

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Data adaptive multi-scale representations for image analysis

Julia Dobrosotskaya, Weihong Guo

Julia Dobrosotskaya, Weihong Guo, "Data adaptive multi-scale representations for image analysis," Proc. SPIE 11138, Wavelets and Sparsity XVIII, 1113807 (9 September 2019); doi: 10.1117/12.2529695

SPIE.

Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States

Data adaptive multi-scale representations for image analysis

Julia Dobrosotskaya and Weihong Guo

Case Western Reserve University, Cleveland, OH 44106, USA

ABSTRACT

Data adaptive tight frame methods have been proven a powerful sparse approximation tool in a variety of settings. We introduce a model of a data adaptive representation that also provides a multi-scale structure. Our idea is to design a multi-scale frame representation for a given data set, with scaling properties similar to the ones of a wavelet basis, but without the necessary self-similar structure. The adaptivity provides better sparsity properties, using Besov-like norm structure both induces sparsity and helps in identifying important features.

We focus on investigating the efficiency of a weighted l^1 constraint in the context of sparse recovery from noisy data and compare it to the weighted l^0 model alongside. Numerical experiments confirm that the recovered frame vectors assigned lower weights correspond to image elements of larger scale and lower local variation, thus indicating that weighted sparsity in natural images leads to a natural scale separation.

1. INTRODUCTION

Recent decades of imaging research demonstrated that efficient systems of representation need to be either redundant or data-adaptive. Sparse representation in redundant systems has been proven to be a powerful tool of image recovery and analysis. Even though wavelets are widely considered a typical example of a system that provides sparse representations for natural images, in practice, we are often facing the fact that the level of sparsity that wavelets or tight frames with low redundancy typically provide is not sufficient for efficient application of many popular compressive recovery techniques. In this work we explore the possibility of a non-redundant yet data adaptive representation that also incorporates multi-scale structure. Our idea is to adaptively design multi-scale representation for a given data set, trying to obtain a basis/frame with scaling properties similar to the ones of a wavelet basis, but without the necessary self-similar structure. The adaptivity will provide better sparsity properties based on data correlation, using Besov-like norm structure will both induce sparsity and identify important features of the data.

Data adaptive tight frame methods have been proven a powerful sparse approximation tool established in a variety of settings^{1,2,3,4,5}. In this work we focus on investigating the efficiency of weighted l^1 constraint in the context of sparse recovery from corrupted data and compare it to the weighted l^0 model alongside. In other words, we incorporate the weighted sparsity constraint into a model class that recovers a tight frame in which a given dataset is represented optimally in a certain sense with the purpose of data analysis and regularization.

After experimenting with several model problems to ascertain the appropriateness of using the Besov 1-1-1 norm as a measure of imposing a sparsity constraint together with regularity (see Sec. 2), we developed a methodology for recovering a tight frame in which a given sequence of correlated images from can be represented optimally in a weighted sparsity sense (see Sec. 3). The weighted norm that we design cannot be used to measure or enforce classical Sobolev regularity since the adaptively recovered tight frame does not incorporate the properties of regular wavelets any more. However, the multiscale structure imposed by the weights scaled analogously to the Besov norm coefficients assigned to different wavelet scales naturally creates an hierarchical structure in the adaptively recovered frame. Numerical experiments (see Sec. 4) confirm that the recovered frame elements assigned lower weights correspond to image elements of larger scale and lower local variation, thus indicating that weighted sparsity in natural images leads to a natural scale separation.

Data adaptive design of a sparse representation becomes especially relevant in the cases when the imaging data involved has a large number of highly correlated details. In that case joint sparsity becomes a constraint implicitly exploiting the similarities between data channels for better recovery and adaptive representation allows

email: julia.dobrosotskaya@case.edu

to identify a frame that incorporates the most dominant features shared between the data channels. We include the results of numerical tests of denoising, as a particular example of image recovery, in the case of images including repetitive elements and texture, and in the case of a hyperspectral stack, where the similarity of features between bands (channels) comes from the acquisition method (Sec. 4). We summarize our findings and outline the future work in Sec. 5.

2. MOTIVATION FOR USING WEIGHTED L^1 REGULARIZATION

Besov $B_{1,1}^1$ space is a subspace of the BV with $|u|_{TV} \leq C\|u\|_{1,1,1}$ ^{6,7}. The Besov 1-1-1 norm of a function $u \in L^2(\mathbb{R})$ can be expressed via the coefficients of its decomposition in an orthonormal r -regular ($r \geq 1$) wavelet basis generated by the wavelet function ψ and scaling function ϕ as follows:

$$\|u\|_{B_{1,1}^1} = \sum_{k \in \mathbb{Z}} |\langle u, \phi_{0,k} \rangle| + \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} 2^j |\langle u, \psi_{j,k} \rangle|, \quad \text{here } \phi_{0,k}(t) = \phi(t-k), \quad \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k).$$

Besov regularity has been successfully used in sparse recovery algorithms (see⁸). Unlike the l^1 norm of the wavelet basis coefficients, the weighted norm imposes sparsity together with the requirement on the coefficient decay, introducing a mild restriction somewhat resembling joint sparsity with additional requirement on non-zero coefficients to be localized at 'lower frequency' positions. Consider using the Besov $B_{1,1}^1$ norm as a regularizer in simple problems like image denoising via minimizing

$$E_1(u) = \|u\|_{B_{1,1}^1} + \frac{\lambda}{2} \|u - u_0\|_2^2,$$

where u_0 is a given noisy image. However, it does not immediately lead to great results due to the absence of translation invariance in the representation, and thus to prioritizing edges in the data positioned along a diadic grid imposed by the non-redundant wavelet decomposition. Numerical experiments confirm what we could expect - it still outperforms the output of the minimization problem without the weighing:

$$E_2(u) = \|Wu\|_1 + \frac{\lambda}{2} \|u - u_0\|_2^2$$

here W is a predetermined wavelet decomposition matrix. See Fig. 1(b) and (c) respectively.

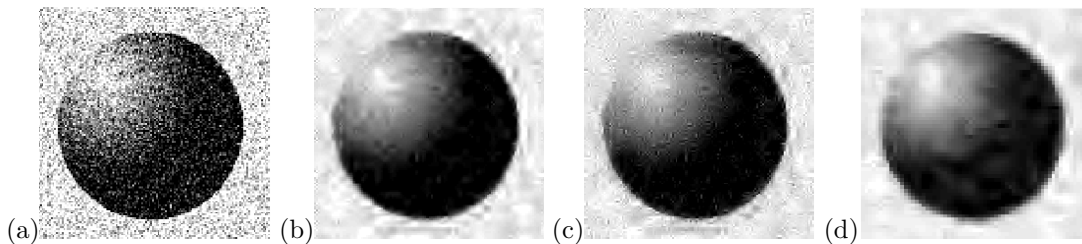


Figure 1. (a) noisy input with $SNR = 4.5dB$, (b) denoising by minimizing E_1 (based on $B_{1,1}^1$, i.e. weighted l^1 norm of non-redundant wavelet transform) with resulting $SNR = 9.75$, (c) denoising by minimizing E_2 (based on l^1 norm of non-redundant wavelet transform) with resulting $SNR = 9.1$, (d) denoising by minimizing E_1 with random shifts

The results were obtained using ADMM implementation,^{9,10} Results can be slightly improved by using a redundant representation (the stationary wavelet transform, which, unfortunately, involves using a 'redundant' rectangular matrix W of much larger size, making such model hard to implement on a computer with RAM of 32Gb or lower) or, as an alternative, introducing random shifts of the image during the ADMM iterations in otherwise identical context - see Fig. 1(d).

Denoising result using $B_{1,1}^1$ model applied to image patches is shown in Fig.3. A 256×256 image was split into overlapping patches of size 16×16 , which formed the image stack. We used matrix W of the 2D wavelet transform associated with the *db2* wavelet and depth of decomposition equal to 3 and D contained the associated Besov weights. The resulting improvement in SNR was from $20dB$ to $23.79dB$, (the parameters had values $\rho = .2$, $\mu = .005$) comparing to the l^1 model with the same basis raising the SNR to $22.97dB$.



Figure 2. (a) Noisy image, $SNR = 20.62dB$, (b) recovery via minimizing E_2 over image patches, $SNR = 22.97$, (c) recovery via minimizing E_1 over image patches, $SNR = 23.79dB$

3. DESCRIPTION OF THE MODEL

Let us assume that Y_0 is a set of q images of size $N \times N$ that have some correlation/common features and that can be expected to have a sparse representation in an orthonormal basis. Let the frame analysis operator be performed by the $N \times N$ matrix W , and the synthesis - by its adjoint W' . Let y_0 denote the matrix of size $N^2 \times q$ containing images reshaped as column vectors. In the process of minimizing the energies described in the next paragraph we would like to recover an updated set of images Y via the corresponding reshaped matrix $y = W'c$, where c is of size $N^2 \times q$. These models require different types of weighted sparsity from the recovered signal: weighted l^1 pseudo-sparsity in the first case and joint weighted l^0 sparsity in the second.

M1. The first model we consider is aimed at recovering the optimal basis W and the sparse representation $Y = W'c$ of the original images Y_0 via solving

$$\min_{W, c: WW' = I} \|W'c - y_0\|_F^2 + \mu \|Dc\|_{1,1}, \quad (P1)$$

where D is a diagonal matrix of nonnegative coefficients weights, typically exponentially increasing from one group of coefficients that we would like to identify as a unified scale to another, as we comment on later. For the convenience of the numerical approximation we introduce an auxiliary variable r and reformulate the problem as

$$\min_{W, c, r} \|W'c - y_0\|_F^2 + \mu \|r\|_1 + \frac{\rho}{2} \|Dc - r\|_F^2. \quad (P2)$$

Since we initialize W as one of the wavelet bases, D can be chosen to be a matrix that makes $\|DW' \cdot\|_1 = \|\cdot\|_{B_{1,1}^1}$.

M2. The second model we test is recovering the optimal basis W and the sparse representation Y of the original images Y_0 via solving

$$\min_{W, c: WW' = I} \|W'c - y_0\|_F^2 + \mu \|Dc\|_{2,0}, \quad \|Dc\|_{2,0} = \|D\chi(c)\|_2, \quad (P3)$$

where $\chi(c)$ is a matrix with rows equal to the rows of the identity whenever the respective rows of c are non-zero, and zero vectors otherwise. This imposes the restriction of an 'hierarchical joint sparsity', as the non-zero coefficient rows are weighted in a predefined scaled manner; $W'c = y$ after reshaping delivers Y - the stack of recovered images.

As before, we introduce an auxiliary r and reformulate the problem as

$$\min_{W, c, r} \|W'c - y_0\|_F^2 + \mu \|r\|_{2,0} + \frac{\rho}{2} \|Dc - r\|_F^2. \quad (P4)$$

Let us notice that in the context of these models a meaningful recovery of the basis(transform) W is only possible when q is comparable to N^2 or $q > N^2$. For instance, recovery of such a basis for the an entire RGB

image (without splitting it into patches) is not plausible. Let us also remark that in the case of the image stack Y_0 being comprised of the image patches the recovered basis for the patches padded by zeros around each respective patch forms a tight frame for the space of matrices of size of the original image. This was proven in,¹ where the authors consider a similar model with unweighted joint sparsity constraint.

4. NUMERICAL IMPLEMENTATION AND RESULTS

4.1 Algorithms

The following algorithm implements the minimization posed in (P2) via ADMM with a proximal regularization term for W . $W_0 = W$ and D are initialized as described above. y_0 is the original stack of images reshaped as vector columns and written as a matrix $N^2 \times q$, c_0 is also a matrix $N^2 \times q$, columns of which are the respective coefficients of decomposition of the columns of y_0 via W_0 and $r_0 = Dc_0$.

$$\begin{aligned} c_{k+1} &= (I_{N^2} + \rho D^2)^{-1} (W_k y_0 + \rho D' (r_k - 1/\rho \lambda_k)) \\ r_{k+1} &= \text{shrink}(Dc_k + \rho^{-1} \lambda_k, \mu/\rho), \quad \text{shrink}(a, b) = \text{sign}(a) \max(\text{abs}(a) - b, 0) \\ \lambda_{k+1} &= \lambda_k + \rho (Dc_k - r_k) \\ y_k &= W' c_k \\ [U, S, V] &= \text{SVD}(c_k y_0' + \beta W_k), \quad W_{k+1} = U(V') \end{aligned}$$

Alternative algorithm implementations we tested did not include Lagrange multipliers, but included proximal terms for all c and r , yielding similar results.

The following algorithm implements the minimization posed in (P4). W , D , y_0 and c_0 are initialized as in the previous case.

$$\begin{aligned} c_{k+1} &= (I_{N^2} + \rho D^2)^{-1} (W_k y_0 + \rho D' r_k) \\ y_k &= W' c_k \\ r_{k+1, \text{row}\#i} &= \gamma_i (Dc_k)_{\text{row}\#i}, \quad \text{where} \\ \gamma_i &= \begin{cases} 1, & \text{if } \|(Dc_k)_{\text{row}\#i}\|_2^2 > 2\frac{\mu}{\rho} \\ 0, & \text{otherwise} \end{cases} \\ [U, S, V] &= \text{SVD}(c_k y_0' + \beta W_k), \quad W_{k+1} = U(V') \end{aligned}$$

4.2 Numerical results: denoising

The major difference between models M1 and M2 lies in the sparsity requirements. While in both cases we are looking for an optimal basis and associated representation, the weighted sparsity constraints incorporate different underlying assumptions about the data. In M1 we are recovering the basis in which all the images in the set are sparse in the weighted l^1 sense. In M2, however, we require *joint* weighted l^0 sparsity, thus enforcing the supports of the coefficient vectors for images to coincide or significantly overlap.

Both models naturally fit into the context of denoising. We consider two examples - denoising a grayscale image that contains large areas of repetitive texture and denoising a stack of hyperspectral images. The grayscale image 256×256 is split into q overlapping patches of smaller size $N \times N$, making sure that $q > N^2$ and applying both models gives meaningful basis recovery. Fig. 3 shows the results of applying M1 and M2 to denoise Barbara image with original noise level at $SNR = 20.62dB$ with $N = 16$, $q = 1024$ resulting in a denoised image with $SNR = 25.01dB$ and $SNR = 24.58dB$ respectively. For comparison we also show the outputs of the respective algorithms without the weighting.

To further explore the effectiveness of the proposed methods on multi-channel data with similarities between channels we applied models M1 and M2 to the problem denoising of a hyperspectral stack. A subset of size $64 \times 64 \times 162$ of the Urban hyperspectral set,¹¹¹² was picked, gaussian noise added to each band independently. Denoising was applied to the sub-stacks of size $8 \times 8 \times 162$, those were chosen with overlapping margins and the results were averaged respectively.

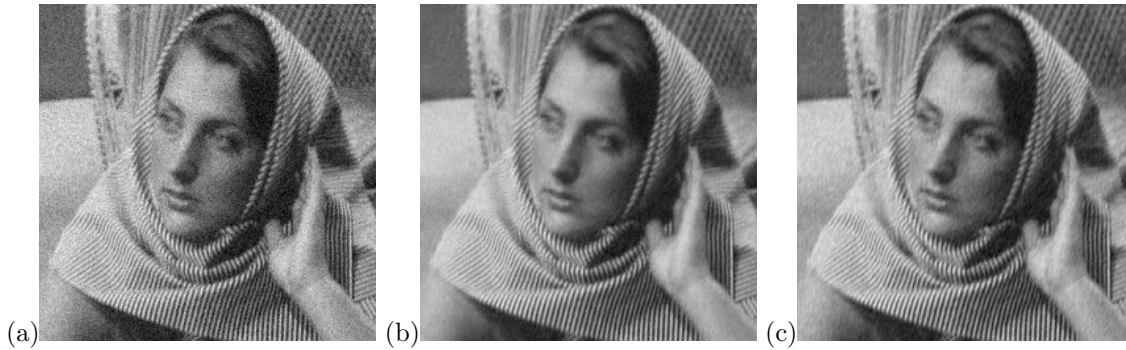


Figure 3. (a) Noisy image, SNR = 20.62dB, (b) recovery via M1 model, output SNR = 25.01dB, (c) recovery via M2 model, output SNR = 24.58dB

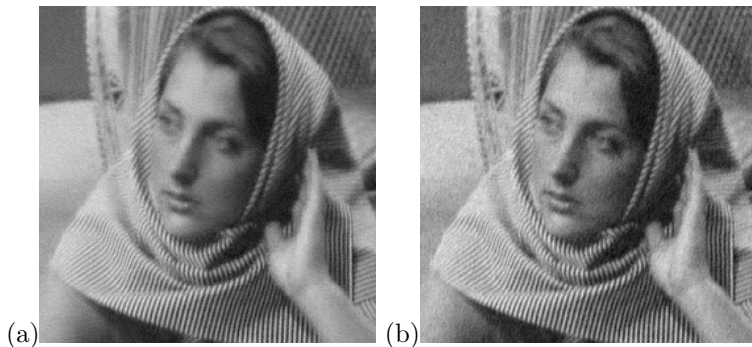


Figure 4. (a) recovery via M1 model without the weights($D = I$), output SNR = 23.69dB, (b) recovery via M2 model without the weights($D = I$), output SNR = 23.7dB

5. CONCLUSIONS AND FUTURE WORK

Smoothness or regularity and/or sparsity are the classical requirements we impose on the recovered signal in the modern image processing world. However, in the presence of additional information, such as similarities between multiple imaging data channels with a significant number of channels in relation to the image size, the notion of joint sparsity arises as a new requirement and a numerical constraint. In our models we implement two scenarios where we use updated versions of the joint sparsity constraint: weighted l^1 relaxed group sparsity and weighted joint sparsity. The former seems to perform better for the purposes of denoising in the context we presented, which is relevant to the problem of careful initialization needed for non-convex methods involving the weighted l^0 norm.

In our future work we would like to investigate the possibility of formulating a non-smooth analogue of Besov regularity for discrete data even when the basic building elements are not regular in the classical sense. We would like to see if the results of the fitting procedure can yield conclusions about the input data - both in terms of smoothness and relation between patches/bands/channels, as well as potential estimation of the intrinsic data dimension.

We will further investigate the applicability of these methods in the context of data with a priori low intrinsic dimension in the presence of high noise or altered via known lossy operators. As a high impact application we would like to apply our model to recovering a basis inducing the hierarchical joint sparsity in an image stack obtained within the MRF (magnetic resonance fingerprinting) context. The highly undersampled in the Fourier domain images in this problem are highly correlated and thus fit our sparsity framework. We would like to properly structure the lossy operators in the matrix form for the use as approximately linear transforms, recovering a patch of a stack at a time in order to fulfill the requirement of the meaningful joint sparsity requirement in the context of the optimal basis recovery.

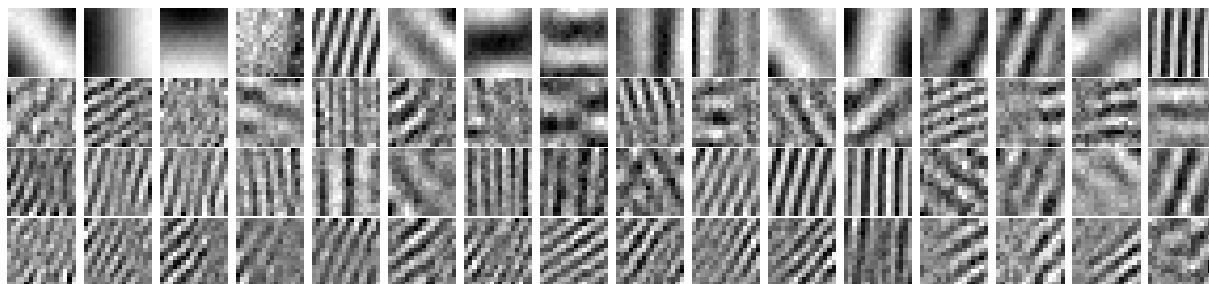


Figure 5. The first four images represent the basis elements that were assigned weight $d_i i = 1, i = 1, \dots, 4$, the next 12 - weight $d_i i = 2, i = 5, \dots, 16$; and the following 48 (assigned the weight $d_i i = 4, i = 17, \dots, 64$) to illustrate the tendencies further; the rest of the basis not shown.

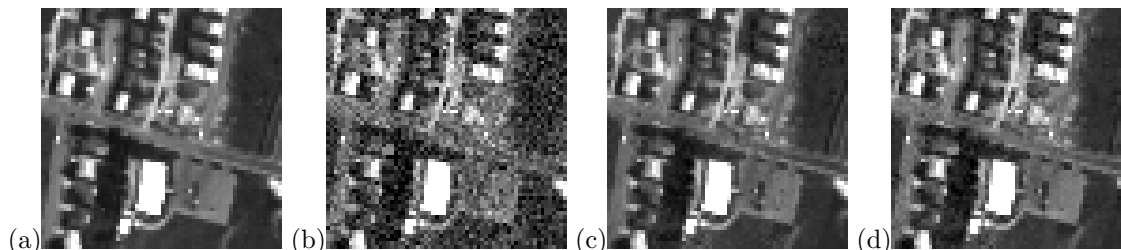


Figure 6. A band (the 7th band) from the hyperspectral stack (a) original image, (b) noisy image with $SNR = 14.27dB$, (c) recovery via M1 model, output $SNR = 22.14dB$ (d) recovery via M2 model, output $SNR = 21.51dB$

6. ACKNOWLEDGMENTS

Dr. Weihong Guo is partially supported by NSF DMS-1521582.

REFERENCES

- [1] Cai, J.-F., Ji, H., Shen, Z., and Ye, G.-B., "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis* **37**(1), 89 – 105 (2014).
- [2] Wang, J. and Cai, J.-F., "Data-driven tight frame for multi-channel images and its application to joint color-depth image reconstruction," *Journal of the Operations Research Society of China* **3**, 99–115 (2015).
- [3] Zhan, R. and Dong, B., "CT image reconstruction by spatial-radon domain data-driven tight frame regularization," *SIAM J. Imaging Sciences* **9**, 1063–1083 (2016).
- [4] Zhou, W., Cai, J.-F., and Gao, H., "Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography," *Inverse Problems* **29** (11 2013).
- [5] Choi, J., Bao, C., and Zhang, X., "PET-MRI joint reconstruction by joint sparsity based tight frame regularization," *SIAM Journal on Imaging Sciences* **11**(2), 1179–1204 (2018).
- [6] Mallat, S., [A Wavelet Tour of Signal Processing], Academic Press, Boston, 3 ed. (2009).
- [7] Meyer, Y., [Wavelets and Operators], Cambridge Univ. Press (1992).
- [8] Fornasier, M., [Theoretical Foundations and Numerical Methods for Sparse Recovery], Radon Series on Computational and Applied Mathematics, De Gruyter (2010).
- [9] Gabay, D. and Mercier, B., "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers and Mathematics with Applications* **2**(1), 17 – 40 (1976).
- [10] Glowinski, R. and Marroco, A., "Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite d'une classe de problemes de Dirichlet non lineaires," *ESAIM: Mathematical Modelling and Numerical Analysis - Revue francaise d'automatique, informatique, recherche operationnelle. Analyse numerique*. **9**(R2), 41–76 (1975).
- [11] Zhu, F., Wang, Y. I., Fan, B., Xiang, S., Meng, G., and Pan, C., "Spectral unmixing via data-guided sparsity," *IEEE Transactions on Image Processing* **23**, 5412–5427 (2014).
- [12] Zhu, F., Wang, Y., Xiang, S., Fan, B., and Pan, C., "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing* **88**, 101–118 (2014).

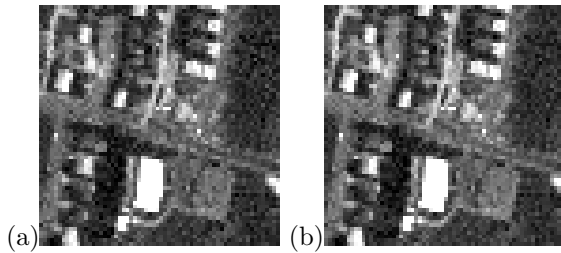


Figure 7. (a) comparison with unweighted l^1 model (7-th band of the stack), recovered $SNR = 17.25dB$ (b) comparison with unweighted l^1 model (7-th band of the stack), recovered $SNR = 18.25dB$

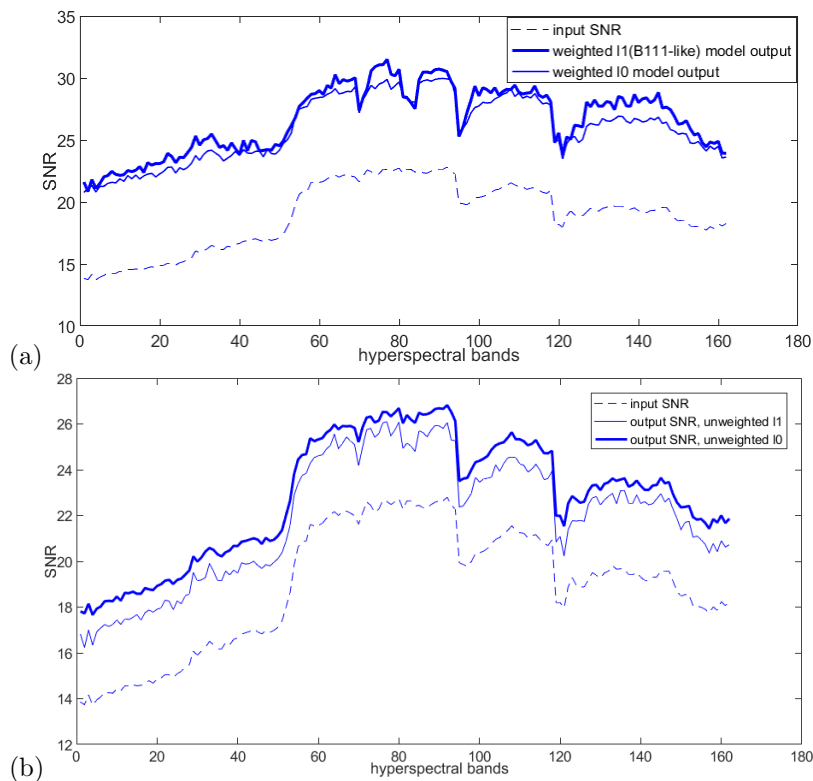


Figure 8. Improvement in SNR per band: (a) the models with weighted coefficients: after applying M1 the average increase in SNR is $7.764dB$, after applying M2 the average increase in SNR is $7.013dB$ (b) similar models with unweighted coefficients: unweighted l^1 minimization leads to the SNR increase averaging $3.015dB$, unweighted l^0 minimization leads to the SNR increase averaging $3.926dB$