dual methods [19], it- erative reweighted norm (IRN) [20], fixed point algorithm [21], graph cuts [4] to (fast) iterative shrinkage-thresholding algo- rithm ((F)ISTA), and so on [5], [22], [23]. It is important to note that basis-pursuit denoising [7] can offer a "synthesis" counterpart of TV; i.e., inverting the derivative operator leads to a dictionary with (shifted) Heaviside step-functions. There- fore, analysis-prior and synthesis-prior methods seem to be equivalent from a theoretical point-of-view, but the represen- tation by the Heaviside dictionary is unstable. For example, in [24], the analysis counterpart of the constructed dictionary is preferred for its simplicity. In some cases, a specific dictionary can be built for the synthesis operator [25], [26], including generalizations of wavelet design [27], [28].

Here, we extend the conventional TV concept and adapt the regularization for any linear differential operator. The choice of the operator can be tuned to the presence of a linear system and can be adapted to the type of the driving signal that is ex- pected. We propose generalized TV regularization for the typ- ical setting in signal processing dealing with uniformly sampled data. We also provide two practical algorithms to solve the op- timization problem at hand.

The paper is organized as follows. We elaborate conventional TV as a mathematical principle and then introduce our gener- alized TV scheme in Section II. Section III is devoted to the algorithms and how to modify these general-purpose solvers for generalized TV regularization. In Section IV we simulate a system of third-order linear differentials and compare the per- formance of proposed generalized TV and conventional TV. Finally, we comment on the numerical results and draw some conclusions.

II. GENERALIZED TOTAL VARIATION

A. Beyond Conventional TV

We introduce a new regularization term that is inspired from the combination of TV with linear differential operators. We first revisit conventional TV, which forms the fundamental mathematical concept behind our methodology.

Definition 1 (Total Variation): TV of a continuous-domain function $x(t)$ on interval $[a, b]$, is defined as the supremum of absolute differences for any finite partition $P = \{t_0 = a, \dots, t_n, \dots, t_m = b\}_n$:

$$\text{TV}\{x\} = \sup_P \sum_{n=1}^{m} |x(t_n) - x(t_{n-1})| \qquad (3)$$

If the first derivative of $x(t)$ is well defined, then TV can be shown to be equivalent to

$$\text{TV}\{x\} = \int_a^b |D\{x\}(t)| \, dt \qquad (4)$$

where is D regular (continuous-domain) derivative [4].

For many applications in signal processing, the data is acquired and available on a uniform grid. Specifically, we consider the series of samples $\mathbf{x} = [x(t_n)]_{n \in \mathbb{Z}}$. In such a case, TV can be computed as the ℓ_1-norm of the finite-difference operator

$$\text{TV}\{\mathbf{x}\} = \sum_{n \in \mathbb{Z}} |\Delta\{\mathbf{x}\}[n]| \qquad$$

where $\Delta\{\mathbf{x}\}[n] = x[t_n] - x[t_{n-1}]$. This point of view clearly reveals the link between the formulations (3) and (4); i.e., the finite difference is Δ the discrete counterpart of the continuous-

domain derivative D. In fact, in the case of uniform discrete data as will be considered further on, the concept of TV can seamlessly be extended for discrete filters that are associated to general linear differential operators.

Definition 2 (Generalized Total Variation): For a discrete signal $\mathbf{x}$ we define the generalized TV as

$$\text{TV}\{\mathbf{x}\} = \sum_{n \in \mathbb{Z}} |\Delta_L\{\mathbf{x}\}[n]| \qquad (5)$$

where Δ_L[1] is the discrete version of the Nth-order differential operator

$$L = \prod_{i=1}^{N} (D - \alpha_i I) \left(\prod_{i=1}^{M} (D - \gamma_i I) \right)^{-1} \qquad (6)$$

with I is the identity operator, $\alpha_i \in \mathbb{C}$, $i = 1, \dots, N$, and $\gamma_i \in \mathbb{C}$, $i = 1, \dots, M$, the zeros and poles of the operator, respectively, and $M < N$. We conveniently characterize the operator by $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_M)$. This operator can also be used to build exponential B-splines [29], [30] and wavelets [27].

Clearly, the definition in (5) includes conventional TV; i.e., the case $N = 1$ with $\alpha_1 = 0$ and $M = 0$ reverts to that case. It should also be noted that for $M = 0$ the discrete operator Δ_L has finite length $(N + 1)$. For $M > 0$, the support of the discrete operator Δ_L becomes infinite in general and the operator can be constructed by a proper combination of causal and anti-causal filtering depending on the poles γ_i of the system. The following proposition summarizes how to obtain the discrete counterpart Δ_L of L The proof can be found in the Appendix.

Proposition 1 (Discrete Implementation of L): Consider the continuous-domain linear differential operator $L = L_n L_d^{-1}$, where L_n and L_d constitute the operators of ; the operator L_d is inverted as before. Therefore,

$$L_n = \prod_{i=1}^{N} (D - \alpha_i I)$$

$$L_d = \prod_{i=1}^{M} (D - \gamma_i I).$$

We separate L_d in its causal (L'_d, characterized by $\boldsymbol{\gamma}' = (\gamma'_1, \dots, \gamma'_{m_1})$, $\text{Re}(\gamma'_k) < 0$) and anti-causal (L''_d, $\boldsymbol{\gamma}'' = (\gamma''_1, \dots, \gamma''_{m_2})$, $\text{Re}(\gamma''_k) > 0$) parts to assure stability. Then, the discrete operator Δ associated to can be obtained by a cascade of filtering operations corresponding to

$$\mathbf{y}'[n] = \mathbf{x}[n] - \sum_{k=1}^{m_1} \mathbf{y}'[n-k] \Delta_{L'_d}[k],$$

$$\mathbf{y}''[n] = \mathbf{y}'[n + m_2] \Delta_{L''_d}[-m_2] - \sum_{k=-m_2}^{-1} \mathbf{y}''[n-k] \Delta_{L''_d}[k],$$

$$\mathbf{y}[n] = \sum_{k \in \mathbb{Z}} \mathbf{y}''[n-k] \Delta_{L_n}[k] \qquad (7)$$

where x is the input, y' is the output of the first recursive filtering step (causal), y'' is the output of the second recursive
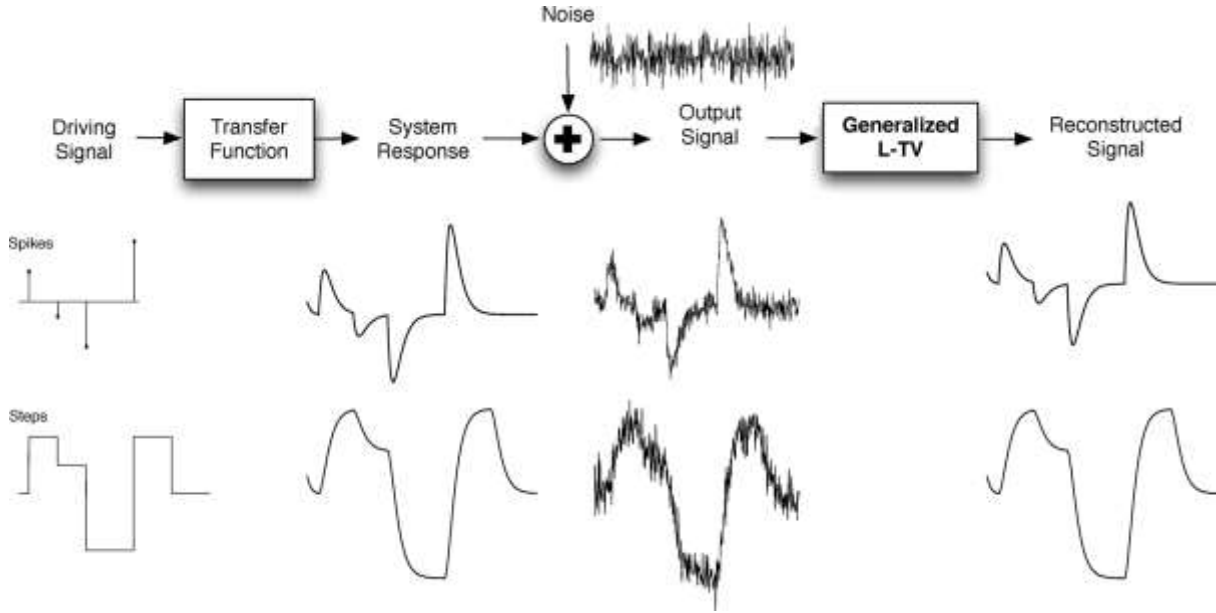
Fig. 1. Illustration showing the observation model and the use of generalized TV. First, the driving signal serves as input to the linear system. Next, the (ideal) system response gets corrupted by noise. Generalized TV regularization aims at minimizing the ℓ_1-norm of the differential operator—which is tuned to the linear system and driving signal—applied to the reconstruction.

filtering step (anti-causal) and $\mathbf{y}$ is the final output. The constituting filters are given in the time domain by

$$\Delta_{L_n}[n] = (-1)^n \sum_{|\mathbf{m}|=n} (e^{\boldsymbol{\alpha}})^{\mathbf{m}}, \mathbf{m} \in [0,1]^N, 0 \le n \le N$$

$$\Delta_{L_d'}[n] = (-1)^n \sum_{|\mathbf{m}|=n} (e^{\gamma'})^{\mathbf{m}}, \mathbf{m} \in [0,1]^{m_1}, 0 \le n \le m_1$$

$$\Delta_{L_d''}[n] = (-1)^n \sum_{|\mathbf{m}|=-n} (e^{-\gamma''})^{\mathbf{m}}, \mathbf{m} \in [0,1]^{m_2}, -m_2 \le n \le 0$$

where we use the multi-index notation $\mathbf{m} = (m_1, \ldots, m_N)$, with $|\mathbf{m}| = \sum_{k=1}^{N} m_k$ and the conventions $c^{\mathbf{m}} = (c^{m_1}, \ldots, c^{m_N})$ and $\mathbf{c}^{\mathbf{m}} = \prod_{k=1}^{N} c_k^{m_k}$. The initial conditions for the recursive implementations (the last two rows of (7)) can be determined with respect to the input $\mathbf{x}$; i.e., $\mathbf{y}'$ and $\mathbf{y}''$ are nonzero within the support of $\mathbf{x}$

B. TV Regularization

As usual for TV regularization, we cast our problem of finding the approximation $\dot{\mathbf{x}}$ from the noisy measurements $\mathbf{y}$ into a variation formulation where we use the generalized TV regularizer from (5). Then, the minimization problem becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \underbrace{\frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\|\Delta_L\{\mathbf{x}\}\|_1}_{J(\mathbf{x})} \quad (8)$$

where λ is the regularization tuning parameter and Δ_L is the discretized form of a differential operator of the form (6). The solution depends on λ which controls the trade off between the data and regularization terms.

The optimal solution to (8) should find a compromise between data fidelity and regularization cost. In the continuous domain, any homogeneous solution of the differential operator L has no cost since $L\{x_0\}(t) = 0$. This property also holds for the sampled version $\mathbf{x}_0$ and the discrete filter Δ_L; i.e., $\Delta_L\{\mathbf{x}_0\} = 0$. Therefore, null-space components of the differential operator can be used at no cost to minimize the residual

error. Moreover, signals with a sparse representation after applying Δ_L have a low cost for the ℓ_1-norm and will be preferred. For the case $L = D$, we retrieve the conventional TV regularization where the constant signal is the (only) null-space component and the lowest global cost is attained for piecewise-constant signals.

C. On the Choice of Differential Operator L

The main working principle of TV is to impose sparsity through the ℓ_1-norm of the derivative of the signal, which is typically satisfied for piecewise-constant signals when combined with the data term. The additional flexibility that we propose by generalizing for any differential operator allows incorporating additional or different prior information.

When dealing with a linear system in Fig. 1, the differential operator can be tuned to the inverse of the system response. In this case, the differential operator will undo the effect of the linear system, and regularization will be guided by the driving signal. The ℓ_1-norm leads to the optimal performance when the signal is spike-like. More complex driving signals can be dealt with by further refining the operator; e.g., for a step-like driving signal a regular derivative can be added to the regularizing operator. We will discuss this point in more detail in Section IV.

III. OPTIMIZATION ALGORITHMS

In what follows, we briefly review two practical algorithms to obtain the minimizer of the cost functional $J(\mathbf{x})$ in (8). For now, we assume that all vectors and matrices involved are finite-dimensional. This minimization problem can be regarded as a generalized form of the analysis-prior denoising problem (see [6] and [26] and the references therein). The first algorithm is essentially an adaptation of the one provided by Chambolle for TV denoising [5], [31] (also see [19] for a slightly different algorithm). In [5], the (fast) gradient projection ((F)GP) algorithm is derived for TV denoising problems ($\Delta_L = \Delta$) whereas we adapted the algorithm for the discrete operator Δ_L.

Algorithm 1: Let us consider the regularization operator Δ_L; its z-transform is given by $\hat{\Delta}_L(z) := \sum_n \Delta_L[n] z^{-n}$ and $z = e^{j\omega}$ reverts to the discrete time Fourier transform of $\Delta_L[n]$. Further more, we choose c subject to

$$c > \sup_\omega \left| \hat{\Delta}_L(e^{j\omega}) \right|^2 = \sup_\omega \frac{\prod_{i=1}^N |1 - e^{\alpha_i} e^{-j\omega}|^2}{\prod_{i=1}^M |1 - e^{\gamma_i} e^{-j\omega}|^2}.$$

The algorithm then iterates the following steps until convergence:

I) Update $\mathbf{z} := \Delta_L\{\mathbf{y}\}/(\lambda c) + (I - \Delta_L \Delta_L^T/c)\{\mathbf{z}\}$, where the adjoint of Δ_L is $\Delta_L^T[n] = \Delta_L^*[-n]$.
II) Update $\mathbf{z} := \text{clip}(\mathbf{z})$ where $\text{clip}(\cdot)$ denotes the element-wise clipping function,

$$\text{clip}(\cdot) = \text{sign}(\cdot) \min(|\cdot|, 1).$$

After convergence, set $\mathbf{x} = \mathbf{y} - \lambda \Delta_L^T\{\mathbf{z}\}$.

With a small variation only, the fast version of the algorithm can be obtained [5].

Algorithm 2: Repeat the following steps until the convergence criterion is met for each ith iteration with $k_1 = 1$:

I) Update $\mathbf{z}_i := \Delta_L\{\mathbf{y}\}/(\lambda c) + (I - \Delta_L \Delta_L^T/c)\{\mathbf{v}_i\}$.
II) Update $\mathbf{z}_i := \text{clip}(\mathbf{z}_i)$.
III) Update $k_{i+1} = \frac{1 + \sqrt{1 + 4k_i^2}}{2}$.
IV) Update $\mathbf{v}_{i+1} = \mathbf{z}_i + \frac{k_i - 1}{k_{i+1}}(\mathbf{z}_i - \mathbf{z}_{i-1})$.

After convergence, set $\mathbf{x} = \mathbf{y} - \lambda \Delta_L^T\{\mathbf{z}\}$.

These algorithms make use of the dual-problem [19] and employs the "fast gradient-projection" (FGP) method [5]. Equivalent algorithms can be derived employing the "majorization-minimization" scheme (see [32]–[34]). Again initiating from the dual form, Chambolle [19] proposes a slightly different gradient-based model; finally, similar algorithms [23], [31] suggest "graph-cut" technique to solve the same problem. We refer to [35] for an extensive overview of recent algorithms.

It is important to note that generalized TV can be combined with the more general deconvolution problem: the data term may include a matrix/operator $\mathbf{A}$ e.g., that corresponds to a blurring filter. Then the optimization problem turns into

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\Delta_L\{\mathbf{x}\}\|_1$$

The solution of this problem is not very practical to achieve directly with Algorithms 1 and 2 since the inverse of $\mathbf{A}^T\mathbf{A}$ may be ill-conditioned. The trick is to apply a two step optimization. The first (outer) loop tackles the deblurring problem while the second (inner) loop solves the denoising problem as proposed here. For more details, we refer to [5 Sec. 5] for combining (F)ISTA/FGP algorithms, and to [36] for the "majorization—minimization" algorithm.

IV. EXPERIMENTAL RESULTS

In this section, we present several examples to illustrate the usefulness and performance of the proposed method. First, we demonstrate signal reconstruction of a simulated linear system response with different driving signals and show how it outperforms conventional TV. Finally, we show waveform analysis of audio signals by tuning the zeros of the operator to the central tone.

TABLE I
THE DIFFERENTIAL OPERATOR L FOR GENERALIZED TV NEEDS TO BE CHOSEN ACCORDING TO THE LINEAR SYSTEM AND THE TYPE OF DRIVING SIGNAL. I IS THE IDENTITY OPERATOR. THE INVERSE IS DEFINED AS EXPLAINED IN SECTION IV-A

driving signal	linear system	optimal TV operator L
spikes	I	I
piecewise constant	I	D (conventional TV)
spikes	$\tilde{h} = \tilde{L}^{-1}\{\delta\}$	$\tilde{L}$
piecewise constant	$\tilde{h} = \tilde{L}^{-1}\{\delta\}$	$D\tilde{L}$
piecewise linear	$\tilde{h} = \tilde{L}^{-1}\{\delta\}$	$D^2\tilde{L}$

A. Proof of Concept

We present synthetic examples according to the model shown in Fig. 1 to demonstrate the usefulness of generalized TV regularization. To focus the attention, we consider a linear system

$$\tilde{h} * s = x \tag{9}$$

where s and x are the input and output signals, respectively, and $*$ denotes the convolution operator. We introduce the continuous Fourier transform of $\tilde{h}(t)$ as $\hat{\tilde{h}}(\omega) := \int_t \tilde{h}(t)e^{-j\omega t}dt$.

We can define a pseudoinverse $\tilde{L}\{x\} = \check{g} * x$, where $\check{g}$ is defined by

$$\hat{\check{g}}(\omega) = \begin{cases} 1/\hat{\tilde{h}}(\omega), & \text{if } \hat{\tilde{h}}(\omega) \neq 0 \\ \hat{\tilde{h}}(\omega) = 0, & \text{if } \hat{\tilde{h}}(\omega) = 0. \end{cases}$$

Then, matching the regularization operator L with $\tilde{L}$ will cancel the effect of the linear system:

$$L\{x\} = L\{\tilde{h} * s\} = \check{g} * \tilde{h} * s = s + s_{\text{null}}$$

where s_{null} is a null-space component of the system $\tilde{h}$; in other words, $\tilde{h} * s_{\text{null}} = 0$. Notice that a null-space component of L cannot be recovered.

Depending on the assumptions on the driving signal s, we have the additional freedom to include derivative operators in L in order to further sparsify the signal and make the ℓ_1-norm effective. Specifically, in Table I, we give an overview of how the operator L should be chosen for various types of driving signals; e.g., spikes, piecewise-constant, and piecewise-linear.

We now illustrate these concepts by considering a third-order system driven by a spike-like signal. In this case, the Fourier domain of the impulse response of the system has the form

$$\hat{\tilde{h}}(\omega) = \frac{(j\omega - \hat{\gamma}_1)}{(j\omega - \hat{\alpha}_1)(j\omega - \hat{\alpha}_2)(j\omega - \hat{\alpha}_3)}$$

with three poles and one zero. Consequently, the differential operator $\tilde{L}$ can be characterized in its turn in the Fourier domain as the inverse of $\tilde{h}$

$$\hat{\tilde{L}}(\omega) = \frac{(j\omega - \hat{\alpha}_1)(j\omega - \hat{\alpha}_2)(j\omega - \hat{\alpha}_3)}{(j\omega - \hat{\gamma}_1)}$$

where the system's poles take the role of the operator's zeros (and vice versa). The time-domain operator then corresponds to

$$\tilde{L} = (D - \hat{\alpha}_1 I)(D - \hat{\alpha}_2 I)(D - \hat{\alpha}_3 I)(D - \hat{\gamma}_3 I)^{-1}.$$

For a practical example, we take $\hat{\alpha} = \hat{\alpha}_1 = \hat{\alpha}_2 = \hat{\alpha}_3 = -2$ and $\hat{\gamma} = -0.1$.
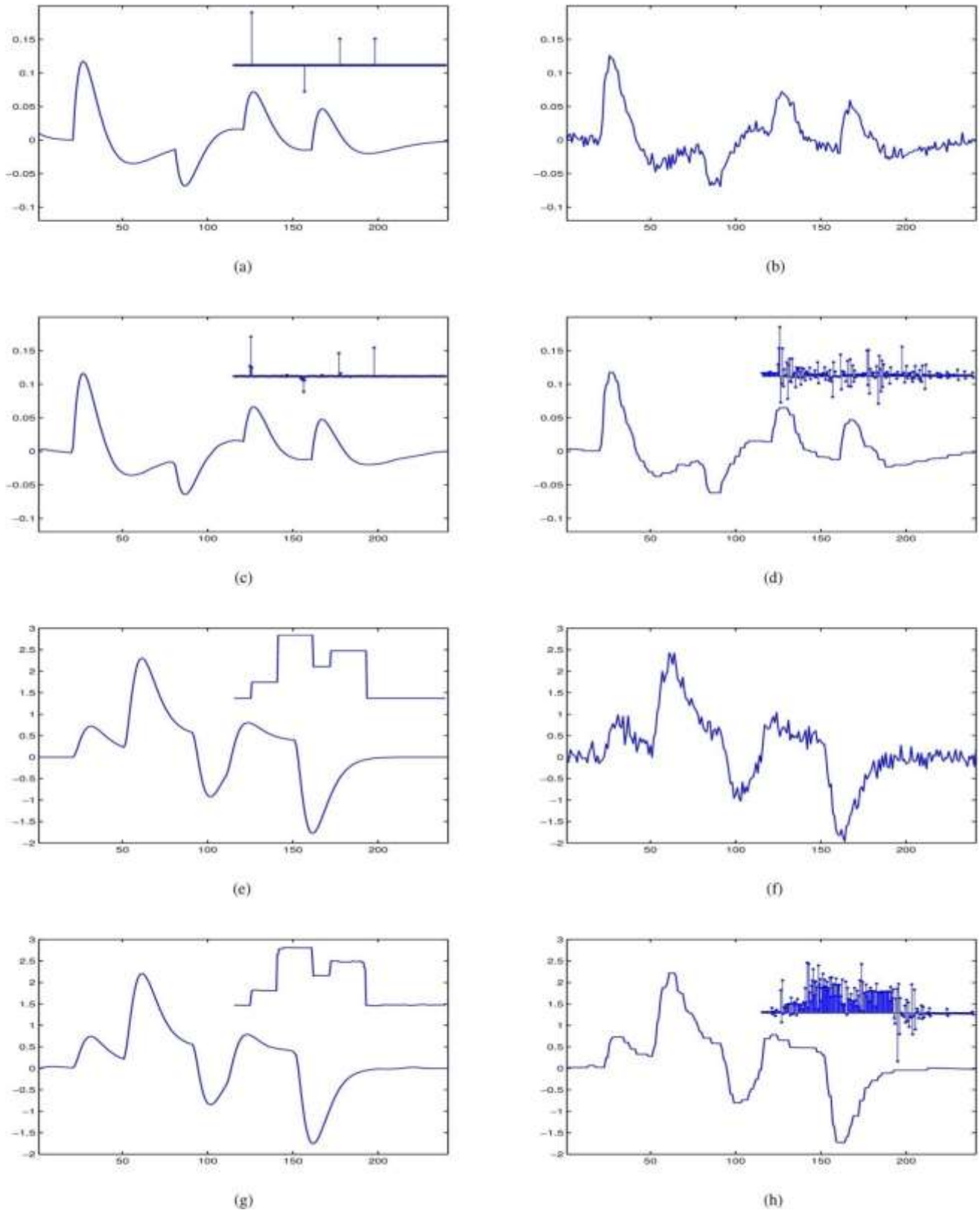
Fig. 2. Generalized TV and conventional TV for a third-order linear system (with three poles and one zero) driven by spike-like signal (first two rows) and piecewise-constant signal (last two rows). (a) Original signal (inset: driving signal); (b) noisy signal; (c) reconstruction using generalized TV regularization; (d) reconstruction using conventional TV regularization; (e) original signal (inset: driving signal); (f) noisy signal; (g) reconstruction using generalized TV regularization; (h) reconstruction using conventional TV regularization.

Since we assume the driving signal to be spike-like, we select $L = \tilde{L}$, for which the discrete version $\Delta_{L_\circ}$ of the

forward operator corresponds to an FIR filter with four taps: $[1, -3e^\alpha, 3e^{2\alpha}, -e^{3\alpha}]$ and the discrete version of the inverse op-

TABLE II
OVERVIEW OF THE PERFORMANCE MEASURED AS SNR (dB). THE OPTIMAL TUNING PARAMETER λ WAS DETERMINED USING AN ORACLE. AVERAGE SNR AND ITS STANDARD DEVIATION ARE REPORTED FOR 100 REALIZATIONS OF THE NOISE

	SNR (dB)	λ	SNR (dB)	λ	SNR (dB)	λ
noisy signal 1	5	—	10	—	20	—
oracle Wiener	14.09 ± 0.87	—	17.45 ± 0.65	—	24.31 ± 0.41	—
conventional TV	12.32 ± 0.68	0.32 ± 0.04	15.89 ± 0.63	0.16 ± 0.05	22.15 ± 0.46	0.10 ± 0.01
generalized TV	14.86 ± 1.96	0.82 ± 0.38	19.58 ± 2.22	0.52 ± 0.25	28.71 ± 1.92	0.17 ± 0.09
noisy signal 2	5	—	10	—	20	—
oracle Wiener	14.17 ± 0.88	—	18.87 ± 0.84	—	26.90 ± 0.59	—
conventional TV	13.08 ± 0.72	1.96 ± 0.36	16.70 ± 0.64	0.98 ± 0.17	23.99 ± 0.41	0.22 ± 0.05
generalized TV	14.83 ± 1.68	10.02 ± 1.86	19.78 ± 1.86	4.90 ± 1.49	29.35 ± 1.60	1.70 ± 0.63

erator corresponds to a causal IIR filtering with $\Delta_{L'} = [1, -e^\gamma]$. In Fig. 2(a), we show the noiseless original signal, the spike-like driving signal in the inset. We also added a random null-space component of the operator $\check{L}$ (i.e., $c_1 e^{\bar{\alpha}t} + c_2 t e^{\bar{\alpha}t} + c_3 t^2 e^{\bar{\alpha}t}$) as a "background" to which the regularization term is insensitive. Next, the signal is corrupted by i.i.d. Gaussian noise 15 dB [see Fig. 2(b)] and approximated using two regularization methods: generalized TV and conventional TV as shown in Fig. 2(c) and (d), respectively. For each method the oracle selected the optimal regularization parameter λ since it has access to the groundtruth. We also compute the filtered version Δ_L of the regularized solution in order to explore how well we reconstruct the underlying driving signal, see the insets in Fig. 2(c) and (d). The choice of the regularization parameter λ is important since the oracle is not available in practice. Different strategies can be adapted for selecting λ (see [37] for classical references to methods such as generalized cross-validation and the L-curve, or [38] for a recent Monte Carlo adaptation of Stein's unbiased risk estimate that works well for TV, and [19], [39]–[42]).

As an additional experiment, we change the driving signal into piecewise-constant, as shown in Fig. 2. Accordingly, we adapt the regularizing operator into $L = D\check{L}$, which only leads to an alteration of the FIR filter, $\Delta_{L_c} = [1, -1 - 3e^\alpha, 3e^{2\alpha} + 3e^\alpha, 3e^\alpha - e^{3\alpha}, e^{3\alpha}]$. The results for the generalized TV and conventional TV are shown in Fig. 2(g) and (h), respectively.

The regularization in (8) is performed using the algorithm described in Section III. We deploy three different noise settings: additive Gaussian noise corresponding to SNR level of 5, 10, and 20 dB. We report average SNR levels with standard deviation (over 100 realizations), with the optimal regularization parameter and maximum SNR, in Table II. We compare the reconstruction quality obtained by oracle-Wiener filtering (optimal for a Gaussian-process driving signal corrupted with additive Gaussian noise), conventional TV, and generalized TV. As expected, the results reveal that generalized TV is superior to conventional TV and the Wiener filter when tuning the operator L different from D is appropriate. Moreover, as it can be appreciated from the corresponding Figures, also the reconstruction of the underlying driving signal has high quality and can be useful for further processing in applications.

An important concern is the robustness of the choice of the regularization operator L with respect to the underlying "true" linear system. To that aim, we generate signals for a spike-like input of a third-order linear system $\check{L} = (D - \bar{\alpha}I)^3(D - \gamma I)^{-1}$ with $\bar{\alpha}$ in the range [-2, 2]. We create 10 different realizations corrupted by additive Gaussian noise resulting into 10 dB SNR. Next, we apply several regularization strategies: 1) generalized TV with α tuned exactly to the system; 2) generalized
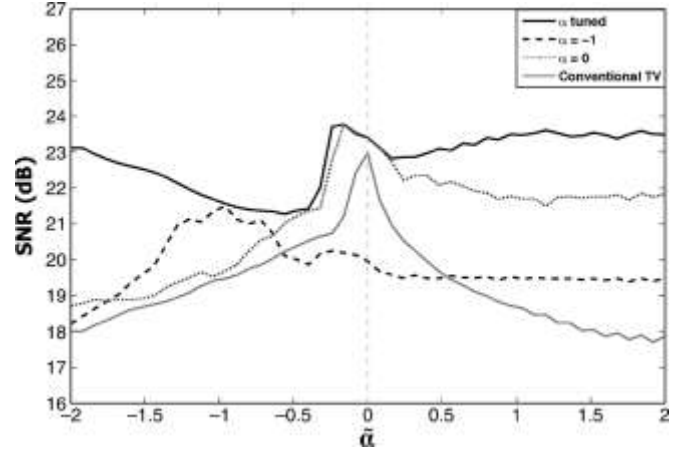

Fig. 3. Performance measured as SNR (dB) for generalized TV and conventional TV regularization of a third-order linear system with equivalent differential operator $\check{L} = (D - \bar{\alpha}I)^3(D - \gamma I)^{-1}$, with $\gamma = 4$, for varying $\bar{\alpha}$ values. The reported SNR measures are averaged over 10 realizations of additive Gaussian noise (noisy signal 10 dB).

TV with $\alpha = -1$; 3) generalized TV with $\alpha = 0$ to illustrate second-order TV; and 4) conventional TV. We kept $\gamma = 4$ constant to eliminate pole-zero cancellation. In Fig. 3, we plot the average SNR (10 regularizations) for different $\bar{\alpha}$ values of the linear system. As expected, we observe that second-order TV and matched generalized TV have equal performance at $\bar{\alpha} = 0$. Similarly, generalized TV with fixed $\alpha = -1$ meets matched generalized TV at $\bar{\alpha} = -1$. Moreover, we notice that SNR levels for matched generalized TV tend to increase further for larger values of $\bar{\alpha}$. The Green's function of the differential operator $\check{L}$ increases which generates high correlation between the samples. Conventional TV underperforms as the derivative operator is not well tuned to the linear system.

B. Sound Waveform Processing

We show that it is possible to tune the operator of generalized TV to include information about modulation, which can be useful for audio signals; e.g., processing tonal and transients layers [43].

Specifically, let us assume the simplified signal model as a sum of shifted decaying exponentials, each one modulated by a high-frequency sinusoidal function:

$$y(t) = \sum_i A_i \sin(\omega_{0,i}(t - t_i)) e^{\alpha(t-t_i)} u(t - t_i) \qquad (10)$$

where α is the decay rate. (10) can be considered as the sum of responses of linear systems with impulse responses $\sin(\omega_{0,i}t)e^{\alpha t}u(t)$ for spikes $A_i\delta(t - t_i)$, respectively. Here we
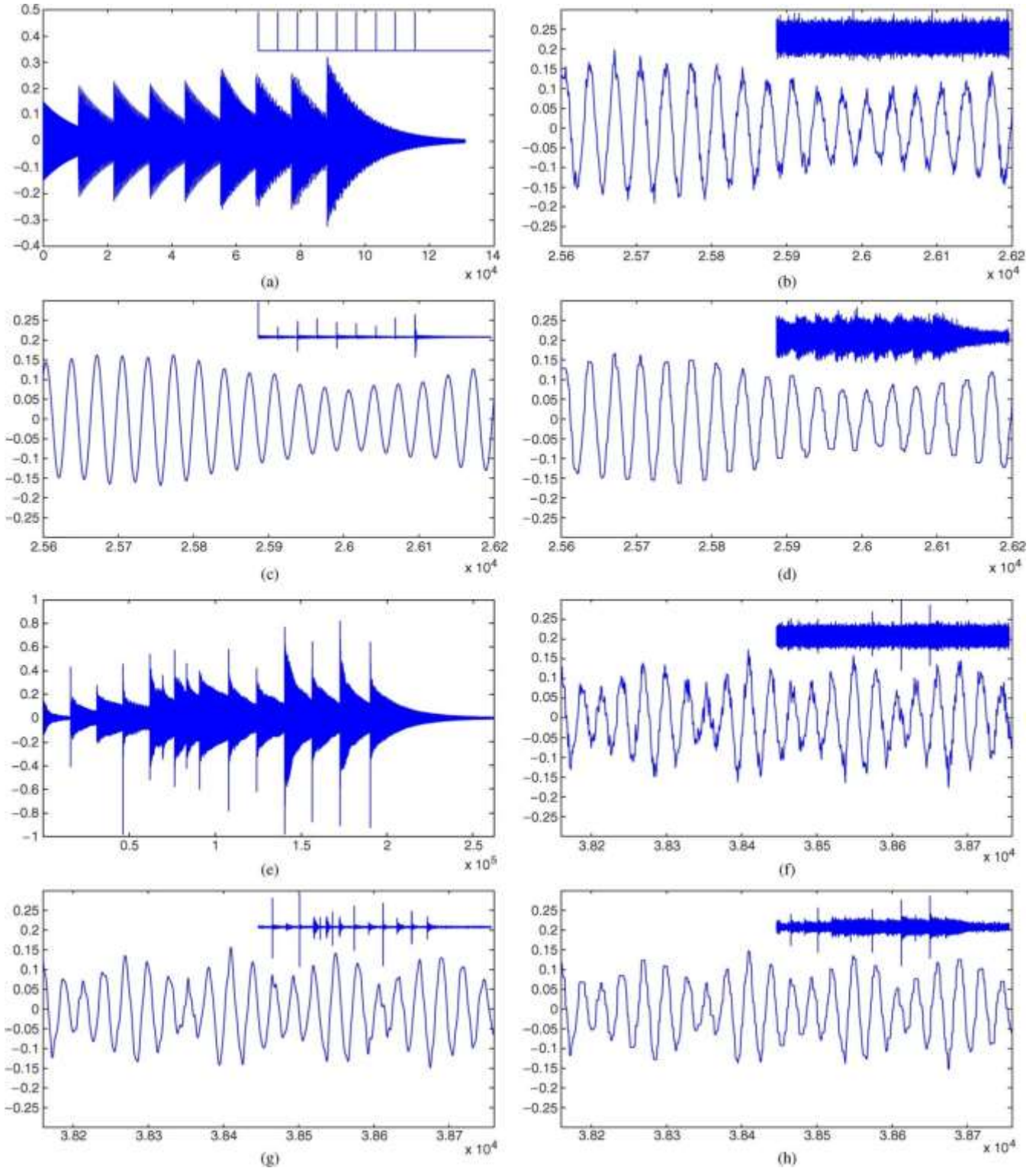
Fig. 4. Zoom of the denoised audio signal, Für Elise (first two rows) and Glockenspiel signal (last two rows) corrupted with i.i.d.\ Gaussian noise 15 dB with generalized TV and conventional TV. We employ differential operator $L = (D + (j\omega_0 + \alpha)I)(D - (-j\omega_0 + \alpha)I)$ and $L = ((D + (j\omega_0 + \alpha)I)(D + (-j\omega_0 + \alpha)I))$ for Für Elise and Glockenspiel signal, respectively, where ω_0 is the average frequency and $\alpha = -4\,s^{-1}$. (a) Original signal; (b) noisy signal; (c) reconstruction using generalized TV regularization; (d) reconstruction using conventional TV regularization; (e) original glockenspiel signal; (f) noisy signal; (g) reconstruction using generalized TV regularization; and (h) reconstruction using conventional TV regularization.

pick one (central) frequency ω_0, for which the corresponding transfer function is

$$\hat{h}(w) = \frac{\omega_0}{(jw + (j\omega_0 + \alpha))(jw + (-j\omega_0 + \alpha))} \quad (11)$$

and the associated second-order differential operator becomes $L = (D + (j\omega_0 + \alpha)I)(D + (-j\omega_0 + \alpha)I)$.

We generated a synthetic signal [Fig. 4(a)] according to (10) with the first 9 notes of Für Elise whose frequencies range from 329–1318 Hz and at sampling frequency $f_s =$ 44 100 Hz. The

decay rate was $\alpha = -4\,s^{-1}$. We used the average frequency of the notes and $\alpha = -4\,s^{-1}$. The effect of tuning α is negligible on the output since the decay is very slow compared to the sampling frequency. We created a noisy realization of the audio signal by corrupting it with i.i.d. Gaussian noise at SNR 15 dB, see Fig. 4(b). We show the output of generalized (25.3 dB) and conventional TV (18.3 dB) in 4(c) and (d), respectively. The regularization parameter λ was chosen using an oracle.

In the supplementary material, we have provided audio files of the example given above; i.e., original, noisy, denoised using conventional and generalized TV, and the spike driving signal can be listened. We also include another example using the real-world test signal "Glockenspiel" [43], where we increase the multiplicity of the zeros to make the frequency response of L increasingly flat around ω_0, and thus cancel also tones with nearby frequencies.[2] Finally, in Fig. 4(e), we show the "Glockenspiel" signal, which is directly adopted from [43]. We degraded the signal with additive i.i.d. Gaussian noise at SNR $=$ 15 dB; see Fig. 4(f). The output for generalized TV (18.3 dB) and conventional TV (16.67 dB) is shown in 4(g) and (h), respectively. Note that the SNR values are computed against the real "Glockenspiel" signal, which contains some noise itself and consequently the comparison with the original signal is not ideal. First, the (noiseless) ground truth is not known. Second, the signal contains different frequency components from perturbed harmonics in the tonal layer and sharp transitions in the transient layer, neither of which are modeled by the operator in generalized TV—nevertheless, the result is still better than for conventional TV. More advanced applications of generalized TV for sound wave processing can be devised in future, such as the inclusion of multiple regularization terms with different operators each (e.g., for different frequencies and harmonics) and an additional model to deal with the transient layer.

V. CONCLUSION

TV regularization has become a widely applied scheme with edge-preserving properties. Extensions have mainly focused on the use of higher order derivatives [12]–[17] and recently also on nonlocal generalizations [44]. In this paper, we extended the basic TV concept further by introducing a general differential operator L instead of the derivative D. This allows a great deal of flexibility since we can take into account the presence of a linear system and different types of driving signals. From digital signal processing view, we build the simplest discrete approximation of the differential operator L and we define them explicitly as FIR/IIR filters. Generalized TV scheme is appealing for many signal processing applications, in particular, when the system can be expressed or approximated in terms of differential operators. We illustrated our framework by considering a third-order linear system. In order to solve the regularization problem, we proposed an alternative algorithm based on the dual form of TV. This work can be considered as the analysis prior counterpart of exponential spline wavelets [27] or generalized Daubechies wavelets [28]; indeed, these wavelets are generalizations that can be tuned to a given differential operator and their use in regularized reconstruction

concurs with a synthesis prior. Despite the fact that our approach is heavily inspired on the continuous domain, future research is needed to tighten the mathematical link between the proposed signal processing approach (in the discrete domain) and proper generalization of TV in the continuous domain. We believe that our approach is promising in this respect because in recent work [45] it was shown that the signal-processing approach for conventional TV (that is, ℓ_1-norm of finite differences) can be linked to proper continuous-domain modeling of stochastic processes.

Finally, we mention that another application of generalized TV could be to perform system identification; i.e., when strong assumptions on the (sparsity of the) driving signal can be made, one can attempt to retrieve the form of L and, for model selection, evaluate the performance of the estimated operator by statistical methods such as Akaike's information criterion and its extension for the state-space representation of linear systems [46].

APPENDIX
PROOF OF PROPOSITION 1

Proof: We make the proof by construction. For the first-order differential operator $L = (D - \alpha_1)$ for which $N = 1$ and $M = 0$, the corresponding discrete operator, Δ_L, becomes $\Delta_L\{\mathbf{x}\}[n] = \mathbf{x}[n] - e^{\alpha_1}\mathbf{x}[n-1]$; see [47].

For the differential operator $L = \prod_{i=1}^{N}(D - \alpha_i I)$ of order $N > 0$ and $M = 0$, we can obtain the filter Δ_L by successive convolutions (leading to support of $N + 1$); the z-transform of $\hat{\Delta}_L(z)$ is then

$$\hat{\Delta}_L(z) := \sum_{n=0}^{N} \Delta_L[n]z^{-n} = \prod_{i=1}^{N}(1 - e^{\alpha_i}z^{-1}) = \prod_{i=1}^{N} p_i(z^{-1}) \quad (12)$$

where p_i is a polynomial with 2 coefficients $p_{i,k} = (-e^{\alpha_i})^k$ with $k \in [0,1]$. Note that we can express (12) benefiting the polynomial multiplication which leads to the convolution as

$$\prod_{i=1}^{N} p_i(z^{-1}) = \sum_{n=0}^{N} \mathbf{P}_N[n]z^{-n} \quad (13)$$

where $\mathbf{P}_N = p_1 * p_2 * \ldots * p_N$ and $*$ is the convolution operator. Therefore, we can express the filter $\Delta_L[n] = \mathbf{P}_N[n]$ as

$$\mathbf{P}_N[n] = \sum_{k_{N-1}=0}^{n} \cdots \sum_{k_2=0}^{k_3} \sum_{k_1=0}^{k_2} (-e^{\alpha_1})^{k_1}(-e^{\alpha_2})^{k_2-k_1}$$
$$\cdots (e^{\alpha_N})^{n-k_{N-1}}, \quad \{k_1,\ldots,n-k_{N-1}\} \in [0,1]^N$$

with a change of variables $m_1 = k_1, m_i = k_i - k_{i-1}, m_N = n - k_{N-1}$ we have

$$\Delta_L[n] = (-1)^n \sum_{\mathbf{m}=n} (e^{\boldsymbol{\alpha}})^{\mathbf{m}}, \quad \mathbf{m} \in [0,1]^N, 0 \le n \le N. \quad (14)$$

Similarly, for the general differential operator L, with $M > 0$ we have in z-domain

$$\hat{\Delta}_L(z) = \hat{\Delta}_{L_n}(z)\hat{\Delta}_{L_d}^{-1}(z) = \frac{\prod_{i=1}^{N}(1 - e^{\alpha_i}z^{-1})}{\prod_{i=1}^{M}(1 - e^{\gamma_i}z^{-1})} \quad (15)$$

where $\Delta_{L_n}[n]$ is represented explicitly in (14). Note that the filter $\hat{\Delta}_L$ in (15) has infinite support in time, therefore stability should be assured. Depending on the poles γ_i of the operator, we

find a combination of causal $(\boldsymbol{\gamma'})$ and anti-causal $(\boldsymbol{\gamma''})$ filters that guarantee stability; e.g., for $M = 1$ and $N = 0$, we have either $\Delta_{\Gamma}[n] = e^{\gamma' n} u[n]$ or $\Delta_{\Gamma}[n] = -e^{\gamma'' n} u[-n-1]$, where $u[n]$ is the unit step function. In practice, the filter with input x and output y can be reformulated and implemented as in a recursive way by providing the realization of causal and anti-causal parts separately. To this aim, we represent the inverse filter

$$\widehat{\Delta}_{L_d}^{-1}(z) = \widehat{\Delta}_{L_d'}^{-1}(z)\widehat{\Delta}_{L_d''}^{-1}(z)\mathbf{e}^{-\gamma''}(-z)^{m_2} \qquad (16)$$

by the causal and anti-causal filters

$$\widehat{\Delta}_{L_d'}(z) = \prod_{i=1}^{m_1}\left(1 - e^{\gamma_i'}z^{-1}\right)$$
$$\widehat{\Delta}_{L_d''}(z) = \prod_{i=1}^{m_2}\left(1 - e^{-\gamma_i''}z\right) \qquad (17)$$

where $\mathbf{e} = (e, \ldots, e)$ is a vector of length m. Then the corresponding recursive algorithm can be easily derived from the z-domain representation. Here we will concentrate on the anti-causal part (the derivation for the causal part is similar). To obtain $\mathbf{y} = \Delta_{L_d''}^{-1}\{\mathbf{x}[n + m_2](-1)^{m_2}\mathbf{e}^{-\gamma''}\}$, we consider

$$\widehat{\mathbf{y}}(z)\widehat{\Delta}_{L_d''}(z) = \mathbf{e}^{-\gamma''}(-z)^{m_2}\widehat{\mathbf{x}}(z)$$

from which we find

$$\sum_k \mathbf{y}[n-k]\Delta_{L_d''}[k] = \mathbf{x}[n+m_2](-1)^{m_2}\mathbf{e}^{-\gamma''}.$$

From (14), we can derive the explicit time domain expression for the anti-causal filter $\Delta_{L_d''}$ as

$$\Delta_{L_d''}[n] = (-1)^n \sum_{|\mathbf{m}|=-n}\left(e^{-\gamma''}\right)^{\mathbf{m}}, \mathbf{m} \in [0,1]^{m_2}, -m_2 < n \le 0.$$
$$\qquad (18)$$

Therefore, we obtain

$$\mathbf{y}[n] = \mathbf{x}[n+m_2]\Delta_{L_d''}[-m_2] - \sum_{k=-m_2}^{1}\mathbf{y}[n-k]\Delta_{L_d''}[k]$$

where we used $\Delta_{L_d''}[0] = 1$. $\blacksquare$

Let us give an example for the third-order differential operator $L - \prod_{i=1}^{3}(D - \alpha_i I)$. The FIR filter $\Delta_{\Gamma}[n]$ then becomes

$$\Delta_{\Gamma}[n] = \Big[1, -(e^{\alpha_1} + e^{\alpha_2} + e^{\alpha_3}), e^{(\alpha_1+\alpha_2)} + e^{(\alpha_1+\alpha_3)}$$
$$+ e^{(\alpha_2+\alpha_3)}, -e^{(\alpha_1+\alpha_2+\alpha_3)}\Big], \quad 0 \le n \le 3.$$

REFERENCES

[1] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*. New York: Springer, 1996.

[2] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," in *Proc. 11th Annu. Int. Conf. Center for Nonlinear Studies on Experimental Math.: Comput. Issues in Nonlinear Sci.*, Amsterdam, The Netherlands, 1992, pp. 259–268, Elsevier North-Holland, Inc..

[4] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 231–241, Feb. 2001.

[5] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[6] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, Jun. 2007.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[8] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 101–126, 1992.

[9] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Process.*, vol. 7, pp. 319–335, 1996.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Series B*, vol. 58, pp. 267–288, 1994.

[11] C. Vonesch and M. Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 509–523, Mar. 2009.

[12] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Handbook of Mathematical Models in Computer Vision*. New York: Springer, 2005, pp. 17–30.

[13] W. Stefan, R. Renaut, and A. Gelb, "Improved total variation-type regularization using higher order edge detectors," *SIAM J. Imag. Sci.*, pp. 232–251, Jan. 2010.

[14] A. Chambolle and P. L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, pp. 167–188, 1997.

[15] J. Yuan, C. Schnörr, and G. Steidl, "Total-variation based piecewise affine regularization," in *Proc. 2nd Int. Conf. Scale Space Variation. Methods Comput. Vis. (SSVM)*, 2009, pp. 552–564.

[16] T. F. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 22, no. 2, pp. 503–516, 2000.

[17] T. Chan, S. Esedoglu, and S. Park, "Image decomposition combining staircase reduction and texture extraction," *J. Vis. Commun. Image Rep.*, pp. 468–486, Jan. 2007.

[18] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magn. Resonance Med.*, vol. 65, no. 2, pp. 480–491, 2011.

[19] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, Jan. 2004.

[20] B. Wohlberg and P. Rodriguez, "An iteratively reweighted norm algorithm for minimization of total variation functionals," *IEEE Signal Proc. Lett.*, vol. 14, no. 12, pp. 948–951, Dec. 2007.

[21] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, 1996.

[22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[23] J. Darbon and M. Sigelle, "Image restoration with discrete constrained total variation part i: Fast and exact optimization," *J. Math. Imag. Vis.*, vol. 26, no. 3, pp. 261–276, Dec. 2006.

[24] M. Elad, J. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, Nov. 2005.

[25] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," *SIAM J. Numer. Anal.*, vol. 42, pp. 686–713, 2004.

[26] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," in *Proc. SPIE, Wavelets XIII*, Aug. 2–4, 2009, vol. 7446.

[27] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1256–1267, Apr. 2006.

[28] C. Vonesch, T. Blu, and M. Unser, "Generalized Daubechies wavelet families," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4415–4429, September 2007.

[29] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1425–1438, Apr. 2005.

[30] M. Unser, "Cardinal exponential splines: Part II—Think analog, act digital," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1439–1449, Apr. 2005.

[31] A. Chambolle, "Total variation minimization and a class of binary MRF models," *Lecture Notes Comput. Sci.*, vol. 3757, pp. 136–152, 2005.

[32] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Stat.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[33] K. Lange, *Optimization*. New York: Springer, 2004.

[34] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[35] M. Zibulevsky and M. Elad, "L1-l2 optimization in signal and image processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 76–88, 2010.

[36] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: A majorization-minimization approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2006, vol. 2, p. II.

[37] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[38] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.

[39] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition-modeling, algorithms, and parameter selection," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 111–136, Apr. 2006.

[40] V. Solo, "Selection of regularisation parameters for total variation denoising," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 1999, vol. 3, pp. 1653–1655.

[41] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 326–339, 2008.

[42] Y. Lin, B. Wohlberg, and H. Guo, "UPRE method for total variation parameter selection," *Signal Process.*, vol. 90, pp. 2546–2551, Aug. 2010.

[43] F. Jaillet and B. Torrésani, "Time-frequency jigsaw puzzle: Adaptive multiwindow and multilayered Gabor expansions," *Proc. Int. J. Wavelets Multiresolution Inf. Process.*, vol. 5, no. 2, pp. 293–315, 2007, 05.

[44] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *SIAM Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, 2007.

[45] M. Unser and P. Tafti, "Stochastic models for sparse and piece- wise-smooth signals," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 989–1006, Mar. 2011.

[46] T. Bengtsson and J. Cavanaugh, "An improved Akaike information criterion for state-space model selection," *Comput. Stat. Data Anal.*, vol. 50, no. 10, pp. 2635–2654, June 2006.

[47] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2146–2159, Jun. 2005.