# MATCHING PURSUIT BASED CONVOLUTIONAL SPARSE CODING

*Elad Plaut and Raja Giryes*

School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

**ABSTRACT**

Convolutional sparse coding using the $\ell_{0,\infty}$ norm has been described as "a problem that operates locally while thinking globally". In this paper, we present a matching pursuit based greedy algorithm specifically tailored to the $\ell_{0,\infty}$ norm. We also propose a corresponding dictionary learning algorithm, which trains a local dictionary on a set of global images. Our approach is based on the convolutional relationship between the local dictionary and the global image. It operates locally while taking into account the global nature of the images. We demonstrate the usage of our proposed strategy for the task of image inpainting.

***Index Terms*—** Convolutional Sparse Coding, Global modeling, Local Processing, Greedy Algorithms, Sparse Representations

## 1. INTRODUCTION

Sparse coding can be described as solving the following minimization problem, known as the $P_0^\epsilon$ problem:

$$(P_0^\epsilon): \quad \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \ s.t. \ \|\mathbf{x} - \mathrm{D}\boldsymbol{\alpha}\|_2^2 \leq \epsilon,$$

where $\boldsymbol{\alpha} \in \mathbb{R}^p$ is a sparse representation of a signal $\mathbf{x} \in \mathbb{R}^N$ in the dictionary $\mathrm{D} \in \mathbb{R}^{N \times p}$. The columns of D are referred to as atoms and are assumed to have unit $\ell_2$ norm. The $\ell_0$ (pseudo-)norm returns the number of nonzero coefficients in a vector, also called the sparsity.

As $P_0^\epsilon$ is an NP-hard problem, several approximation approaches have been proposed. One of them is the greedy approach, which includes matching pursuit [1], and another one uses a convex relaxation to the $\ell_1$ norm [2].

When training a dictionary on a set of vectors $\{\mathbf{x}_i\}_{i=1}^s$, one typically solves the minimization problem

$$\min_{\mathrm{D},\{\boldsymbol{\alpha}_i\}_{i=1}^s} \sum_{i=1}^{s} \|\mathbf{x}_i - \mathrm{D}\boldsymbol{\alpha}_i\|_2^2 \ s.t. \ \|\boldsymbol{\alpha}_i\|_0 \leq k \ , \ 1 \leq i \leq s.$$

A common approach to solve this problem is repeatedly alternating between optimizing the dictionary with the sparse representations held fixed, and optimizing the sparse representations with the dictionary held fixed.

This work is supported in part by the ERC-StG SPADE grant and the MAGNET MDM grant.

### 1.1. Local Sparse Coding

In practice, dictionaries are only learned on relatively small signals due to computational complexity, memory requirement and the required quantity of training signals. Thus, an image of size $M \times N$ is usually divided into overlapping patches of size $m \times n$. Sparse coding is performed on each patch separately with a local dictionary $\mathrm{D}_L \in \mathbb{R}^{mn \times p}$, and the reconstructed image is taken as the average over the overlapping patches. This approach is fundamentally limited as it applies the sparse model to patches only, and does not take into account the dependencies between them. While some methods have been proposed for globalizing the prior [3], [4], [5], the current dominant solution seeks a shift-invariant dictionary.

### 1.2. Convolutional Sparse Coding

In order to address these problems, a convolutional approach has recently been developed [6], [7], [8], [9], demonstrating state-of-the-art performance in various applications [10], [11], [12]. A local dictionary $\mathrm{D}_L$ is used to represent a global signal by including all shifts of the local atoms: $\mathbf{x} = \mathrm{D}\boldsymbol{\alpha} = \sum_{j=1}^{p} \mathrm{D}_j \boldsymbol{\alpha}_j = \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j$, where $\boldsymbol{\alpha}$ is the global sparse representation vector, $\mathrm{D}_j$ is the convolution matrix of the local atom $\mathbf{d}_j$, $\boldsymbol{\alpha}_j$ is the vector of the coefficients multiplying $\mathbf{d}_j$ at each of its shifts within the global signal, and $\mathrm{D} = \begin{bmatrix} \mathrm{D}_1 & \mathrm{D}_2 & \dots & \mathrm{D}_p \end{bmatrix}$ is the global dictionary, also referred to as the convolutional dictionary (in the case of two-dimensional signals it is a block-Toeplitz matrix). The projections of the global signal onto D can be computed efficiently using convolution operations: $\mathbf{b}_j = \mathrm{D}_j^T \mathbf{x} = \overleftrightarrow{\mathbf{d}_j} * \mathbf{x}$, where $\overleftrightarrow{\mathbf{d}_j}$ is the column-stacked horizontally and vertically flipped atom $\mathbf{d}_j$. Thus, there is no need to store or use the explicit form of D for matrix multiplication.

With the above formulation, the sparse coding problem becomes:

$$\min_{\{\boldsymbol{\alpha}_j\}} \left\| \mathbf{x} - \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j \right\|_2^2 \ s.t. \ \sum_{j=1}^{p} \|\boldsymbol{\alpha}_j\|_0 \leq k \ . \quad (1)$$

Greedy solutions to (1) and similar forms have been proposed in [13], [14], [15], [16], and [17]. They are based on matching pursuit or orthogonal matching pursuit, with the projections computed efficiently as convolutions. However, even with efficient implementations, the complexity of such a greedy algorithm is prohibitive: $O(MNpk)$. This is especially true for large global images, where the global sparsity is large. Consequently, all other algorithms surveyed in [18] and [19] use an $\ell_1$ relaxation and minimize its unconstrained Lagrangian.

By imposing a constraint on the $\ell_0$ or $\ell_1$ norm of a global image, we can achieve a globally sparse representation while allowing all possible shifts of the local dictionary. However, the selected atoms may be concentrated in some areas of the image, while others may be very sparse. Thus, when using a convolutional dictionary, the global sparsity alone is a poor indication of error measures of images which take into account the structure of the whole image.

In addition, previous works [20], [21] have shown that pursuit algorithms are guaranteed to succeed as long as the sparsity is lower than a certain threshold. Therefore, a local pursuit method that is able to succeed in the sparser patches, which might cover most of the global image, could fail in its denser patches. Thus, it is tempting to impose a constraint on sparsity at the local level, while taking into account the global structure. For this purpose, we need to constrain a norm that induces sparsity at the local level across the whole global signal. Such an approach was proposed in [22], which defines the $\ell_{0,\infty}$ norm.

### 1.3. The $P^{\epsilon}_{0,\infty}$ problem

The $P^{\epsilon}_{0,\infty}$ problem is described in [22] as "a problem that operates locally while thinking globally":

$$\left(P^{\epsilon}_{0,\infty}\right): \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_{0,\infty} \ s.t. \ \|\mathbf{x} - \mathrm{D}\boldsymbol{\alpha}\|_2^2 \leq \epsilon. \qquad (2)$$

Recall that the global dictionary D is a concatenation of the convolution matrices of the local atoms of size $m \times n$, zero-padded to the size of the global image ($M \times N$) and column-stacked. We define a stripe $\Omega_i$ as a set of $(2m-1)(2n-1)p$ indices of columns in D who may have nonzero values at the row corresponding to the pixel $i$ ($1 \leq i \leq MN$). Only nonzero coefficients of $\boldsymbol{\alpha}$ with these indices contribute to the pixel $x_i$. The $\ell_{0,\infty}$ (pseudo-)norm of the vector $\boldsymbol{\alpha}$, called the *stripe-sparsity* and denoted $\|\boldsymbol{\alpha}\|_{0,\infty}$, is the number of nonzero coefficients in the densest stripe, or equivalently the maximum number of atoms contributing to any pixel. By limiting the sparsity of the densest stripe, we are effectively limiting the sparsity of all stripes, and therefore the number of overlaps in all pixels of the global representation.

The work in [22] proposes two optimization algorithms for solving the $\ell_1$ relaxed unconstrained version of (2). In [23], a corresponding dictionary learning algorithm is proposed, which also minimizes an unconstrained penalized Lagrangian. In [24], the $\ell_{1,\infty}$, which is a convex relaxation of the $\ell_{0,\infty}$ norm, is shown to underperform the simple $\ell_1$ norm regularization for white Gaussian noise denoising.

### 1.4. Contribution

Greedy strategies have so far been proposed for solving the standard convolutional sparse coding problem (1), which as mentioned above is computationally demanding. In this work, we propose to overcome this issue by introducing a novel greedy sparse coding scheme for the $\ell_{0,\infty}$ norm constrained minimization problem (2). To the best of our knowledge, all strategies proposed for this scheme so far have relied on $\ell_1$ relaxation and unconstrained optimization [22], [23]. The convolutional sparse recovery scheme proposed in this work has the following desirable properties: (i) as a greedy algorithm, it allows direct control over the squared error and the $\ell_{0,\infty}$ sparsity; (ii) it recovers efficiently multiple sparse representation coefficients at the scale of the image by projecting onto the local dictionary only once; (iii) it introduces an MOD-like dictionary learning approach for a dictionary update step, separate from the sparse coding step.

An extended version of this paper with another dictionary learning approach and additional experiments appears in [25].

## 2. A GLOBAL CONVOLUTIONAL GREEDY PURSUIT

Recall the standard matching pursuit algorithm: (i) calculate projections of the signal onto all atoms; (ii) select the one with the largest absolute value and use the inner product as a coefficient; (iii) subtract the projection from the signal, creating residual vector; and (iv) repeat steps (i)-(iii) for the residual until a stopping condition is reached.

In the case of the $P_{0,\infty}$ problem, once we select a single atom, we might as well add many more atoms until we can no longer add any more without creating overlaps. After selecting the atom with the most significant projection, we enforce the non-overlapping constraint by discarding all atoms that overlap it (all indices in its stripe, denoted $\Omega_{i^*}$). We exclude atoms by zeroing their projections, thus ensuring they will not be selected. Notice that the selected atoms are exactly the same ones that would have been selected if we had recomputed the projection onto the dictionary for each added atom separately. All atoms that remain in the global dictionary are orthogonal to the selected atom, so there is no need to recalculate the projections. Note also that each atom increases the $\ell_0$ norm while holding the $\ell_{0,\infty}$ norm fixed and reducing the representation error. We define a layer as such a representation without overlaps.

If the stopping condition has not yet been reached, we add the next series of atoms, thus increasing the $\ell_{0,\infty}$ norm by one. We do this by recomputing the projections of the residual onto the dictionary (using convolution operations), sorting them, and constructing an additional layer of non-overlapping atoms. When there remain no atoms in the dictionary to be selected, we have completed one layer. Then we subtract each reconstructed layer from the image before calculating the projections of the residual onto the convolutional dictionary, and we repeat until a stopping condition is reached. The resulting algorithm is summarized in Algorithm 1, which we call Global Convolutional Matching Pursuit (GCMP).

**Algorithm 1** Global Convolutional Matching Pursuit

$\boldsymbol{\alpha} \leftarrow \mathbf{0}$
$\mathbf{r} \leftarrow \mathbf{x}$
$k \leftarrow 0$
**while** $\|\mathbf{r}\|_2 > \epsilon$ **do**
  $\mathbf{b} \leftarrow \left[ \begin{array}{cccc} \overleftrightarrow{\mathbf{d}}_1 * \mathbf{r} & \overleftrightarrow{\mathbf{d}}_2 * \mathbf{r} & \dots & \overleftrightarrow{\mathbf{d}}_p * \mathbf{r} \end{array} \right]^T$
  **while** $\max_i \ |b_i| > 0$ **do**
    $i^* \leftarrow \arg\max_i \ |b_i|$
    $\alpha_{i^*} \leftarrow \alpha_{i^*} + b_{i^*}$
    **for** $i \in \Omega_{i^*}$ **do**
      $b_i \leftarrow 0$
    **end for**
  **end while**
  $\mathbf{r} \leftarrow \mathbf{x} - \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j$
  $k \leftarrow k + 1$
**end while**

Note that we compute the projection of the residual onto the dictionary only once per outer loop, and not per atom. Consequently, the order of complexity is $O\left(MNp \|\boldsymbol{\alpha}\|_{0,\infty}\right)$, which is much lower than the $O(MNpk)$ required by MP. Moreover, by construction, the resulting solution obeys

$$\frac{MN}{2mn}\left(\|\boldsymbol{\alpha}\|_{0,\infty} - 1\right) \leq k \leq \frac{MN}{mn}\|\boldsymbol{\alpha}\|_{0,\infty} \ . \tag{3}$$

## 3. DICTIONARY LEARNING

We propose a dictionary learning approach, which learns offline a local dictionary on a training set of global signals (rather than on patches). Let $\left\{\mathbf{x}^{(i)}\right\}$ be a set of $s$ global signals and let $\mathrm{D}_L$ be a local dictionary with $p$ columns $\{\mathbf{d}_j\}$. Each global signal is represented as a sum of convolutions:

$$\mathbf{x}^{(i)} = \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j^{(i)}$$

The dictionary is trained by optimizing both the sparse representation and the dictionary:

$$\begin{gathered}\min_{\left\{\boldsymbol{\alpha}^{(i)}\right\},\{\mathbf{d}_j\}} \sum_{i=1}^{s} \left\| \mathbf{x}^{(i)} - \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j^{(i)} \right\|_2^2 \\ s.t. \ \left\|\boldsymbol{\alpha}^{(i)}\right\|_{0,\infty} \leq K, \ 1 \leq i \leq s,\end{gathered} \tag{4}$$

where $\boldsymbol{\alpha}^{(i)}$ is a vector formed by vertically stacking $\boldsymbol{\alpha}_j^{(i)}$ for all $j$. To solve (4), we repeatedly alternate between a sparse coding step with a constraint on the $\ell_{0,\infty}$ norm, and a dictionary update step. The sparse coding problem is solved by using GCMP (Algorithm 1) for each of the $s$ global images separately. The dictionary update step minimizes the sum of total squared representation errors for the whole dictionary:

$$\min_{\{\mathbf{d}_j\}} \sum_{i=1}^{s} \left\| \mathbf{x}^{(i)} - \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j^{(i)} \right\|_2^2 . \tag{5}$$

The convolution operation can be written as a matrix multiplication, where one of the operands is converted into a matrix, and we may write (5), as:

$$\min_{\{\mathbf{d}_i\}} \sum_{i=1}^{s} \left\| \mathbf{x}^{(i)} - \sum_{j=1}^{p} \mathrm{A}_j^{(i)} \mathbf{d}_j \right\|_2^2 . \tag{6}$$

In (6), $\mathrm{A}_j^{(i)}$ is the convolution matrix of $\boldsymbol{\alpha}_j^{(i)}$. Defining the matrix $\mathrm{A}^{(i)}$ as the horizontal concatenation of $\mathrm{A}_j^{(i)}$ for all $j$ and the vector $\mathbf{d}$ as a vertical concatenation of the local dictionary, the inner sum in (6) may be replaced by a single matrix multiplication and the minimization is over a single vector:

$$\min_{\mathbf{d}} \sum_{i=1}^{s} \left\| \mathbf{x}^{(i)} - \mathrm{A}^{(i)} \mathbf{d} \right\|_2^2 .$$

Next, we define the matrix X by stacking the training set $\left\{\mathbf{x}^{(i)}\right\}$ horizontally, and we define the matrix A by stacking the matrices $\mathrm{A}^{(i)}$ vertically. Then, the sum of squared $\ell_2$ norms can be rewritten as a single squared Frobenius norm:

$$\min_{\mathbf{d}} \ \|\mathrm{X} - \mathrm{A}\mathbf{d}\|_F^2 . \tag{7}$$

An analytical solution to (7) is achieved by taking the gradient and setting it to zero. This leads to the following solution:

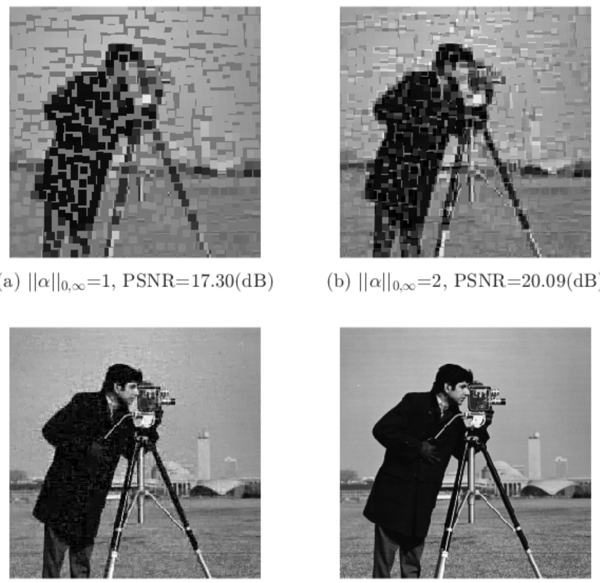
$$\mathbf{d} = \left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T\mathrm{X} = \left(\sum_{i=1}^{s} \mathrm{A}^{(i)T}\mathrm{A}^{(i)}\right)^{-1} \sum_{i=1}^{s} \mathrm{A}^{(i)T}\mathbf{x}^{(i)}.$$

Instead of using a direct matrix inversion, which is computationally prohibitive, $\mathbf{d}$ can be computed by a variety of numerical methods. For example, gradient descent methods only require computations of $\mathrm{A}^T\mathrm{A}\mathbf{d}$ and $\mathrm{A}^T\mathrm{X}$, both of which can be constructed as convolution operations. We use Conjugate Gradient Least Squares (CGLS) with these convolution operations.

After updating the dictionary, we go back to sparse coding with the updated dictionary using GCMP. Then, the next dictionary update occurs using the new A. Thus, a local dictionary is trained on a set of global images and optimized for solving the $P_{0,\infty}$ problem. This is a convolutional version of Method of Optimal Directions (MOD) [27].

**Table 1**. PSNR of inpainted images

| Method | Barbara | Boat | House | Lena | Peppers | Cameraman | Couple |
|---|---|---|---|---|---|---|---|
| Patch based K-SVD [26] | **13.33** | **11.05** | **10.61** | 11.96 | **10.78** | 10.06 | 12.07 |
| Heide et al. [9] | 11.00 | 10.29 | 10.18 | 11.77 | 9.41 | 9.74 | 11.99 |
| Papayan et al. [23] | 11.67 | 10.33 | 10.56 | 11.92 | 9.18 | 9.95 | **12.25** |
| GCMP (our method) | 11.94 | 10.53 | 10.51 | **11.99** | 9.93 | **10.24** | 11.61 |



(a) $\|\alpha\|_{0,\infty}$=1, PSNR=17.30(dB)

(b) $\|\alpha\|_{0,\infty}$=2, PSNR=20.09(dB)

(c) $\|\alpha\|_{0,\infty}$=8, PSNR=27.27(dB)

(d) original image

**Fig. 1**. Representations of the *cameraman* image for several values of $\|\boldsymbol{\alpha}\|_{0,\infty}$.

## 4. EXPERIMENTS

### 4.1. Accuracy vs. Sparsity

To evaluate the effect of the $\ell_{0,\infty}$ sparsity on the reconstruction of GCMP, we applied GCMP to the *cameraman* test image. We used the two-dimensional Haar wavelet dictionary of size $8 \times 8$ as the local dictionary.

Fig. 1 shows the original image and the reconstructed image for several values of $\|\boldsymbol{\alpha}\|_{0,\infty}$. For $\|\boldsymbol{\alpha}\|_{0,\infty} = 1$ there are no overlaps, and the selected atoms are mostly the DC atom and low-frequency atoms. As $\|\boldsymbol{\alpha}\|_{0,\infty}$ increases, the representation becomes more accurate.

### 4.2. Inpainting

Inpainting is the task of recovering an image from its corrupted version, which has missing pixels. We assume a convolutional sparse representation of the corrupted image: $\mathbf{y} = \mathrm{C}\mathbf{x} = \mathrm{CD}\boldsymbol{\alpha} = \mathrm{C}\sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j$, where $\mathbf{y}$ is the corrupted image, D is a convolutional dictionary, C, referred to as the subsampling matrix, is a diagonal binary matrix with $c_{ii} = 0$ for a corrupted pixel $i$, $c_{ii} = 1$ for an uncorrupted pixel $i$, and $c_{ij} = 0$ for $i \neq j$.

We find the sparse representation of the corrupted image by solving the $P_{0,\infty}^{\epsilon}$ problem with the global dictionary CD. The projections of the corrupted image onto CD are the same as onto D due to C being symmetric and idempotent: $\mathbf{b} = \mathrm{D}^T\mathrm{C}^T\mathrm{C}\mathbf{x} = \mathrm{D}^T\mathbf{y}$. The only effect C has on Algorithm 1 is when computing the residual, which becomes: $\mathbf{r} \leftarrow \mathbf{y} - \mathrm{C}\sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j$. After computing the sparse representation of $\mathbf{y}$, we estimate the original image as: $\hat{\mathbf{x}} = \mathrm{D}\boldsymbol{\alpha} = \sum_{j=1}^{p} \mathbf{d}_j * \boldsymbol{\alpha}_j$.

As in [9] and [23], we trained a dictionary of 100 atoms of size $11 \times 11$ on the *Fruit* training set, which contains ten images of fruit. We trained five dictionaries, each constrained to a different $\ell_{0,\infty}$ sparsity: 8, 16, 32, 64 and 128. We corrupted several test images by removing 50% of the pixels from each image at random. We inpainted the test images using GCMP[1], and averaged the reconstructions from the five dictionaries. Table 1 compares the PSNR of the inpainted images to the results reported in [23] (including their results for the method from [9]) and to the patch-based method from [26], which uses K-SVD to train on overlapping patches of the corrupted image. The patch-based method is much slower than the convolutional methods (the sparse coding step alone requires $O(MNmnpk)$ operations), and it is an online method (unlike the other methods we present here, which train an offline dictionary on the *Fruit* dataset). Online inpainting using our method appears in [25] and achieves better performance. We used $\mathrm{PSNR} = 20\log\left(\frac{\sqrt{MN}}{\|\mathbf{x}-\hat{\mathbf{x}}\|_2}\right)$ as defined in [23].

## 5. CONCLUSION

This work proposes a greedy strategy for the convolutional sparse coding problem based on the constrained $P_{0,\infty}^{\epsilon}$ problem. It offers an alternative to the approach in [22] and [23], which minimizes an unconstrained penalized Lagrangian with a convex relaxation to the $\ell_1$ norm. Shifting from the $\ell_0$ norm to $\ell_{0,\infty}$ makes the usage of greedy algorithms for convolutional sparse coding computationally feasible. We also propose a dictionary learning method based on $P_{0,\infty}^{\epsilon}$, which trains a local dictionary on global images. It is useful for image inpainting with offline pretrained dictionaries, giving results comparable to [23] and to patch-based methods.

[1] Code available at http://web.eng.tau.ac.il/~raja

## 6. REFERENCES


[1] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Transactions on signal processing*, vol. 41.12, pp. 3397–3415, 1993.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43.1, pp. 129–159, 2001.

[3] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," *ICCV*, pp. 479–486, 2011.

[4] M. Elad, "Sparse and redundant representations: From theory to applications in signal and image processing, chapter 14.3," *Springer*, pp. 278–297, 2010.

[5] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, May 2012.

[6] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng, "Shift-invariant sparse coding for audio classification," *Uncertainty in Artificial Intelligence*, 2007.

[7] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "De-convolutional networks," *CVPR*, pp. 2528–2535, 2010.

[8] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," *CVPR*, pp. 391–398, 2013.

[9] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," *CVPR*, pp. 5135–5143, 2015.

[10] H. Zhang and V. M. Patel, "Convolutional sparse coding based image decomposition," *BMVC*, 2016.

[11] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," *ICCV*, pp. 1823–1831, 2015.

[12] Y. Liu, X. Chen, R. K. Ward, , and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. doi:10.1109/lsp.2016.2618776, 2016.

[13] M. S. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," *NIPS*, vol. 11, pp. 730–736, 1999.

[14] A. Szlam, K. Kavukcuoglu, and Y. LeCun, "Convolutional matching pursuit and dictionary training," *arXiv abs/1010.0422*, 2010.

[15] B. Mailh, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," *EUSIPCO*, pp. 1–5, 2008.

[16] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "Shift-invariant sparse representation of images using learned dictionaries," *Proc. IEEE Workshop Mach. Learn. Signal Process. (MLSP)*, pp. 145–150, Oct. 2008.

[17] Q. Barthlemy, A. Larue, A. Mayoue, D. Mercier, and J. I. Mars, "Shift & 2d rotation invariant sparse coding for multivariate signals," *IEEE Trans. Signal Process.*, vol. 60 no. 4, pp. 1597–1611, 2012.

[18] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Signal Processing*, vol. 25(1), pp. 301–315, 2016.

[19] H. Bristow and S. Lucey, "Optimization methods for convolutional sparse coding," *arXiv abs/1406.2407*, 2014.

[20] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50(10), pp. 2231–2242, 2004.

[21] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell_1$ minimization," *Proceedings of the National Academy of Sciences*, vol. 100(5), pp. 2197–2202, 2003.

[22] V. Papyan, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Transactions on Signal Processing*, vol. 65.21, pp. 5687–5701, 2017.

[23] V. Papyan, J. Sulam Y. Romano, and M. Elad, "Convolutional dictionary learning via local processing," *ICCV*, 2017.

[24] B. Wohlberg, "Convolutional sparse coding with overlapping group norms," *arXiv abs/1708.09038*, 2017.

[25] E. Plaut and R. Giryes, "A greedy approach to convolutional sparse coding," *arXiv*, 2018.

[26] M. Elad, "Sparse and redundant representations: From theory to applications in signal and image processing, chapter 15," *Springer*, pp. 338–341, 2010.

[27] K. Engan, S. O. Aase, and J. H. Hakon-Husoy, "Method of optimal directions for frame design," *IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, pp. 2443–2446, 1999.