Recovery of Discontinuous Signals Using Group Sparse Higher Degree Total Variation

Greg Ongie*, *Student Member, IEEE*, Mathews Jacob, *Senior Member, IEEE*

***Abstract*—We introduce a family of novel regularization penalties to enable the recovery of discrete discontinuous piecewise polynomial signals from undersampled or degraded linear measurements. The penalties promote the group sparsity of the signal analyzed under a nth order derivative. We introduce an efficient alternating minimization algorithm to solve linear inverse problems regularized with the proposed penalties. Our experiments show that promoting group sparsity of derivatives enhances the compressed sensing recovery of discontinuous piecewise linear signals compared with an unstructured sparse prior. We also propose an extension to 2-D, which can be viewed as a group sparse version of higher degree total variation, and illustrate its effectiveness in denoising experiments.**

***Index Terms*—Higher degree total variation (HDTV), group sparsity, analysis models, compressive sensing, denoising**

I. Introduction

Total variation (TV) has had great success as a regularization penalty for a diversity of inverse problems, especially in its ability to preserve sharp edges in recovered signals. However, a piecewise constant prior is not appropriate for all types of signals. For example, many natural images contain smoothly varying regions due to lighting gradients, and are better modeled as piecewise linear or piecewise polynomial. To address this issue, researchers have proposed various extensions of TV that promote sparsity of higher order derivatives [1]–[5], which are closely related to spline representations in 1-D [6].

However, these higher order schemes can bias the results towards continuous (with derivatives of order $n = 2$) or continuously differentiable signals ($n > 2$), since these signals have sparser representations under higher order derivatives. In contrast, to represent a discontinuous 1-D signal as a spline function under an nth order derivative requires n adjacent coefficients per discontinuity, assuming the discontinuities are sufficiently separated. Yet the preservation of these discontinuities is a key concern in many signal recovery settings, such as in image reconstruction where discontinuities represent sharp image edges. Hence, we propose extending these higher order schemes to accommodate discontinuous signals by exploiting the group structure of higher order derivatives of the signal.

Specifically, we derive a model for discrete 1-D discontinuous piecewise polynomial signals with maximum degree $(n-1)$ based on the group sparsity of the nth order derivative. This can be viewed as an instance of a *group sparse analysis model*, a variant of the recently proposed cosparse analysis signal model [7]. We introduce a family of non-convex regularization penalties to promote group sparsity of derivatives. We also propose an efficient algorithm to solve linear inverse problems regularized with these penalties based on the smooth continuation method in [8]. Our experiments on the compressed sensing (CS) recovery of 1-D synthetic piecewise linear signals show that our proposed scheme outperforms the corresponding cosparse scheme.

We also investigate an extension of the penalties to 2-D, which can be viewed as group sparse versions of our recently introduced higher degree total variation (HDTV) penalties [5] and their generalizations [9]. Note that neither [5] nor [9] address the case of group sparse derivatives or non-convex penalties. Our experiments show the group sparse HDTV penalty improves over regular HDTV and other related higher order TV penalties in denoising natural images.

A. Related work

Many other researchers have proposed non-convex penalties for promoting group sparsity, showing substantial improvements over convex formulations, which motivates their use in this work. Notably [10] investigates an overlapping group sparsity prior for denoising, and proposes using certain non-convex penalty functions such that the overall cost function is convex. While this approach has advantages over the fully non-convex formulation that we pursue, the theory in [10] is not easily extended the general linear inverse problem setting we consider, nor to a general analysis prior. See also [11] for a non-convex group sparse analysis approach to compressive color imaging.

This work also has similarities to [12] which introduces a group sparse version of 1-D TV to allow for the recovery of smooth edges in signals that are approximately piecewise constant. However, the signal model and goals of this work are fundamentally different, as we are concerned with *preserving* sharp edges in signals that are piecewise polynomial, not piecewise constant.

II. Signal Model

A. 1-D signal model

We begin by considering the class of discrete 1-D signals consisting of a small number of polynomial segments with discontinuities, or "jumps", allowed at the interfaces. In particular, we define $\mathcal{P}^n(k,d)$ to be space of all signals $\mathbf{x} \in \mathbb{R}^d$

G. Ongie is with the Department of Mathematics, University of Iowa, and M. Jacob is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, 52245 USA (e-mail: gregory-ongie@uiowa.edu; mathews-jacob@uiowa.edu)

This work is supported by grants NSF CCF-0844812, NSF CCF-1116067, NIH 1R21HL109710-01A1, ACS RSG-11-267-01-CCE, and ONR-N000141310202.

consisting of at most k polynomial segments, each having maximum degree $(n-1)$; see Fig. 1. Note that signals whose piecewise polynomial segments have different degrees can be modeled as elements of $\mathcal{P}^n(k,d)$, where $(n-1)$ is equal to the highest degree of a segment.

More formally, we may define $\mathcal{P}^n(k,d)$ as all $\mathbf{x} \in \mathbb{R}^d$ such that for some constants $a_{i,j} \in \mathbb{R}$ and knots $1 \leq m_1 < ... < m_k \leq d$,

$$\mathbf{x} = \sum_{i=1}^{k} \left(a_{i,0}\mathbf{r}_{m_i}^0 + a_{i,1}\mathbf{r}_{m_i}^1 + \cdots + a_{i,n-1}\mathbf{r}_{m_i}^{n-1} \right), \quad (1)$$

where $\mathbf{r}_m^\ell := (1/\ell!)\max(0, j-m+1)^\ell$ are powers of a linear ramp. Signals in the form of (1) have a particularly simple representation under the the discrete nth derivative operator $\mathbf{D} : \mathbb{R}^d \to \mathbb{R}^d$, defined as n applications of the backwards finite difference operator $\mathbf{F}$, defined as $[\mathbf{Fz}]_j = z_j - z_{j-1}$, for $1 \leq j \leq d$ with $z_0 = 0$. Note that $\mathbf{D}\mathbf{r}_m^\ell = \boldsymbol{\delta}_m^{(n-\ell-1)}$ where $\boldsymbol{\delta}_m^{(p)}$ is a pth order derivative of a Dirac delta centered at m, which has support $\{m, ..., m+p\}$. By linearity, for any $\mathbf{x} \in \mathcal{P}^n(k,d)$ this gives

$$\mathbf{Dx} = \sum_{i=1}^{k} \left(a_{i,0}\boldsymbol{\delta}_m^{(n-1)} + a_{i,1}\boldsymbol{\delta}_m^{(n-2)} + \cdots + a_{i,n-1}\boldsymbol{\delta}_m \right). \quad (2)$$

When the number of jumps k is much less than than the signal length d, the analyzed signal in (2) is sparse, having at most $n \cdot k$ nonzero entries, with support contained within the union of all groups of indices $\{m_i, ..., m_i+n-1\}$, $i = 1, ..., k$. In other words, $\mathbf{Dx}$ is *group sparse* under the nth order derivative, where the groups are defined as all overlapping sets of n consecutive indices within $\{1, ..., d\}$.

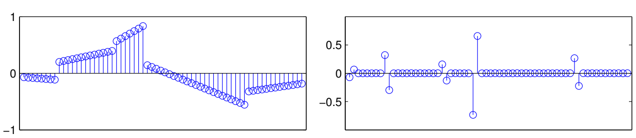


Fig. 1: Example of a signal belonging to $\mathcal{L}^2(5, 64)$ (left) and the signal analyzed under the second derivative operator (right). Note the second derivative is pairwise group sparse.

B. Relation to cosparse analysis models

The above signal model is closely related to the recently proposed *cosparse analysis* model [7]. A cosparse model encompasses signals that are sparse after the application of an *analysis operator* $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$, $p \geq d$, *i.e.* the signal $\mathbf{y} = \mathbf{\Omega x} \in \mathbb{R}^p$ is sparse. The proposed model can be viewed as extension of a cosparse model to signals that are group sparse under an analysis operator, thus we call this a *group sparse analysis* model.

As observed in [7], cosparse analysis models face the problem of having a combinatorial number of subspaces of low dimension in their union-of-subspaces formulation, which has bearings on the robustness of CS-recovery of signals belonging to these models [13]. Enforcing group sparsity of the analyzed signal can significantly reduce the number of these subspaces, since they are determined by the allowed locations of the non-zero entries of the analyzed signal, which are more heavily restricted in the group sparse case.

C. Group sparse regularization penalties

A standard approach to promote group sparsity is to penalize the mixed ℓ^2-ℓ^1 norm of the signal restricted to the groups $\mathcal{G}$:

$$\Phi_1(\mathbf{y}) = \sum_{g \in \mathcal{G}} \|\mathbf{y}_g\|_2, \quad (3)$$

where $\mathbf{y}_g \in \mathbb{R}^d$ denotes the vector whose entries are the same as $\mathbf{y} \in \mathbb{R}^d$ on the index set $g \in \mathcal{G}$, and zeros elsewhere. This penalty is also known as the *group lasso*, and its ability to recover group sparse signals been studied extensively [14], including the case of overlapping groups [15]–[17]. To better approximate the ideal ℓ^0 measure of sparsity, we also consider the following non-convex variant:

$$\Phi_p(\mathbf{y}) = \frac{1}{p} \sum_{g \in \mathcal{G}} \|\mathbf{y}_g\|_2^p, \quad 0 < p \leq 1. \quad (4)$$

In our setting, we may give a more compact representation of the penalties in (3) and (4) by introducing an *expansion operator* $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^{(d-n+1)\times n}$ defined by

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_d \end{bmatrix}^T \mapsto \frac{1}{\sqrt{n}} \begin{bmatrix} y_1 & y_2 & \cdots & y_{d-n+1} \\ \vdots & \vdots & & \vdots \\ y_n & y_{n+1} & \cdots & y_d \end{bmatrix}^T,$$

where the rows of $\mathcal{E}\mathbf{y}$ collect the coefficients of $\mathbf{y}$ belonging to each group of n consecutive indices. Up to a scaling, we may recast (3) and (4) as $\Phi_p(\mathbf{y}) = \|\mathcal{E}\mathbf{y}\|_{2,p}^p$ for $0 < p \leq 1$, where $\|\mathbf{U}\|_{2,p}^p := \frac{1}{p}\sum_i \|\mathbf{u}_i\|_2^p$, and $\mathbf{u}_i$ is a row of $\mathbf{U}$.

Therefore, to promote the group sparsity of derivatives $\mathbf{y} = \mathbf{Dx}$, we propose the following family of regularization penalties:

$$\Psi_p(\mathbf{x}) := \Phi_p(\mathbf{Dx}) = \|\mathcal{E}\mathbf{Dx}\|_{2,p}^p, \quad 0 < p \leq 1. \quad (5)$$

Note that by setting $\mathcal{E} = I$, the identity operator, (5) reduces to a cosparse penalty $\|\mathbf{Dx}\|_p^p$. In this way, our algorithm in Sec. III can be adapted to the cosparse case.

D. Extension to 2-D: Group sparse HDTV

Let $\mathbf{x} \in \mathbb{R}^d$ now represent a vectorized 2-D discrete image. One simple extension to 2-D is to apply the proposed 1-D penalty to the rows and columns of the image in a separable fashion:

$$\Psi_p(\mathbf{x}) = \|\mathcal{E}_1\mathbf{D}_1\mathbf{x}\|_{2,p}^p + \|\mathcal{E}_2\mathbf{D}_2\mathbf{x}\|_{2,p}^p, \quad 0 < p \leq 1. \quad (6)$$

Here $\mathbf{D}_1$ and $\mathbf{D}_2$ are the horizontal and vertical finite difference operators, respectively, and the expansion operators $\mathcal{E}_1$ and $\mathcal{E}_2$ collect all overlapping sets of n indices in the horizontal and vertical direction, respectively. This may be viewed as a two-angle group sparse version of the anisotropic HDTV penalty introduced in [5], hence we label the penalty in (6) as group sparse HDTV (GS-HDTV). Note we may also express (6) as $\|\mathcal{E}\mathbf{Dx}\|_{2,p}^p$, for $\mathcal{E} = \mathrm{diag}(\mathcal{E}_1, \mathcal{E}_2)$ and $\mathbf{D} = [\mathbf{D}_1^T, \mathbf{D}_2^T]^T$, which allows us to directly apply our algorithm in Sec. III to this case as well.

III. ALGORITHM

Here our goal is to recover a signal $\mathbf{x} \in \mathbb{R}^d$ from its possibly degraded or undersampled linear measurements $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{n} \in \mathbb{R}^m$, where $\mathbf{A} \in \mathbb{R}^{m \times d}$ with $m \leq d$ models the signal acquisition, and $\mathbf{n}$ is a vector of noise. For simplicity we assume the noise is additive white Gaussian (AWGN), $\mathcal{N}(0, \sigma^2)$. We pose the recovery as an optimization problem:

$$\min_{\mathbf{x}} \|\mathcal{E}\mathbf{D}\mathbf{x}\|_{2,p}^p + \frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \tag{7}$$

where $\lambda > 0$ is a regularization parameter whose optimal value depends on the noise level σ.

The cost function in (7) is convex when $p = 1$, and may be efficiently minimized using a variable splitting method [18] or augmented Lagrangian/split Bregman iterations [19]. For the non-convex case $0 < p < 1$ we adopt an approach proposed in [8] that generalizes the splitting method in [18] to non-convex ℓ^p penalties. The adaptation of this algorithm to our setting (7) is straightforward and is summarized in Alg. 1. The key ingredient is a generalization of the vectorized ℓ^1 soft-shrinkage rule to ℓ^p penalties for $0 < p \leq 1$:

$$\text{shrink}_p(\mathbf{t}, \gamma) := \max\{0, 1 - \gamma\|\mathbf{t}\|_2^{p-2}\} \cdot \mathbf{t}, \quad \forall \mathbf{t} \in \mathbb{R}^n, \gamma > 0.$$

The algorithm consists of a inner loop, which solves a smooth approximation to (7), and an outer loop, which refines the approximation by incrementing a parameter β. The inner loop of the algorithm alternates between two efficiently solved subproblems: 1) A shrinkage operation on the rows of $\mathcal{E}\mathbf{D}\mathbf{x}$ and 2) a matrix inversion step that can be solved analytically in the Fourier domain for operators $\mathbf{A}$ diagonalizable under the DFT, or more generally by a fast iterative method. We note the complexity of the algorithm is essentially the same as TV minimization [18], aside from the extra overhead in computing the expansion operator $\mathcal{E}$ and its transpose $\mathcal{E}^T$.

Algorithm 1 Non-convex group sparse analysis recovery

Choose parameters $\lambda > 0, r > 1, \beta_0 > 0$.
Initialize $\mathbf{x} \leftarrow \mathbf{x}_0 \in \mathbb{R}^d$, $\mathbf{Z} \leftarrow \mathbf{0} \in \mathbb{R}^{d \times n}$, $\beta \leftarrow \beta_0$.
for $j = 1$ to M_{outer} **do** ▷ β-continuation
  $\alpha \leftarrow \lambda/\beta$
  for $i = 1$ to M_{inner} **do** ▷ Alternating minimization
    $\mathbf{Z} \leftarrow \text{shrink}_p(\mathcal{E}\mathbf{D}\mathbf{x}, 1/\beta)$
    $\mathbf{x} \leftarrow (\mathbf{D}^T\mathbf{D} + \alpha\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{D}^T\mathcal{E}^T\mathbf{Z} + \alpha\mathbf{A}^T\mathbf{b})$
  end for
  $\beta \leftarrow r \cdot \beta$
end for
return $\mathbf{x}$

We find parameter choices $1.5 \leq r \leq 1.75$ and $10 \leq \beta_0 \leq 25$, $M_{\text{outer}} = 30$, $M_{\text{inner}} = 15$, work well for the experiments considered in this work. While a smaller value of p can further enhance sparsity, we observe taking p too small ($p < 0.5$) requires many more inner iterations M_{inner} for stable convergence. We choose $p = 0.5$ as a compromise between sparsity enhancement and computation time.

IV. EXPERIMENTS

A. 1-D CS-recovery

We compare our proposed group sparse analysis penalties with the corresponding cosparse penalties (i.e. (5) with $\mathcal{E} = \mathbf{I}$), in both the convex ($p = 1$) and non-convex ($p = 0.5$) case, for the CS-recovery of discontinuous synthetic piecewise linear signals. First we determine the probability of exact recovery of a signal with k jumps from from noiseless random samples. In each trial we generate a random signal $\mathbf{x} \in \mathcal{P}^2(k, 256)$ by selecting the indices of the k jump discontinuities uniformly at random, with jump magnitudes distributed as $\mathcal{N}(0, 1)$, and the slopes of each line segment distributed as $\mathcal{N}(0, 0.01)$; see Fig. 3(a) for a representative signal generated by this procedure. We also randomly generate measurement matrices $\mathbf{A} \in \mathbb{R}^{150 \times 256}$ with i.i.d. $\mathcal{N}(0, 1)$ entries, each column rescaled to have unit norm. The convex recovery schemes were solved with the MATLAB `cvx` package to ensure optimally accurate solutions [20]. The non-convex schemes were solved with Alg. 1, combined with Augmented Lagrangian iterations [21] to enforce the constraint $\mathbf{A}\mathbf{x} = \mathbf{b}$. Here we used the same parameters and initialization $\mathbf{x}_0 = \mathbf{A}^\dagger\mathbf{b}$, where $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$. The recovered signal $\mathbf{x}^*$ is judged to be recovered "exactly" if $\|\mathbf{x}^* - \mathbf{x}\|_\infty / \|\mathbf{x}\|_\infty < 10^{-6}$. The results of the experiment are shown in Fig. 2.

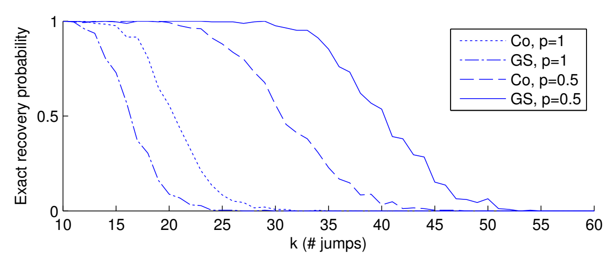


Fig. 2: Exact CS-recovery probabilities for 1-D piecewise linear signals of length $d = 256$ from $m = 150$ as a function of jumps k from noiseless random measurements ($N = 250$ trials), using group sparse (GS) and cosparse (Co) penalties.

We observe that the convex group sparse penalty performs marginally worse than the convex cosparse penalty. However, this trend is reversed with the non-convex penalties, such that the range of exact recovery is extended nearly two-fold in the group sparse case. This result demonstrates the effectiveness of a non-convex penalty for promoting analysis group sparsity, and is consistent with other comparisons of non-convex and convex penalties in this context (see e.g. [10]).

In Fig. 3 we illustrate the case of noisy measurements, as specified by (7). The quality of the recovery $\mathbf{x}^*$ is measured by the signal-to-noise ratio, $\text{SNR} = 20\log_{10}(\|\mathbf{x}\|_2 / \|\mathbf{x}^* - \mathbf{x}\|_2)$, where $\mathbf{x}$ is the original signal. Note that Alg. 1 depends on a regularization parameter λ which needs to be tuned for optimal results. In each case we choose the λ that maximizes the SNR of the recovery. Table I reports the SNR's obtained by each penalty over a range of undersampling factors (m) and noise levels (σ), averaged over $N = 50$ trials. Random signals and measurement matrices are generated as in the noiseless setting. Again, we find the non-convex group sparse penalty consistently outperforms the non-convex cosparse penalty in

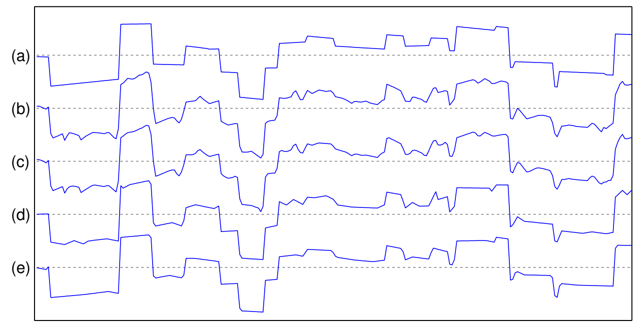


Fig. 3: Example of noisy CS-recovery. (a) original piecewise linear signal of length $d = 256$ with $k = 25$ jumps. Signals recovered from $m = 150$ measurements corrupted by AWGN, $\sigma = 0.05$: (b) convex cosparse (SNR=17.5dB); (c) convex group sparse (SNR=17.4 dB); (d) non-convex cosparse (SNR=22.0dB); (e) non-convex group sparse (SNR=26.4dB).

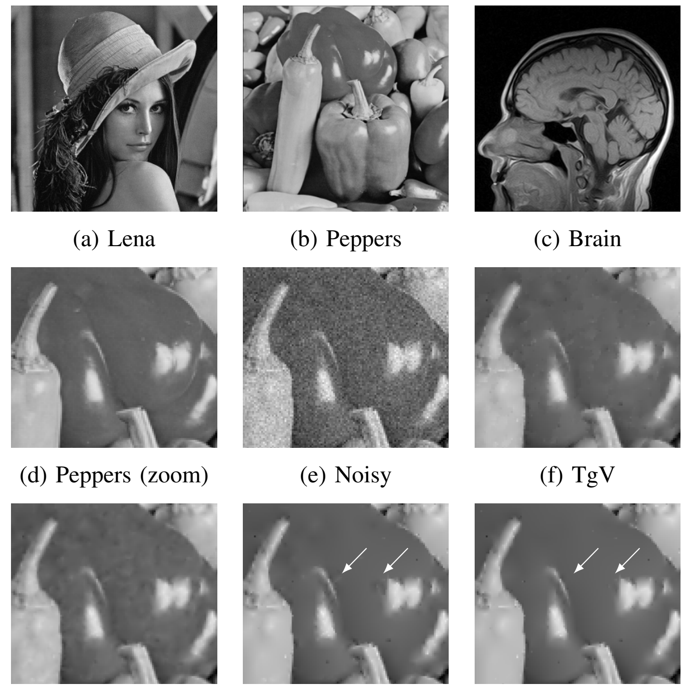
	σ = 0.05			σ = 0.10		
Penalty	m=125	m=150	m=175	m=125	m=150	m=175
Co, p=1	15.2 (3.2)	15.3 (3.4)	15.6 (3.3)	15.5 (3.3)	15.5 (3.4)	15.2 (3.2)
GS, p=1	15.7 (2.7)	15.7 (2.8)	15.9 (2.8)	15.9 (2.8)	15.9 (2.9)	15.6 (2.8)
Co, p=0.5	18.2 (5.1)	18.3 (5.3)	18.8 (5.3)	18.5 (5.2)	18.4 (5.4)	18.2 (5.3)
GS, p=0.5	21.4 (4.4)	21.1 (4.5)	21.1 (4.6)	20.9 (4.8)	20.8 (4.9)	20.8 (4.7)

TABLE I: Comparison of group sparse (GS) and cosparse (Co) penalties for CS-recovery of 1-D piecewise linear signals ($k = 25, d = 256$) from m random measurements corrupted with AWGN, $\mathcal{N}(0, \sigma^2)$. Reported values are the average SNR's (in dB) from $N = 50$ realizations, with the standard deviation (in dB) in parenthesis.

terms of average SNR, and both significantly outperform the convex penalties. Here the convex group sparse penalty also shows a slight SNR improvement over the convex cosparse penalty. However, the effect is small relative to the variance, and insignificant compared to the improvement offered by non-convex penalties.

B. 2-D denoising

We demonstrate the utility of the non-convex (p=0.5) second degree GS-HDTV penalty (6) for denoising natural images. The three 256×256 test images we consider are shown in Fig. 4(a)-(c), which we corrupt with various levels of AWGN. Here we compare against standard isotropic total variation (TV), and two second degree TV extensions: total generalized variation (TgV) [22] and anisotropic second degree HDTV [5]. Also, we compare with the non-convex (p=0.5) second degree cosparse penalty (i.e. (6) with $\mathcal{E}_1 = \mathcal{E}_2 = \mathbf{I}$), which we label Co-HDTV. The GS-HDTV and Co-HDTV schemes are solved with Alg. 1; average computation time of our unoptimized MATLAB implementation running on a desktop computer (Intel i5-3470 CPU, 3.20 GHz, 8 GB RAM) is ∼25s. The SNR's of the denoised images are reported in Table II. Regularization parameters were tuned to optimize the SNR in each case. Note the non-convex schemes GS-HDTV and Co-HDTV typically outperform the convex ones, demonstrating the benefit of non-convex penalty functions. However, GS-HDTV also shows a consistent ∼0.3dB gain in SNR over Co-HDTV, indicating the benefit of GS-HDTV is not due to non-convexity alone. Indeed, we see the most improvement



(a) Lena (b) Peppers (c) Brain

(d) Peppers (zoom) (e) Noisy (f) TgV

(g) HDTV (h) Co-HDTV (i) GS-HDTV

Fig. 4: Denoising comparison. (a)-(c) Original images. (d) Peppers image, zoomed for detail, (e) corrupted with AWGN (SNR=20.0dB). (f) Denoised with TgV regularization (SNR=26.8dB), (g) convex HDTV (SNR=26.7dB), (h) non-convex Co-HDTV (SNR=26.9dB), and (j) non-convex GS-HDTV (SNR=27.2dB). The arrows in (h) and (i) identify noise artifacts successfully removed by GS-HDTV but not Co-HDTV.

	Lena			Peppers			Brain		
Penalty	17dB	20dB	25dB	17dB	20dB	25dB	17dB	20dB	25dB
TV	22.4	24.3	25.5	24.7	26.4	27.7	22.7	24.7	26.0
TgV	22.6	24.5	25.7	24.9	26.8	28.2	23.2	25.1	26.3
HDTV	22.4	24.3	25.5	25.0	26.7	28.1	23.3	**25.3**	**26.6**
Co-HDTV	22.5	24.4	25.7	25.1	26.9	28.3	23.1	25.0	26.2
GS-HDTV	**22.8**	**24.7**	**26.0**	**25.4**	**27.2**	**28.6**	**23.4**	**25.3**	**26.6**

TABLE II: (SNR in dB) 2-D denoising results. Images were corrupted with AWGN to match the indicated SNR (17dB, 20dB, or 25dB). Regularization parameters were optimized to obtain the best SNR in all cases. Original images are displayed in Fig. 4.

in the test images with significant piecewise linear regions, *Lena* and *Peppers* (∼0.4dB SNR gain over HDTV). Contrast this with the mostly piecewise constant *Brain* image (no SNR improvement over HDTV). Visually, the GS-HDTV penalty also appears better at removing localized inhomogeneities in piecewise linear regions; see Fig. 4(d)-(i).

V. Conclusion

Most higher order generalizations of TV implicitly consider a continuous piecewise polynomial model. We extend these schemes to discontinuous signal models by enforcing group sparsity of higher order derivatives. The proposed non-convex group sparse derivative penalties show significant improvement for the CS-recovery of 1-D discontinuous piecewise linear signals compared to a cosparse prior. We also show a similar improvement for the denoising natural images over other second order extensions of TV.

REFERENCES

[1] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2000.

[2] M. Lysaker and X.-C. Tai, "Iterative image restoration combining total variation minimization and a second-order functional," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 5–18, 2006.

[3] J. Yuan, C. Schnörr, and G. Steidl, "Total-variation based piecewise affine regularization," in *Scale Space and Variational Methods in Computer Vision*. Springer, 2009, pp. 552–564.

[4] S. Lefkimmiatis, J. Paul Ward, and M. Unser, "Hessian schatten-norm regularization for linear inverse problems," *IEEE transactions on image processing*, vol. 22, no. 5-6, pp. 1873–1888, 2013.

[5] Y. Hu and M. Jacob, "Higher degree total variation (HDTV) regularization for image recovery," *Image Processing, IEEE Transactions on*, vol. 21, no. 5, pp. 2559–2571, 2012.

[6] G. Steidl, S. Didas, and J. Neumann, "Splines in higher order TV regularization," *International journal of computer vision*, vol. 70, no. 3, pp. 241–255, 2006.

[7] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[8] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. IEEE, 2009, pp. 262–265.

[9] Y. Hu, G. Ongie, S. Ramani, and M. Jacob, "Generalized higher degree total variation (hdtv) regularization," *Image Processing, IEEE Transactions on*, vol. 23, no. 6, pp. 2423–2435, June 2014.

[10] P.-Y. Chen and I. W. Selesnick, "Group-sparse signal denoising: Nonconvex regularization, convex optimization," *Signal Processing, IEEE Transactions on*, vol. 62, no. 13, pp. 3464–3478, 2014.

[11] A. Majumdar and R. K. Ward, "Compressed sensing of color images," *Signal Processing*, vol. 90, no. 12, pp. 3122–3127, 2010.

[12] I. W. Selesnick and P.-Y. Chen, "Total variation denoising with overlapping group sparsity," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 5696–5700.

[13] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *Information Theory, IEEE Transactions on*, vol. 55, no. 4, pp. 1872–1882, 2009.

[14] J. Huang, T. Zhang *et al.*, "The benefit of group sparsity," *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010.

[15] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *The Journal of Machine Learning Research*, vol. 12, pp. 2777–2824, 2011.

[16] G. Peyré, J. Fadili *et al.*, "Group sparsity with overlapping partition functions," *Proc. EUSIPCO 2011*, pp. 303–307, 2011.

[17] M. Kowalski, K. Siedenburg, and M. Dorfler, "Social sparsity! neighborhood systems enrich structured shrinkage operators," *Signal Processing, IEEE Transactions on*, vol. 61, no. 10, pp. 2498–2511, 2013.

[18] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[19] J.-F. Cai, S. Osher, and Z. Shen, "Split bregman methods and frame based image restoration," *Multiscale modeling & simulation*, vol. 8, no. 2, pp. 337–369, 2009.

[20] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[21] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 681–695, 2011.

[22] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.