ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10446773

LEARNING A CONVEX PATCH-BASED SYNTHESIS MODEL VIA DEEP EQUILIBRIUM

Stanislas Ducotterd *Sebastian Neumayer* *Michael Unser*

Biomedical Imaging Group,
École polytechnique fédérale de Lausanne (EPFL),
CH-1015 Lausanne, Switzerland

ABSTRACT

We investigate the learning of a convex patch-based synthesis model for the reconstruction of images. In essence, we propose to learn a dictionary via bilevel optimization for denoising. Using implicit differentiation, we find a closed-form formula of the derivative of the minimizer of an objective with respect to the atoms of the dictionary. We also propose a novel way to handle the mean of each patch of the predictions, which improves our model when it is applied to other inverse problems. For minimizing the objective involving the learnt dictionary, we propose an early stopping criterion to further improve the performance of the model for denoising. Finally, we assess our model in a compressed sensing MRI inverse problem and show that, despite being trained on denoising only, our model yields good reconstruction performances.

***Index Terms*—** Optimization, dictionary learning, denoising, image reconstruction

1. INTRODUCTION

This work aims at the learning of a synthesis model for the resolution of linear inverse problems [1]. Specifically, we consider the variational formulation

$$\mathbf{x}^* = \underset{\mathbf{x}\in C}{\arg\min} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + R(\mathbf{x}), \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n}$ denotes the measurement operator, $\mathbf{y} \in \mathbb{R}^m$ denotes the associated data, $R\colon \mathbb{R}^n \to \mathbb{R}^+$ is a regularizer that has to be specified, and $C \subseteq \mathbb{R}^n$ imposes constraints on the (unknown) reconstruction $\mathbf{x}^*$. Among the many possible choices for R, we pursue a synthesis-based approach [2], essentially assuming that each overlapping patch of the reconstruction $\mathbf{x}^*$ has a sparse representation in some dictionary. Note that, for this setting, the regularization in (1) can be also studied from an analysis perspective [3, 4] (see [5] for an in-depth discussion of the relation between the two approaches).

The research leading to these results was supported by the European Research Council (ERC) under European Union's Horizon 2020 (H2020), Grant Agreement - Project No 101020573 FunLearn and by the Swiss National Science Foundation, Grant 200020 184646/1.

Within the context of learned synthesis models (see [6, 7] for a survey), there are several approaches to find the dictionary in an unsupervised way, such as patch-based dictionary learning [8, 9, 10] and convolutional dictionary learning [11, 12]. While the former approaches usually extract patches from the image and enforce that these should have some sparse representation, the latter ones directly define the whole image in terms of a convolution of a dictionary with a sparse vector of coefficients. Even though the patch-based methods have substantially more parameters, they actually result in a more compact dictionary.

In recent years, data-driven approaches for solving inverse problems have become increasingly popular [13, 14]. There, a dictionary is usually learned in a supervised way such that the solution $\mathbf{x}^*$ of (1) minimizes an empirical loss over a dataset. In principle, this requires one to determine the derivative of $\mathbf{x}^*$ with respect to the dictionary, a feat which is not easy to achieve a priori. To the best of our knowledge, the supervised approaches in the literature are all convolutional, and they all rely on unrolling techniques to learn the dictionary [15, 16].

In this work, we present a scheme for the learning of a convex patch-based synthesis model. It relies on a combination of unrolling and deep equilibrium (DEQ) [17]. Moreover, we handle the mean of each patch of $\mathbf{x}$ in a novel way, which improves the performance of our model. Once the dictionary is learned, we evaluate the model using different optimization algorithms on denoising and compressed sensing MRI (CS-MRI). We also compare our results with CRR [4], a learned convex analysis model. The code is available on Github[1].

2. METHOD

The matrix $\mathbf{P}_k$ extracts a patch of the image $\mathbf{x}$ at a given location. We seek to learn a dictionary of zero-mean atoms $\mathbf{D} \in \mathbb{R}^{d\times p}$ such that any minimizer $(\mathbf{x}^*, \{\boldsymbol{\alpha}_k^*\}_{k=1}^N) \in C \times (\mathbb{R}^p)^N$ of the convex objective

$$J_{\mathbf{D},\mathbf{y}}(\mathbf{x}, \boldsymbol{\alpha}) = \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + \sum_{k=1}^{N} \beta \mathcal{S}_{\mathbf{D},k}(\mathbf{x}, \boldsymbol{\alpha}_k) + \lambda\|\boldsymbol{\alpha}_k\|_1 \tag{2}$$

[1] https://github.com/StanislasDucotterd/Dictionnary_Learning

with

$$\mathcal{S}_{\mathbf{D},k}(\mathbf{x},\boldsymbol{\alpha}) = \min_{c\in\mathbb{R}} \frac{1}{2}\|\mathbf{P}_k\mathbf{x} - \mathbf{D}\boldsymbol{\alpha} - \mathbf{1}c\|^2 \tag{3}$$

corresponds to a high-quality reconstruction $\mathbf{x}^*$ of the data $\mathbf{y}$. For a given $\mathbf{x}$, it holds that $c_k^* = \frac{1}{d}\mathbf{1}^T\mathbf{P}_k\mathbf{x}$. Hence, we replace $\mathbf{P}_k$ by $\widetilde{\mathbf{P}}_k = (\mathbf{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^T)\mathbf{P}_k$ to get $\mathcal{S}_{\mathbf{D},k}(\mathbf{x},\boldsymbol{\alpha}) = \|\widetilde{\mathbf{P}}_k\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|^2$. While this approach is similar to the usual strategy that sets $c = \frac{1}{d}\mathbf{1}^T\mathbf{P}_k\mathbf{H}^T\mathbf{y}$ when $\mathbf{H} = \mathbf{I}$ [10, 18], it leads to very different results when the task differs from a basic denoising. Throughout this paper, we denote every method that uses our novel handling of the mean with the suffix ++.

The learning of the dictionary $\mathbf{D}$ can be achieved via bilevel optimization. More precisely, given some training samples $(\mathbf{x}_l, \mathbf{y}_l) \in \mathbb{R}^n \times \mathbb{R}^m$, we minimize the reconstruction loss

$$\mathcal{L}(\mathbf{D}) = \sum_l \|\mathbf{x}_l - \mathbf{x}_l^*\|_1, \tag{4}$$

where

$$\mathbf{x}_l^* \in \arg\min_{\mathbf{x}\in C}\big(\min_{\boldsymbol{\alpha}} J_{\mathbf{D},\mathbf{y}_l}(\mathbf{x},\boldsymbol{\alpha})\big). \tag{5}$$

This task consists of two subproblems:

- the inner problem (5), namely a search for the optimal $(\mathbf{x}^*, \{\boldsymbol{\alpha}_k^*\}_{k=1}^N)$ with classical optimization algorithms;
- the outer problem (4), which consists of the learning of the dictionary $\mathbf{D}$.

Note that [19] considers an objective very similar to (2), but relies on unrolling to learn the dictionary $\mathbf{D}$ instead.

2.1. Training of the Model—Inner Optimization

We minimize (2) by successive coordinate descent as;

$$\boldsymbol{\alpha}_k^{(m+1)} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \mathcal{S}_{\mathbf{D},k}(\mathbf{x}^{(m)},\boldsymbol{\alpha}) + \frac{\lambda}{\beta}\|\boldsymbol{\alpha}\|_1 \tag{6}$$

$$\mathbf{x}^{(m+1)} = \arg\min_{\mathbf{x}\in C} \|\mathbf{H}\mathbf{x}-\mathbf{y}\|^2 + \beta\sum_{k=1}^N \mathcal{S}_{\mathbf{D},k}(\mathbf{x},\boldsymbol{\alpha}_k^{(m+1)}). \tag{7}$$

The solutions of (6) and (7) are computed with FISTA [20]. Specifically, the coefficients $\boldsymbol{\alpha}_k^{(m+1)}$ are the fixed points of

$$f_1(\boldsymbol{\alpha}) = S_{\frac{\lambda}{\beta L_1}}\Big(\boldsymbol{\alpha} - \frac{1}{L_1}\boldsymbol{\nabla}_{\boldsymbol{\alpha}}\mathcal{S}_{\mathbf{D},k}(\mathbf{x}^{(m)},\boldsymbol{\alpha})\Big), \tag{8}$$

$k = 1,\dots,N$, while the image $\mathbf{x}^{(m+1)}$ is the fixed point of

$$f_2(\mathbf{x}) = \mathcal{P}_C\Big(\mathbf{x} - \frac{1}{L_2}\boldsymbol{\nabla}_{\mathbf{x}}J_{\mathbf{D},\mathbf{y}}(\mathbf{x},\boldsymbol{\alpha}^{(m+1)})\Big). \tag{9}$$

Here, S_λ denotes the soft-thresholding function and is the proximal operator of the ℓ_1 norm, $\mathcal{P}_C$ denotes the projection onto the set C, $L_1 = \|\mathbf{D}^T\mathbf{D}\|_2$, and $L_2 = \|\mathbf{H}^T\mathbf{H} + \beta\sum_{k=1}^N \widetilde{\mathbf{P}}_k^T\widetilde{\mathbf{P}}_k\|_2$. The fixed-point iterations for (8) and (9) are run until the relative difference in the norms of the iterates is smaller than 10^{-4} and 10^{-6}, respectively.

2.2. Training of the Model—Outer Optimization

Let $\mathbf{d}_k$ denote the kth column vector of $\mathbf{D}$. During optimization, the dictionary is constrained to be in

$$\mathcal{B} = \left\{\mathbf{D}\in\mathbb{R}^{d\times p} : \|\mathbf{D}\|_2 = 1, \|\mathbf{d}_k\| = \|\mathbf{d}_1\|\,\forall 1\le k\le p\right\}. \tag{10}$$

In this way, we do not have to deal with the Lipschitz constant L_1 in the inner optimization. Moreover, this ensures that all atoms have the same norm.

As outlined in the DEQ framework, we can differentiate any $\boldsymbol{\alpha}_k^{(m+1)}$ with respect to $\mathbf{D}$ for any $\mathbf{x}^{(m)}$ through the implicit differentiation. More precisely, we obtain

$$\frac{\partial\boldsymbol{\alpha}_k^{(m+1)}}{\partial\mathbf{D}} = \big(\mathbf{I} - \mathbf{J}_{f_1}(\boldsymbol{\alpha}_k^{(m+1)})\big)^{-1}\frac{\partial f_1}{\partial\mathbf{D}} \tag{11}$$

with the Jacobian

$$\mathbf{J}_{f_1}(\boldsymbol{\alpha}_k^{(m+1)}) = \mathbf{I} - (\mathbf{I} - \mathbf{D}^T\mathbf{D})\mathbf{V} \tag{12}$$

of the function (8), and where $\mathbf{V}$ is a diagonal matrix whose values are $S'_{\lambda/\beta}(\boldsymbol{\alpha}_k^{(m+1)} - \mathbf{D}^T(\mathbf{D}\boldsymbol{\alpha}_k^{(m+1)} - \widetilde{\mathbf{P}}_k\mathbf{x}^{(m)}))$. Unlike most DEQ models, which use Anderson iterations to approximate (11), we rely on a closed-form solution instead. To update the dictionary $\mathbf{D}$, we run two iterations of the coordinate descent (6)-(7). Hence, computing the gradient with respect to $\mathbf{D}$ involves computing the inverse of the Jacobian (12) twice. During the training, we set $C = \mathbb{R}^n$. This allows us to obtain the minimizer of (7) with respect to $\boldsymbol{\alpha}_k^{(m+1)}$ in terms of a simple linear operator and makes the training simpler than using an implicit differentiation step instead.

2.3. Full Minimization

At inference, we minimize the objective (2) up to numerical precision using the iPALM algorithm [21] with the updates

$$\boldsymbol{\beta}_k^{(m)} = \boldsymbol{\alpha}_k^{(m)} + \frac{m-1}{m+2}\big(\boldsymbol{\alpha}_k^{(m)} - \boldsymbol{\alpha}_k^{(m-1)}\big) \tag{13}$$

$$\boldsymbol{\alpha}_k^{(m+1)} = S_{\frac{\lambda}{\beta c_1}}\Big(\boldsymbol{\beta}_k^{(m)} - \frac{1}{c_1}\boldsymbol{\nabla}_{\boldsymbol{\alpha}}\mathcal{S}_{\mathbf{D},k}(\mathbf{x}^{(m)},\boldsymbol{\beta}_k^{(m)})\Big) \tag{14}$$

$$\mathbf{z}^{(m)} = \mathbf{x}_k^{(m)} + \frac{m-1}{m+2}\big(\mathbf{x}_k^{(m)} - \mathbf{x}_k^{(m-1)}\big) \tag{15}$$

$$\mathbf{x}^{(m+1)} = \mathcal{P}_C\Big(\mathbf{z}^{(m)} - \frac{1}{c_2}\boldsymbol{\nabla}_{\mathbf{x}}J_{\mathbf{D},\mathbf{y}}(\mathbf{z}^{(m)},\boldsymbol{\alpha}^{(m+1)})\Big), \tag{16}$$

where $c_1 = 1.01L_1$ and $c_2 = 1.01L_2$. This approach is orders of magnitude faster than coordinate descent (6)-(7). The convergence of iPALM is guaranteed by [21, Theorem 4.1].

3. EXPERIMENTS

3.1. Denoising

We learn the dictionary $\mathbf{D}$ based on denoising. The training dataset consists of 238400 patches of size (40×40) taken

from the BSD500 dataset [22], which consists of images with values in $[0, 1]$. For our experiment, we add Gaussian noise with $\sigma = 5/255, 25/255$, we use 128 atoms of size (10×10), and we set $C = [0, 1]^n$. We report the performance of three different algorithms.

- 2-Step Unrolling: two steps of the coordinate descent (6)-(7);
- Early Stopping: coordinate descent until the relative difference in the norms of the iterates is less than 10^{-4};
- Full Minimization: iPALM (13)-(16) until the relative difference in the objective value (2) is less than 10^{-8}.

The 2-step unrolling is also used for learning $\mathbf{D}$, and the full minimization finds a global minimum of the associated objective (2) up to numerical precision. Indeed, we observe no change in objective value or peak signal-to-noise ratio (PSNR) if we continue the optimization after the convergence criterion is met. Refinements of our simple early stopping strategy will be considered in future research. The inference time on GPU for the 2-step unrolling and the full minimization is in the order of seconds, and in the order of minutes for the early stopping. We compare the results with TV regularization (using the algorithm proposed in [23]), and the learned convex analysis model CRR [4]. The regularization parameters β and λ are tuned on the BSD validation set of 12 images via a grid search, and the denoising performances are then reported on the BSD68 test set in Table 1. The same is done for the hyperparameters of the competing methods. Surprisingly, despite of it being trained on two unrolled steps of coordinate descent, the model performs better if we run more steps at inference. However, the performance eventually decreases if we push the optimization all the way to the minimum. For illustration purposes, we plot the PSNR of $\mathbf{x}^{(m)}$ as a function of the objective value for the coordinate descent and the iPALM algorithms on the *boat* image from the BSD validation dataset in Figure 1. Despite reaching the same minimum, the two algorithms have a very different relation between objective value and PSNR. A similar behavior was observed for the remaining test images. The analysis CRR model perfoms better by 0.1 dB on the higher noise level but is outperformed by 0.2 dB on the lower one.

3.2. Compressed Sensing MRI

We now look at the CS-MRI recovery of an image $\mathbf{x} \in \mathbb{R}^n$ from its measurements $\mathbf{y} = \mathbf{M}\mathbf{F}\mathbf{x} + \mathbf{n} \in \mathbb{C}^m$, where $\mathbf{M}$ is a subsampling mask (identity matrix with some missing entries), $\mathbf{F}$ is the discrete Fourier transform, and $\mathbf{n}$ is a complex Gaussian noise with variance $\sigma_{\mathbf{n}}^2$ (not to be confused with σ, the noise variance used for the denoising) for both the real and imaginary parts. To reconstruct $\mathbf{x}$, we minimize the objective (2) with $\mathbf{H} = \mathbf{M}\mathbf{F}$ using the dictionary $\mathbf{D}$ learned for the denoising. The involved regularization parameters β and λ are tuned for each image via a grid search. The same is done for the hyperparameters of the competing methods.

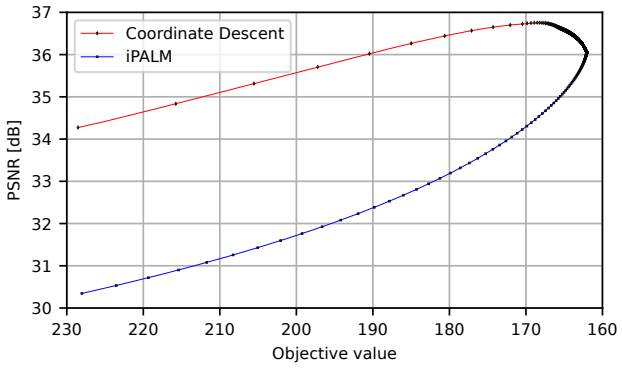


Fig. 1. Relation between objective value and PSNR. The markers correspond to 5 iterations of the algorithm. For better visualization, we only show the end of the optimization process, where coordinate descent iterations decrease the PSNR.

We run the CS-MRI experiment on two images of size (256×256) that take values between $[0, 1]$. Moreover, we consider two subsampling masks (radial and Cartesian), each with a subsampling ratio of 0.3 with $\sigma_{\mathbf{n}} = 10/255$. The corresponding results are reported in Table 2 and a reconstruction example is given in Figure 2. We see that our handling of the mean improves the performance by 0.4 to 1.5 dB, depending on the image and the mask. For the TV regularization, we get the typical staircasing artifacts, which do not occur with our learned synthesis approach. Further, note that our results are comparable to the analysis-based CRR approach.

Unlike denoising, we achieved the best performances while running the optimization until the global minimum of (2) is reached. Hence, we only report the results of Full Minization. Surprisingly, even though our models were trained on denoising, the performance gap between TV regularization and our proposed method increases for image reconstruction. The analysis CRR model performs better by 0.3 dB on the bust/radial problem (which is the simplest one) and is outperformed by 0.15 dB on the brain/Cartesian one (which is the hardest one). The performance metrics behave similarly on the two other problems.

3.3. Visualization of the Atoms

We show the learned atoms of $\mathbf{D}$ in Figure 3. They look quite different from the ones learned in an unsupervized way with the ℓ_0 regularizer in [8]. Most of the atoms for σ=5/255 seem to only have structure in the middle of the patch, which is surprising considering that the performances dropped when using smaller patches. The atoms for σ=25/255 are surprisingly noisy, which potentially allows to better fit some of the noise in the training data. In summary, the atoms learned for the two different noise levels are rather different.

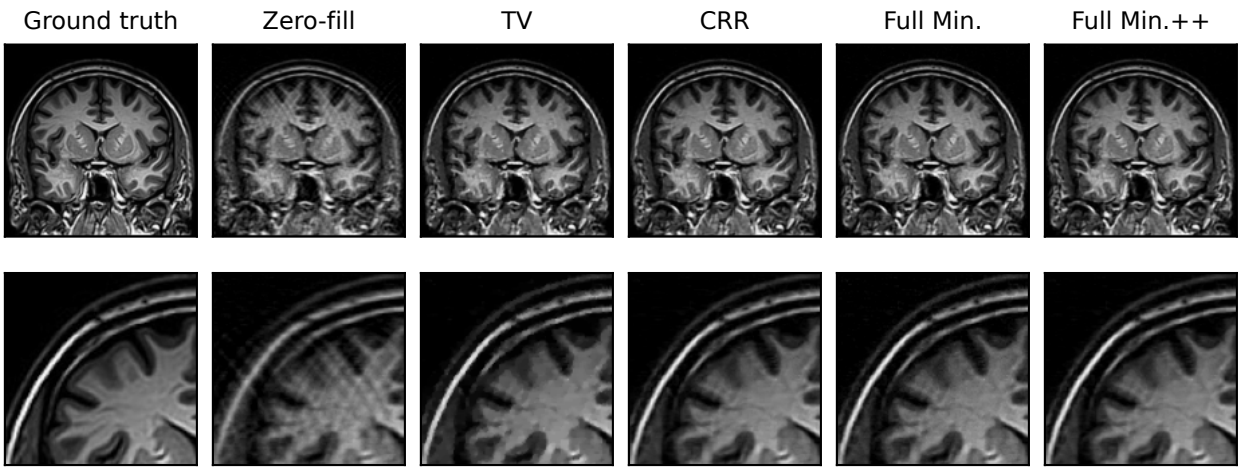


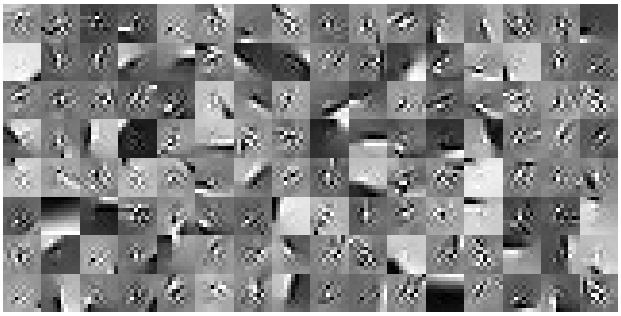
Fig. 2. MRI reconstructions on the brain image with the Cartesian subsampling mask. The PSNR values are given in Table 2.

Noise level	σ=5/255	σ=25/255
TV	36.41	27.48
CRR	36.96	**28.11**
2-Step Unrolling	36.91	28.01
2-Step Unrolling++	36.86	27.93
Early Stopping	**37.16**	27.81
Early Stopping++	**37.16**	27.99
Full Min.	36.82	27.56
Full Min.++	36.83	27.64

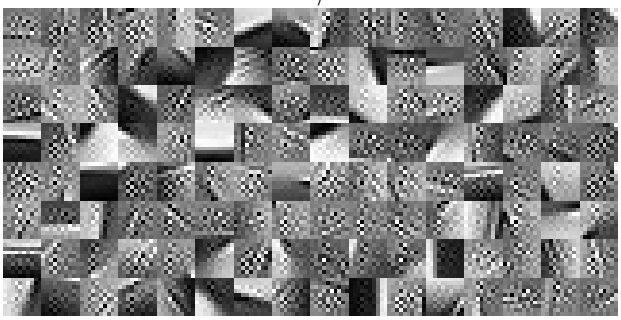
Table 1. PSNR values for denoising the BSD68 dataset.

Subsampling mask	Radial		Cartesian	
Image type	Brain	Bust	Brain	Bust
Zero-filling	24.01	25.32	21.66	23.55
TV	29.46	31.58	24.43	27.69
CRR (σ=5/255)	31.21	**32.71**	25.10	28.37
CRR (σ=25/255)	**31.30**	32.65	25.35	**28.46**
Full Min. (σ=5/255)	30.33	30.94	24.78	27.64
Full Min.++ (σ=5/255)	31.05	32.42	25.21	28.31
Full Min. (σ=25/255)	30.47	31.02	25.11	27.90
Full Min.++ (σ=25/255)	31.27	32.41	**25.48**	28.39

Table 2. PSNR values for MRI reconstruction.



$\sigma = 5/255$



$\sigma = 25/255$

Fig. 3. The learned atoms for two noise levels.

4. CONCLUSION AND FUTURE WORK

We learned a convex patch-based synthesis model on a denoising task by using a combination of unrolling and deep equilibrium. Remarkably, our novel mean handling improves the performance of our model when it is applied to other tasks than denoising, at no extra computational cost.

A future direction is to directly learn the atoms such that the global minimum of (2) found with the iPALM algorithm corresponds to the denoised image. Compared to the deployed unrolled coordinate descent, this is computationally more expensive and involves a more challenging joint implicit differentiation with respect to all the variables $(\mathbf{x}, \{\boldsymbol{\alpha}_k\}_{k=1}^N)$. Another direction would be to complexify the regularizer. In this paper, we worked with a very simple ℓ_1 norm regularizer that penalizes each atom the same way. An improved model would be to adapt the penalization to each atom. To go even further, one could learn the whole regularizer R with linear splines as in the CRR analysis model [4].

5. REFERENCES

[1] Alejandro Ribes and Francis Schmitt, "Linear inverse problems in imaging," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 84–99, 2008.

[2] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[3] Yunjin Chen, Rene Ranftl, and Thomas Pock, "Insights into analysis operator learning: From patch-based sparse models to higher order MRFs," *IEEE Transactions on Image Processing*, vol. 23, pp. 1060–72, 2014.

[4] Alexis Goujon, Sebastian Neumayer, Pakshal Bohra, Stanislas Ducotterd, and Michael Unser, "A neural-network-based convex regularizer for inverse problems," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 781–795, 2023.

[5] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[6] Ivana Tošić and Pascal Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[7] Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[8] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[9] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[10] Yangyang Xu and Wotao Yin, "A fast patch-dictionary method for whole image recovery," *Inverse Problems and Imaging*, vol. 10, no. 2, pp. 563–583, 2016.

[11] Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad, "Convolutional dictionary learning via local processing," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5296–5304.

[12] Cristina Garcia-Cardona and Brendt Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[13] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[14] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A. Fessler, "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2019.

[15] Benoît Malézieux, Thomas Moreau, and Matthieu Kowalski, "Understanding approximate and unrolled dictionary learning for pattern recovery," in *International Conference on Learning Representations*, 2022.

[16] Bahareh Tolooshams and Demba E. Ba, "Stable and interpretable unrolled dictionary learning," *Transactions on Machine Learning Research*, 2022.

[17] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun, "Deep equilibrium models," in *Advances in Neural Information Processing Systems*. 2019, vol. 32, Curran Associates, Inc.

[18] Bihan Wen, Saiprasad Ravishankar, and Yoram Bresler, "Structured Overcomplete Sparsifying Transform Learning with Convergence Guarantees and Applications," *International Journal of Computer Vision*, vol. 114, no. 2, pp. 137–167, 2015.

[19] Andreas Kofler, Marie-Christine Pali, Tobias Schaeffter, and Christoph Kolbitsch, "Deep supervised dictionary learning by algorithm unrolling—Application to fast 2D dynamic MR image reconstruction," *Medical Physics*, vol. 50, no. 5, pp. 2939–2960, 2023.

[20] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[21] Thomas Pock and Shoham Sabach, "Inertial Proximal Alternating Linearized Minimization (iPALM) for Nonconvex and Nonsmooth Problems," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1756–1787, 2016.

[22] Pablo Arbeláez, Michael Maire, Charless Fowlkes, and Jitendra Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.

[23] Antonin Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.