

A GREEDY ALGORITHM WITH LEARNED STATISTICS FOR SPARSE SIGNAL RECONSTRUCTION

Lucas Rencker, Wenwu Wang, Mark D. Plumbley

Center for Vision, Speech and Signal Processing, University of Surrey, UK
{l.rencker, w.wang, m.plumbley}@surrey.ac.uk

ABSTRACT

We address the problem of sparse signal reconstruction from a few noisy samples. Recently, a Covariance-Assisted Matching Pursuit (CAMP) algorithm has been proposed, improving the sparse coefficient update step of the classic Orthogonal Matching Pursuit (OMP) algorithm. CAMP allows the a-priori mean and covariance of the non-zero coefficients to be considered in the coefficient update step. In this paper, we analyze CAMP, which leads to a new interpretation of the update step as a maximum-a-posteriori (MAP) estimation of the non-zero coefficients at each step. We then propose to leverage this idea, by finding a MAP estimate of the sparse reconstruction problem, in a greedy OMP-like way. Our approach allows the statistical dependencies between sparse coefficients to be modelled, while keeping the practicality of OMP. Experiments show improved performance when reconstructing the signal from a few noisy samples.

Index Terms— Sparse representations, signal reconstruction, inpainting, Orthogonal Matching Pursuit

1. INTRODUCTION

Sparse representations have proved to be an efficient tool for the reconstruction of a signal from a few noisy measurements [1–5]. Many natural signals such as images and audio can be represented in an appropriate dictionary, where most coefficients are zero. The problem of sparse signal reconstruction is that of estimating the sparse coefficients from the available noisy samples. Finding a sparse decomposition is known to be an NP-hard problem [6], but a variety of algorithms have been proposed in the literature. Matching Pursuit (MP) [7] is a greedy algorithm that iteratively selects an atom from the dictionary, and updates its corresponding coefficient. Orthogonal Matching Pursuit (OMP) [8] improves the coefficient update step, by performing an orthogonal projection onto the support set, ensuring a steepest descent of the residual error.

The research leading to these results has received funding from the European Union’s H2020 Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet.

MDP is also partly supported by EPSRC grants EP/L027119/2 and EP/N014111/1.

WW is also partly supported by EPSRC grant EP/K014307/1.

One limitation of these algorithms is that they do not make use of the a priori distribution of the coefficients. “Spike-and-slabs” prior models have been introduced [9–11], but usually assume that the coefficients are independent and identically distributed, so they do not make use of the statistical dependencies between coefficients. Graphical models have been proposed to model these dependencies [12–14]. However this approach involves the design of specialised, and often computationally expensive optimization methods. Recently, a Covariance-Assisted Matching Pursuit (CAMP) has been proposed [15]. The author introduced the mean and covariance of the non-zero coefficients into the coefficient update step of OMP, which allows the statistical dependencies between coefficients to be considered. CAMP demonstrates significant improvement on image reconstruction tasks. However, the atom selection step does not take into account the new coefficient update.

CAMP was derived using the Gauss-Markov theorem [16], which does not need an explicit probabilistic model. In the present paper, we instead provide a Bayesian interpretation of the new coefficient update in [15]. The coefficient update can be interpreted as a maximum-a-posteriori (MAP) estimate of the non-zero coefficients, with a Gaussian prior. We then propose to leverage this interpretation, by solving a MAP estimation to the sparse reconstruction problem. In order to enforce sparsity of the solution, we solve the problem in a greedy OMP-like way. Our proposed algorithm shows significant improvement compared to OMP and CAMP, when reconstructing signals from only a few noisy measurements.

This paper is organized as follows: The problem of sparse signal reconstruction from noisy samples is described in Section 2. The OMP and CAMP algorithms are discussed in Section 3. We then analyze CAMP in Section 4, and describe our proposed algorithm in Section 5. The performance evaluation is presented in Section 6, before the conclusion is drawn.

2. PROBLEM FORMULATION

We consider the case of a signal $y \in \mathbb{R}^N$, corrupted by additive noise $v \in \mathbb{R}^N$, and where only a subset $I^r \subset \{1, \dots, N\}$ of

the N samples is observed, such that the received signal y^r is:

$$y^r = M^r(y + v), \quad (1)$$

where $M^r \in \mathbb{R}^{|I^r| \times N}$ is a *measurement* matrix defining the observed samples. We assume that the original signal y can be decomposed on an overcomplete dictionary $D \in \mathbb{R}^{N \times M}$ as $y = Dx$, with x sparse. The reconstruction problem can then be formulated as:

$$\begin{aligned} \hat{x} &= \underset{x}{\operatorname{argmin}} \|x\|_0, \text{ s.t. } \|M^r y - M^r D x\|^2 < \epsilon \\ &= \underset{x}{\operatorname{argmin}} \|x\|_0, \text{ s.t. } \|y^r - D^r x\|^2 < \epsilon, \end{aligned} \quad (2)$$

where $D^r \triangleq M^r D$ is the *reduced* dictionary, obtained by keeping only the relevant rows from the original dictionary. The constant ϵ is the approximation error, here set to be proportional to the noise level. The clean signal can then be reconstructed from the full dictionary D and the estimated coefficients \hat{x} as:

$$\hat{y} = D\hat{x}. \quad (3)$$

3. BASELINE APPROACHES

3.1. Orthogonal Matching Pursuit

The OMP algorithm [8] iteratively selects the most correlated atom with the residual vector at step k , and adds it to the support set Ω^k . It then performs an orthogonal projection of the signal onto the subspace generated by the sub-dictionary $D_{\Omega^k}^r$, defined as:

$$D_{\Omega^k}^r \triangleq D^r S_k^T, \quad (4)$$

where $S_k \in \mathbb{R}^{k \times M}$ is the *selection matrix*, that extracts the k relevant rows corresponding to the selected atoms. OMP for signal reconstruction is presented in Algorithm 1.

The atom selection step (5) is derived so that the algorithm performs the steepest descent of the residual error after each coefficient update (6) (see e.g. [17]). Moreover, as a result of the orthogonal projection (6) and the residual update (7), we have after each iteration:

$$\begin{aligned} D_{\Omega^k}^T r_k &= D_{\Omega^k}^T (y^r - D_{\Omega^k} x_k) \\ &= D_{\Omega^k}^T y^r - D_{\Omega^k}^T D_{\Omega^k} (D_{\Omega^k}^T D_{\Omega^k})^{-1} D_{\Omega^k}^T y^r \\ &= 0. \end{aligned} \quad (8)$$

This ensures that at the next iteration, during the selection step (5), $|\langle d_i^r, r_{k-1} \rangle|^2$ is zero for all $i \in \Omega^{k-1}$. In other words, if an atom has already been previously added to the support Ω^{k-1} , it will not be selected again. This property of the OMP makes it practical, as the number of selected atoms (i.e. the sparsity level) is equal to the number of iterations. Sparsity of the reconstructed signal is ensured by stopping the algorithm after a few iterations.

In the literature, the selection step (5) is often written as $\hat{i} = \operatorname{argmax}_i |\langle d_i^r, r_{k-1} \rangle|$, assuming that the dictionary atoms

Algorithm 1 OMP for signal reconstruction

Require: $y^r, D^r = \{d_i^r\}_{i=1..M}, \epsilon, K_{\max}$
initialize: residual $r_0 = y^r, x_0 = 0, \Omega^0 = \emptyset, k = 0$

while $\|r_k\|^2 > \epsilon$ and $\|x\|_0 < K_{\max}$ **do**
 $k \leftarrow k + 1$

Atom selection step:

$$\hat{i} = \operatorname{argmax}_i \frac{|\langle d_i^r, r_{k-1} \rangle|^2}{\|d_i^r\|^2} \quad (5)$$

Support set update: $\Omega^k = \Omega^{k-1} \cup \{\hat{i}\}$

Coefficient update step:

$$\begin{aligned} x_k &= \underset{u}{\operatorname{argmin}} \|y^r - D_{\Omega^k}^r u\|_2^2 \\ &= (D_{\Omega^k}^{rT} D_{\Omega^k}^r)^{-1} D_{\Omega^k}^{rT} y^r \end{aligned} \quad (6)$$

Residual update:

$$r_k = y^r - D_{\Omega^k}^r x_k \quad (7)$$

return $\hat{y} = D\hat{x}$

are normalized to unit ℓ_2 -norm. In this paper we keep the general formulation (5), so that the discussed algorithm can be used regardless of the measurement M^r and the norm of the atoms d^r . For convenience, we drop the index r in the remainder of the paper.

3.2. Covariance-Assisted Matching Pursuit

The author in [15] recently proposed a novel approach to improve the coefficient update step (6), using a-priori information on the mean and covariance of the non-zero sparse coefficients. The assumption is that the *non-zero* (“nz”) coefficients of the sparse vector $x \in \mathbb{R}^M$ are generated from a random distribution with mean $\mu^{\text{nz}} \in \mathbb{R}^M$ and covariance $\Lambda^{\text{nz}} \in \mathbb{R}^{M \times M}$. As a consequence, at the k -th iteration of the OMP algorithm, the coefficient vector $x_k \in \mathbb{R}^k$ corresponding to the k non-zero elements of x has a mean μ_k and covariance Λ_k , where $\mu_k \in \mathbb{R}^k$ and $\Lambda_k \in \mathbb{R}^{k \times k}$ are defined as:

$$\mu_k \triangleq S_k \mu^{\text{nz}} \quad (9)$$

$$\Lambda_k \triangleq S_k \Lambda^{\text{nz}} S_k^T, \quad (10)$$

In (7), the OMP algorithm usually approximates the coefficients x_k , assuming the following model:

$$y = D_{\Omega^k} x_k + v. \quad (11)$$

The model can be augmented as:

$$\tilde{y} = \tilde{D} x_k + \tilde{v}, \quad (12)$$

with $\tilde{y} = \begin{bmatrix} y \\ \mu_k \end{bmatrix}$, $\tilde{D} = \begin{bmatrix} D_{\Omega^k} \\ I \end{bmatrix}$ and $\tilde{v} = \begin{bmatrix} v \\ \mu_k - x_k \end{bmatrix}$. Let us assume that the noise v has zero mean and covariance

Σ . Using the Gauss-Markov theorem [16], the Best Linear Unbiased Estimator (BLUE) of x_k in (12), is the solution of the weighted least squares problem:

$$x_k = \underset{u}{\operatorname{argmin}} (\tilde{y} - \tilde{D}u)^T W (\tilde{y} - \tilde{D}u), \quad (13)$$

where $W = \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Lambda_k^{-1} \end{bmatrix}$. Solving (13) by setting the derivative with respect to u to zero, gives [15]:

$$x_k = (D_{\Omega^k}^T \Sigma^{-1} D_{\Omega^k} + \Lambda_k^{-1})^{-1} (D_{\Omega^k}^T \Sigma^{-1} y + \Lambda_k^{-1} \mu_k). \quad (14)$$

The atom selection and residual update are performed in a similar way as in OMP, so that the CAMP algorithm can be summarized by simply replacing the coefficient update step (6) by the new update (14) in Algorithm 1.

4. ANALYSIS OF CAMP

CAMP provides an easy way to incorporate the mean and covariance of the non-zero coefficients in the update step of OMP. The covariance reflects the statistical dependencies of the non-zero coefficients, but also the expected energy of each atom. Although no explicit probability distribution of the coefficients was mentioned in [15] (the covariance was introduced by invoking the Gauss-Markov theorem), the weighted least squares (13) can be interpreted as a Bayesian linear regression, with a Gaussian prior. To see this, (13) can be expanded as:

$$x_k = \underset{u}{\operatorname{argmin}} \left[(y - D_{\Omega^k} u)^T \Sigma^{-1} (y - D_{\Omega^k} u) + (\mu_k - u)^T \Lambda_k^{-1} (\mu_k - u) \right]. \quad (15)$$

We can recognize the maximum-a-posteriori (MAP) [16] estimate of x_k , assuming a Gaussian model $y \sim \mathcal{N}(D_{\Omega^k} x_k, \Sigma)$ (i.e. a Gaussian noise $v \sim \mathcal{N}(0, \Sigma)$), and a Gaussian prior on the non-zero coefficients $x_k \sim \mathcal{N}(\mu_k, \Lambda_k)$.

While the CAMP algorithm improves the coefficient update step of the OMP, it suffers from a few drawbacks:

- The coefficient update step has been modified, but the residual update and selection step are identical to those of the OMP. As a result, the derivations in (8) do not hold anymore. This means that, unlike for the OMP, there is no theoretical guarantee that the same atom cannot be selected twice. In [15], the atom selection is explicitly restrained to $i \notin \Omega^{k-1}$. However this also means that some residual energy might be lost at each iteration.
- The atom selection step does not take into account the new update step (13). This means that there is no guarantee that the selected atom will lead to a steepest descent of the residual error, as for the OMP.

In the next section we propose a new formulation of the sparse signal reconstruction problem, inspired by the model in (15), which allows to capture the statistical dependencies between atoms while keeping the practicality of OMP.

5. PROPOSED ALGORITHM

We propose to regularize the problem in (2), in a similar way as (15). Instead of minimizing $\|y - Dx\|^2$ iteratively, we propose instead to solve:

$$\hat{x} = \underset{x}{\operatorname{argmin}} [(y - Dx)^T \Sigma^{-1} (y - Dx) + (x - \mu)^T \Lambda^{-1} (x - \mu)], \quad (16)$$

where $\mu \in \mathbb{R}^M$ and $\Lambda \in \mathbb{R}^{M \times M}$ are the mean and covariance of the *full* sparse coefficient vector x . This can be seen as adding a multivariate Gaussian prior on the sparse coefficient $x \sim \mathcal{N}(\mu, \Lambda)$, and solving the MAP-estimation of the coefficients x .

The minimizer \hat{x} could be found in a closed form, but the solution would not be *sparse*. Instead, we propose to solve (16) in a greedy OMP-like way. For this, we notice that (16) can be easily reformulated as:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|\Sigma^{-1/2} (y - Dx)\|^2 + \|\Lambda^{-1/2} (x - \mu)\|^2 \quad (17)$$

$$= \underset{x}{\operatorname{argmin}} \left\| \begin{bmatrix} \Sigma^{-1/2} y \\ \Lambda^{-1/2} \mu \end{bmatrix} - \begin{bmatrix} \Sigma^{-1/2} D \\ \Lambda^{-1/2} \end{bmatrix} x \right\|^2, \quad (18)$$

where we introduce $\Sigma^{-1/2}$ and $\Lambda^{-1/2}$, the Choleski factorization of Σ^{-1} and Λ^{-1} , defined as $\Sigma^{-1} = (\Sigma^{-1/2})^T \Sigma^{-1/2}$ and $\Lambda^{-1} = (\Lambda^{-1/2})^T \Lambda^{-1/2}$. Finally, the regularized problem (16) can be formulated as:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|\tilde{y} - \tilde{D}x\|^2 \quad (19)$$

with $\tilde{y} \triangleq \begin{bmatrix} \Sigma^{-1/2} y \\ \Lambda^{-1/2} \mu \end{bmatrix}$, and $\tilde{D} \triangleq \begin{bmatrix} \Sigma^{-1/2} D \\ \Lambda^{-1/2} \end{bmatrix}$. We can now easily derive the OMP algorithm from the augmented model (19). The residual update is now:

$$\begin{aligned} \tilde{r}_k &= \tilde{y} - \tilde{D} S_k^T x_k \\ &= \begin{bmatrix} \Sigma^{-1/2} (y - D_{\Omega^k} x_k) \\ \Lambda^{-1/2} (\mu - S_k^T x_k) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma^{-1/2} r_1^k \\ \Lambda^{-1/2} r_2^k \end{bmatrix}, \end{aligned} \quad (20)$$

where $r_1^k \triangleq y - D_{\Omega^k} x_k$ and $r_2^k \triangleq \mu - S_k^T x_k$ are the residuals of the cost function and the regularization term respectively (one can recognize r_1^k as the residual of the classic OMP al-

gorithm). The new selection step is:

$$\begin{aligned}
\hat{i} &= \operatorname{argmax}_i \frac{|\langle \tilde{d}_i, \tilde{r}_{k-1} \rangle|^2}{\|\tilde{d}_i\|^2} \\
&= \operatorname{argmax}_i \frac{\left| \begin{bmatrix} \Sigma^{-1/2} D e_i \\ \Lambda^{-1/2} e_i \end{bmatrix}^T \cdot \begin{bmatrix} \Sigma^{-1/2} r_1^{k-1} \\ \Lambda^{-1/2} r_2^{k-1} \end{bmatrix} \right|^2}{\|\Sigma^{-1/2} D e_i\|^2 + \|\Lambda^{-1/2} e_i\|^2} \\
&= \operatorname{argmax}_i \frac{|e_i^T D^T \Sigma^{-1} r_1^{k-1} + e_i^T \Lambda^{-1} r_2^{k-1}|^2}{e_i^T D^T \Sigma^{-1} D e_i + e_i^T \Lambda^{-1} e_i}, \quad (21)
\end{aligned}$$

where $e_i \in \mathbb{R}^M$ is the column vector that is one at position i and zero elsewhere, such that $D e_i$ is the i -th atom from the dictionary. Once the atom is selected, the new coefficient update at iteration k is:

$$\begin{aligned}
x_k &= (S_k \tilde{D}^T \tilde{D} S_k^T)^{-1} S_k \tilde{D}^T \tilde{y} \\
&= (S_k (D^T \Sigma^{-1} D + \Lambda^{-1}) S_k^T)^{-1} S_k (D^T \Sigma^{-1} y + \Lambda^{-1} \mu) \\
&= (D_{\Omega^k}^T \Sigma^{-1} D_{\Omega^k} + S_k \Lambda^{-1} S_k^T)^{-1} (D_{\Omega^k}^T \Sigma^{-1} y + S_k \Lambda^{-1} \mu). \quad (22)
\end{aligned}$$

5.1. Discussions

As the proposed algorithm has been directly derived from the OMP algorithm, it inherits all the properties of the OMP. The proposed selection step (21), along with the new coefficient update (22) and residual update (20), unlike in [15], ensure that the same atom is not selected twice. The proposed method thus allows to incorporate the covariance and mean of the sparse coefficients, while keeping the practicality of OMP.

Note that the major difference between the CAMP and our algorithm, is that the CAMP uses statistics of the *non-zero* coefficients (μ^{nz} and Λ^{nz}), while our method uses the statistics of the full vectors (μ and Λ). Assuming a Gaussian distribution on the sparse coefficients might not reflect the actual distribution of the coefficients, but allows one to capture the statistical dependencies between atoms in a simple way.

Also note that the augmented dictionary \tilde{D} is of size $(N + M) \times M$, hence the new selection step is of computational cost $\mathcal{O}((N + M) \times M)$, compared to $\mathcal{O}(N \times M)$ for the OMP and CAMP. The proposed approach is thus computationally more expensive.

6. PERFORMANCE EVALUATION

To compare with the CAMP algorithm, we reproduced the inpainting experiment presented in [15], using the code and test data kindly provided by the author. The test data consists of 150,000 patches taken from 6 images (25,000 patches per image), each patch being of size $N = 8 \times 8$ pixels. In a similar fashion as in [15], the experiment was realized in an oracle way (the covariance matrix was learned from the

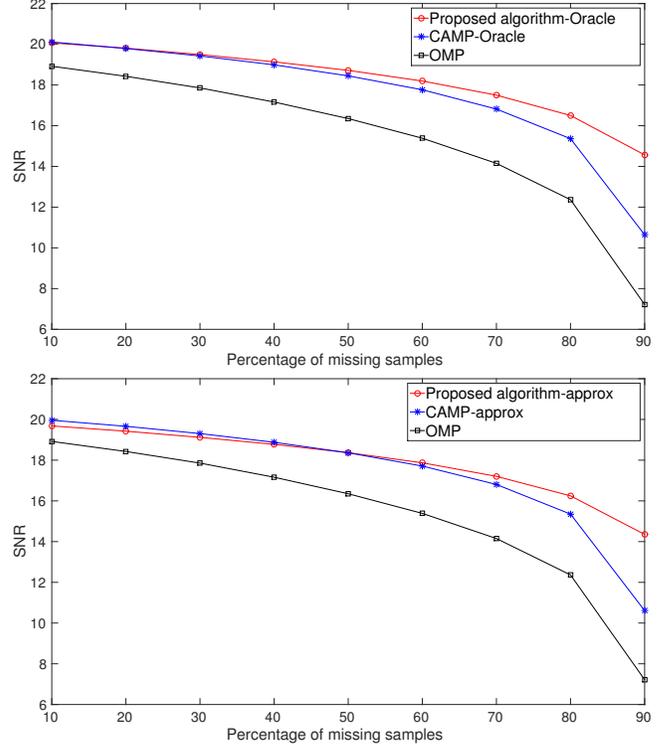


Fig. 1: Comparison of OMP, CAMP and proposed algorithm on image restoration from a few noisy samples. Top: Oracle reconstruction. Bottom: Approximate reconstruction.

same test images), and in an approximate way (the covariance was learned from another image, not in the test set). Each patch was decomposed onto an overcomplete DCT dictionary $D \in \mathbb{R}^{16 \times 256}$. The performance is shown in Figure 1. The experiment was realised by adding white noise ($\Sigma = \sigma^2 I$) with $\sigma = 30$, and randomly deleting samples, from 10% to 90%. The OMP, CAMP and the proposed algorithm were performed with a stopping criterion of $\epsilon = N \times 1.1 \times \sigma$, and a maximum number of atoms $K_{\max} = 32$. The proposed algorithm shows improved performance when more than 60% of the pixels are missing, with an improvement of more than 6 dB compared to the baseline OMP, and 4 dB compared to the CAMP when 90% of the pixels are missing. The increased performance can be explained by the data augmentation performed in (19): adding prior information results in an increased robustness when a lot of samples are missing.

7. CONCLUSION

In this paper, we have analyzed the recent CAMP algorithm, and presented a new greedy sparse reconstruction algorithm. The proposed algorithm provides an MAP solution for the sparse reconstruction problem, in a greedy OMP-like way. Experiments show improved performance over the classic OMP and the CAMP algorithms.

8. REFERENCES

- [1] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [2] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [3] P. Boufounos, M. F. Duarte, and R. G. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *2007 IEEE/SP 14th Workshop on Statistical Signal Processing (SSP 2007)*. IEEE, 2007, pp. 299–303.
- [4] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A reweighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.
- [5] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.
- [6] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.
- [7] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [8] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44 vol.1.
- [9] L. Chaari, J. Y. Tournet, and H. Batatia, "Sparse Bayesian regularization using Bernoulli-Laplacian priors," in *21st European Signal Processing Conference (EUSIPCO 2013)*, Sept 2013.
- [10] C. Herzet and A. Drémeau, "Bayesian pursuit algorithms," in *18th European Signal Processing Conference*. IEEE, 2010, pp. 1474–1478.
- [11] M. Korki, J. Zhangy, C. Zhang, and H. Zayyani, "An iterative Bayesian algorithm for block-sparse signal reconstruction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2174–2178.
- [12] T. Peleg, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2286–2303, May 2012.
- [13] A. Drémeau, C. Herzet, and L. Daudet, "Boltzmann machine and mean-field approximation for structured sparse decompositions," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3425–3438, 2012.
- [14] C. Févotte, B. Torrèsani, L. Daudet, and S. J. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 174–185, 2008.
- [15] A. Adler, "Covariance-assisted matching pursuit," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 149–153, Jan 2016.
- [16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer New York Inc., 2001.
- [17] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer Publishing Company, Incorporated, 2010.