# Multivariate Iterative Hard Thresholding for sparse decomposition with flexible sparsity patterns

Lucas Rencker, Wenwu Wang, Mark D. Plumbley
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, UK
Email: {l.rencker, w.wang, m.plumbley}@surrey.ac.uk

*Abstract*—We address the problem of decomposing several consecutive sparse signals, such as audio time frames or image patches. A typical approach is to process each signal sequentially and independently, with an arbitrary sparsity level fixed for each signal. Here, we propose to process several frames simultaneously, allowing for more flexible sparsity patterns to be considered. We propose a multivariate sparse coding approach, where sparsity is enforced *on average* across several frames. We propose a Multivariate Iterative Hard Thresholding to solve this problem. The usefulness of the proposed approach is demonstrated on audio coding and denoising tasks. Experiments show that the proposed approach leads to better results when the signal contains both transients and tonal components.

## I. INTRODUCTION

Sparse decomposition has proved to be an efficient tool for many signal processing applications such as coding, compression, denoising, inpainting and source separation [1]–[5]. Most natural signals can be efficiently represented as a linear combination of a few basis signals called "atoms", taken from an overcomplete dictionary. Finding a sparse decomposition of signal is known to be an NP-hard problem [6], but a range of algorithms have been proposed in the literature, such as Basis Pursuit [7], Matching Pursuit [8], Orthogonal Matching Pursuit [9] and Iterative Hard Thresholding (IHT) [10], [11].

Many applications require to compute the sparse decomposition of many successive signals, e.g. short-time frames in audio, or small 2D patches in image processing. Usually, each frame is processed independently and sequentially, with a sparsity constraint on each frame (see e.g. [3], [4], [12], [13]). One of the main difficulties in sparse decomposition problems is to find an appropriate sparsity level for the signal, i.e. an appropriate number of non-zero atoms [13]. In audio for example, the sparsity level depends on the nature, and the number of sources. Tonal sources present only a few frequency components, i.e. sparsity over frequencies. Transients or onsets, on the other hand, present a dense frequency content, very localized in time, thus sparsity over time [2]. One way to address the diversity of sparsity levels (i.e. very sparse and very dense frames) is to use hybrid representations: a union of two dictionaries can be used, one adapted to tonals and

one to transients [1], [2]. However, this results in very large dictionaries, and a much higher computational cost. Instead, we propose to decompose several frames simultaneously on the same dictionary, in order to allow us to enforce sparsity *across frames*, as well as sparsity of each frame.

Multiple Measurement Vectors (MMV) algorithms [14], [15], also called simultaneous sparse coding [16] or joint sparse coding algorithms [17], process several vectors at the same time, assuming that each vector is a measurement of the same signal. *Non-local* sparse models [18], or *collaborative filtering* approaches [19] also proposed to code several image patches simultaneously. However these approaches only make use of similar signals, assuming that they share the same support set. Matrix factorization methods, such as Dictionary learning or Non-negative Matrix Factorization (NMF) consider several frames simultaneously. However the sparse coding step in dictionary learning is usually computed frame by frame, assuming a fixed sparsity level for each frame [3], [12]. NMF algorithms process several non-negative vectors simultaneously using multiplicative update rules. This allows for more flexible sparsity patterns to be enforced, such as sparsity over the whole activation matrix (instead of sparsity on each column) [20], [21]. Moreover, this multiframe approach allows for structure *between* frames to be enforced, such as smoothness constraints [21].

In this paper, we propose to extend this idea to the multiple sparse signal decomposition problem. We propose a multivariate sparse decomposition framework, in order to decompose several frames simultaneously over the same dictionary. We then introduce a simple sparsity inducing matrix norm that counts the number of non-zero coefficients *over several frames*. We extend the classic IHT algorithm to the multivariate case, and introduce a masking matrix in order to respect the stopping criterion of each frame. The proposed algorithm processes several frames simultaneously, and chooses the best coefficients across several frames, instead of the best coefficients in each frame. This allows for more flexible sparsity patterns to be enforced, e.g. sparsity over frequencies (tonals) or sparsity over time (transients). We demonstrate how this framework is more adapted to the particular time-frequency structure of audio signals. Experimental results show that the proposed approach leads to better performance in terms of audio coding and denoising, especially for signals containing both tonals and transients.

The paper is organized as follows: in Section II we present the baseline approach, and the classic IHT algorithm. In Section III we propose a multivariate sparse coding framework, that allows to process several frames simultaneously. We later propose a simple sparsity-inducing matrix norm that extends the $\ell_0$ pseudo-norm to matrices, and a multivariate IHT algorithm to solve the multivariate sparse decomposition problem. Experiments are shown in Section IV, before the conclusion is drawn.

## II. BASELINE APPROACH

We consider a signal $y$, that can be decomposed as $T$ vectors $\{y_t\}_{t=1...T}$ each of size $N$, typically short-time frames in audio or 2D patches in image. In this paper we will simply refer to $\{y_t\}_{t=1...T}$ as *frames*. A typical sparse decomposition algorithm assumes that each frame $y_t \in \mathbb{R}^N$ can be sparsely represented in an overcomplete dictionary $D \in \mathbb{R}^{N \times M}$ ($N < M$). The sparse decomposition problem can be written as:

$$\hat{x}_t = \underset{x_t}{\operatorname{argmin}} \|y_t - Dx_t\|_2^2, \text{ s.t. } \|x_t\|_0 < K \quad \forall t = 1, ..., T \tag{1}$$

where $x_t \in \mathbb{R}^M$ is the sparse coefficient vector associated to frame $t$, $\|x_t\|_0$ is the pseudo-norm that counts number of non-zero elements in $x_t$, and $K$ is the *a-priori* sparsity level.

The IHT algorithm [10], [11] is a simple iterative algorithm that alternates between a gradient descent step, and a hard thresholding. The gradient descent ensures that the reconstructed signal fits the observation. The hard thresholding keeps the $K$ largest coefficients at each iteration, and thresholds the remainings to zero, thus ensuring sparsity of the solution. The algorithm stops once it has reached a maximum number of iterations, or once a desired reconstruction error is reached.

A typical approach to deal with multiple frames is to process them one by one, see e.g. [3], [4], [12], [13]. IHT for multiple vectors is presented in Algorithm 1, where $\mu$ is the gradient descent step size, and $\epsilon$ is the desired reconstruction error.

---

**Algorithm 1** IHT for multiple signals

**Require:** $\{y_t\}_{t=1...T}, D, \epsilon, K, \mu$
    **for** $t = 1, ..., T$ **do**
        **initialize:** $x_t^0 = 0$
        **while** $\|y_t - Dx_t^n\|_2^2 > \epsilon$ **do**
            $x_t^{n+1} = x_t^n + \mu D^T(y_t - Dx_t^n)$   ▷ Gradient descent
            $x_t^{n+1} \leftarrow \mathcal{H}_K(x_t^{n+1})$          ▷ Hard Thresholding
    **return** $\{x_t\}_{t=1..T}$

---

IHT is a simple, efficient algorithm with theoretical convergence guarantees for K-sparse signals [10], [11]. In practice however, the sparsity level might vary greatly between consecutive frames [13]. In audio, tonal signals exhibit localized energy around a few frequencies, which leads to an efficient sparse decomposition in a time-frequency dictionary such as a Gabor or DCT dictionary. However the sparsity level depends on the number of harmonics, or the number of sources. Transient components, on the other hand, have a dense frequency spectrum, very localized in time. In this case enforcing sparsity within a frame might not be an appropriate assumption. One should instead enforce sparsity over time, i.e. across consecutive time frames. Enforcing structure across frames requires a procedure where frames are processed simultaneously, instead of one by one.

In the next section we propose to reformulate the sparse decomposition problem (1), in order to process several frames simultaneously. We show how this allows us to enforce more flexible sparsity patterns, such as sparsity over time as well as sparsity over frequencies.

## III. MULTIVARIATE SPARSE CODING

Instead of solving (1) successively for each time frame $y_t$, we propose to reformulate the problem in order to process several frames simultaneously. We define $Y = [y_1, ..., y_T] \in \mathbb{R}^{N \times T}$ as the matrix containing $T$ adjacent frames concatenated altogether, and $X = [x_1, ..., x_T] \in \mathbb{R}^{M \times T}$ the corresponding sparse activation matrix. The multivariate sparse coding problem can then be formulated as:

$$\underset{X}{\operatorname{argmin}} \|Y - DX\|_F^2, \text{ s.t. } \Omega(X) < K_{\text{tot}}, \tag{2}$$

where $\|.\|_F$ is the Frobenius norm, $\Omega$ is a sparsity-inducing norm or pseudo-norm, and $K_{\text{tot}}$ is the overall sparsity level of $X$.

### A. Related works

The multivariate formulation in (2) has already been used in MMV algorithms [14], [15], simultaneous sparse coding [16] or joint sparse coding [17]. However, these algorithms assume that each column of $X$ shares the same support, i.e. that entire rows of $X$ should be set to zero. Some matrix-norms used are then non-zero rows counting norms, such as the $\ell_{0,\infty}$ pseudo-norm [15], [16], [18]:

$$\Omega(X) = \|X\|_{0,\infty} \triangleq \sum_{m=1}^{M} \|X^m\|_\infty^0, \tag{3}$$

with $X^m$ being the $m$-th row of $X$.

The multivariate sparse coding approach (2) also looks similar to the popular sparse NMF approach for *non-negative* data [20], [22]. In [20] for example the sparsity inducing norm was chosen as:

$$\Omega(X) = \sum_{m,t} |X_{m,t}|, \tag{4}$$

where $X$ is a *non-negative* activation matrix, thus enforcing sparsity on the whole activation matrix instead of sparsity on each column. Sparse NMF demonstrated high performance on many audio reconstruction tasks [21], [23]–[25], when working on non-negative data. Moreover, processing several frames simultaneously allows to enforce structure between frames, such as smoothness over time [21]. Here we propose to extend this idea to sparse data that are not necessarily non-negative.

## B. Proposed approach

To illustrate the usefulness of the multivariate approach (2), we propose a simple sparsity inducing matrix norm. We define $\|X\|_0 \triangleq \|\text{vec}(X)\|_0$, the simple "entrywise" matrix norm, extending the $\ell_0$ pseudo-norm to matrices. We propose to solve (2), with $\Omega(X) = \|X\|_0$:

$$\underset{X}{\operatorname{argmin}} \|Y - DX\|_F^2, \text{s.t.} \|X\|_0 < K_{\text{tot}}. \qquad (5)$$

This formulation thus enforces sparsity on the *whole* activation matrix $X$, instead of sparsity on each column (i.e. each frame). Here, we refer to the new sparsity prior $\|X\|_0 < K_{\text{tot}}$ as a *global* sparsity prior. This sparsity prior can also be seen as enforcing sparsity *on average* across several frames, instead of fixing an arbitrary sparsity level on each frame. Consequently, this approach allows for more "flexible" sparsity patterns, such as sparsity over time as well as sparsity over frequencies. It can also potentially efficiently represent very sparse frames, along with dense frames, and different sparsity patterns such as row sparsity and column sparsity.

We propose to solve (5) by reformulating the classic IHT algorithm into a multivariate way. The gradient descent step can easily be reformulated in its vectorized form as:

$$X^{n+1} = X^n + \mu D^T(Y - DX^n). \qquad (6)$$

The hard thresholding can then be performed by simply keeping the $K_{\text{tot}}$ highest elements of $X$ overall. Consequently, the algorithm chooses the best $K_{\text{tot}}$ coefficients across multiple frames, instead of enforcing $K$ elements in each frame. In other words, this means that frames that have a *dense* representation could "borrow" some coefficients from more sparse frames.

One shortcoming of a batch processing is that every frame is processed with the same number of iteration, and the stopping criterion $\|y_t - Dx_t^n\|_2^2 > \epsilon$ cannot be readily applied. This is of utmost importance in the denoising case, since the condition $\|y_t - Dx_t^n\|_2^2 > \epsilon$ reflects our knowledge about the noise level, and can lead to artifacts when it is not respected [26]. In order to deal with the stopping criterion for each frame $t$, we add a *masking* matrix $\mathcal{M} \in \{0,1\}^{M \times T}$ in the gradient descent step. $\mathcal{M}$ is initialized to the all-ones matrix $\mathbb{1}_{M \times T}$ and each column $t$ is set to zero once frame $t$ reaches the stopping criterion. This ensures that frame $t$ stops being updated once it reaches the stopping criterion.

The proposed algorithm is described in Algorithm 2, where we have defined $\mathcal{H}_{K_{\text{tot}}}^{\text{global}}$ as the hard thresholding operator that keeps the $K_{\text{tot}}$ largest components of $X$ overall, $X_t$ is the $t$-th column of $X$, and $\otimes$ is the Hadamard product. The proposed algorithm is thus a batch version of the classic IHT algorithm. Note that batch versions of other sparse coding algorithms have already been proposed to solve (1), such as Batch OMP [27] and Batch FISTA [28].

Figure 1 shows how the proposed approach performs on an audio coding task. We decomposed a glockenspiel signal, containing both tonals and transients. IHT provides a good reconstruction for tonals, but very poor reconstruction of

---

**Algorithm 2** Multivariate IHT with global sparsity

**Require:** $Y, D, \epsilon, K_{\text{tot}}, \mu, \mathcal{M} = \mathbb{1}_{M \times T}$
  **initialize:** $X^0 = 0, n = 0$
  **while** $\exists\, t$ s.t. $\|Y_t - DX_t^n\|_2^2 > \epsilon$ **do**
    $X^{n+1} = X^n + \mu \mathcal{M} \otimes \left(D^T(Y - DX^n)\right)$
    $X^{n+1} \leftarrow \mathcal{H}_{K_{\text{tot}}}^{\text{global}}(X^{n+1})$    ▷ Global Hard Thresholding
    **for** t = 1,...,T **do**       ▷ Check stopping criterion
      **if** $\|Y_t - DX_t^{n+1}\|_2^2 \leq \epsilon$ **then**
        $\mathcal{M}_t \leftarrow \mathbb{0}_M$          ▷ Update mask
  **return** $X$

---

transients. Multivariate IHT with global sparsity (i.e. sparsity on the overall activation matrix) achieves good reconstruction performance on both tonals and transients, with an overall improvement of more than 10 dB. Both algorithms were computed with the same number of non-zero coefficients overall, i.e. same compression performance. More experiments are presented in Section IV.
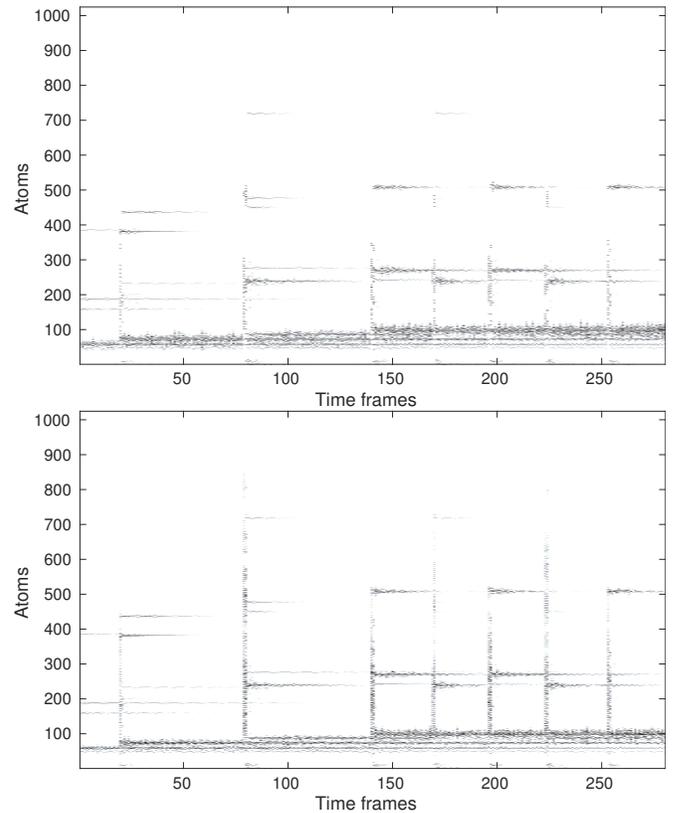


Fig. 1: Sparse activation matrix of $T = 280$ frames from a glockenspiel signal, containing both tonals and transients. Top: IHT (SNR = 21.7 dB). Bottom: Multivariate IHT with global sparsity (SNR = 31.8 dB). Both representations have exactly 8960 non-zero coefficients overall. The proposed approach manages to recover tonal components as well as transients.

## IV. EXPERIMENTS

In this section we present experiments on audio coding and denoising tasks. The test dataset consists of 10 music and 10 speech signals, sampled at 16 kHz. The music dataset contains both tonal signals, transients (e.g. drums), and mixtures of both. The speech dataset contains both male and female speech. Each signal was decomposed into $T = 624$ overlapping time frames of size $N = 256$, with 50% overlap. We used a twice-redundant DCT dictionary of $M = 512$ atoms. The gradient descent step size $\mu$ was fixed and chosen as $\mu = 1/\lambda_{\max}(D^T D)$, where $\lambda_{\max}(D^T D)$ is the highest singular value of $D^T D$. As suggested in [29], [30], we implemented each algorithm using *sparsity relaxation*: the sparsity level was initialized to 1, and increased periodically after a few iterations until the maximum sparsity level was reached. Experiments showed that 10 iterations were enough for the gradient descent step to converge, so we increased the sparsity level every 10 iterations.

### A. Audio coding experiment

We first show the result on an audio coding task. The coding performance is measured as the SNR of the reconstructed signal, as a function of the number of non-zero coefficients. The percentage of non-zero coefficients is simply computed as $100 \times \|X\|_0/(M \times T)$. We compared the IHT algorithm with the proposed multivariate IHT with global sparsity (Alg. 2). Both algorithms were computed so that each representation had exactly the same number of non-zero coefficients overall, and an accuracy of $\epsilon = 10^{-4}$. The results computed from 1% to 10% of non-zero coefficients are presented in Figure 2, for music and speech signals respectively. The proposed algorithm shows an overall improvement of 2 dB on music signals, and up to 5 dB of improvement on speech signals. This shows that the proposed approach leads to much better quality of reconstructed signals for the same number of coefficients, thus improving coding performance. Informal listening tests also revealed that the reconstructed audio was of much better perceptual quality, mostly because transient components and speech onsets were more efficiently coded.

### B. Denoising experiment

We evaluated the denoising performance of both algorithms, on signals corrupted by additive zero-mean Gaussian noise. The algorithms were computed in the same way as with the coding experiment, except that here we fix the stopping criterion $\epsilon$ as $\epsilon = k\sigma^2$, with $\sigma^2$ being the power of the noise and $k$ a constant set arbitrarily. $\epsilon$ is set proportional to the noise level in order to avoid noise components from being picked up during the reconstruction process (see e.g. [3], [18], [31]). The maximum sparsity level was chosen as $K = 32$ for the IHT (i.e. 32 atoms per frame), or $K_{\text{tot}} = 32 \times T$ for the global sparsity, thus enforcing the same number of non-zero coefficients on average. The results are shown is Figure 3. The proposed algorithm shows a slight improvement compared to the classic IHT, with an improvement of 0.2 dB overall, and up to 3 dB for speech signals in high SNR environments.
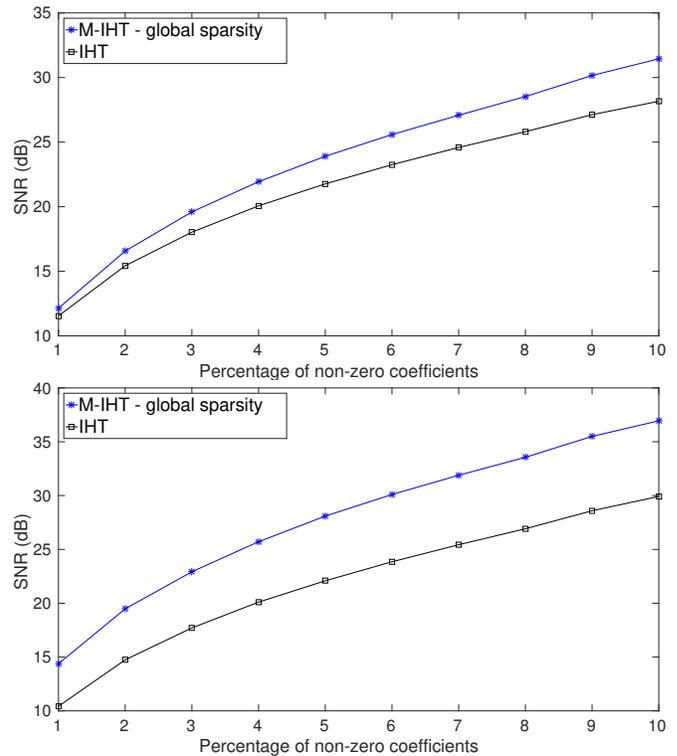


Fig. 2: Audio coding experiment. SNR of reconstructed signal, as a function of the percentage of non-zero coefficients. Top: music signals. Bottom: speech signals.

### C. Computational time

We compare empirically the computational time of IHT and the proposed multivariate IHT with global sparsity. The algorithms were implemented in Matlab, and the average CPU time to process one second of signal is reported in Table I.

|  | IHT | Multivariate IHT with global sparsity |
|---|---|---|
| CPU time (s) | 20.8 | 2.0 |

TABLE I: Average CPU time to process 1s of signal.

The vectorized gradient descent in Alg. 2 leads to a much faster implementation in Matlab, since it avoids an expensive iteration across fames.

## V. CONCLUSION

In this paper, we proposed a multivariate sparse decomposition framework that allows to process several frames simultaneously. We proposed a simple sparsity-inducing matrix norm, and a new Multivariate IHT algorithm. We proposed to extend the hard thresholding operator to matrices, in order to enforce sparsity on the whole activation matrix, instead of sparsity of each frame. This approach allows us to efficiently represent signals with very different sparsity patterns, such as tonals and transients. Experiments showed a significant improvement compared to the classic IHT in terms of audio coding, and moderate improvement in terms of audio denoising. Future
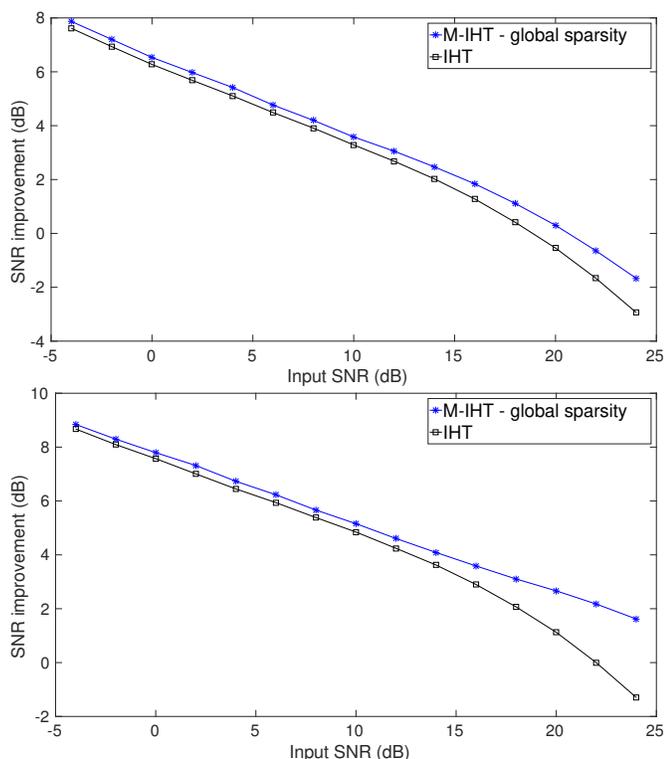
Fig. 3: Audio denoising experiment. SNR improvement of the reconstructed signal, as a function of the input SNR. Top: music signals. Bottom: speech signals.

work will evaluate to proposed approach to dictionary learning tasks, and try to extend the proposed framework to more complex sparsity-inducing matrix norm.

## REFERENCES

[1] E. Ravelli, G. Richard, and L. Daudet, "Union of MDCT bases for audio coding," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1361–1372, 2008.

[2] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies, "Sparse representations in audio and music: From coding to source separation," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 995–1005, 2010.

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

[4] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, 2012.

[5] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[6] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[8] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[9] Y. C. Pati, R. Rezaiifar, 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*, pp. 40–44.

[10] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[11] ——, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[12] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[13] M. G. Christensen, J. Østergaard, and S. H. Jensen, "On compressed sensing and its application to speech and audio signals," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 356–360.

[14] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[15] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[16] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[17] S. Sarvotham, D. Baron, M. Wakin, M. F. Duarte, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conference on Signals, Systems, and Computers*, 2005, pp. 1537–1541.

[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 2272–2279.

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[20] J. Eggert and E. Korner, "Sparse coding and NMF," in *2004 IEEE International Joint Conference on Neural Networks*, vol. 4, pp. 2529–2533.

[21] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

[22] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, no. Nov, pp. 1457–1469, 2004.

[23] M. N. Schmidt and R. K. Olsson, "Single-channel speech separation using sparse non-negative matrix factorization," in *ISCA International Conference on Spoken Language Proceesing (INTERSPEECH)*, 2006.

[24] M. N. Schmidt, J. Larsen, and F.-T. Hsiao, "Wind noise reduction using non-negative sparse coding," in *2007 IEEE Workshop on Machine Learning for Signal Processing*, pp. 431–436.

[25] H. Hu, N. Mohammadiha, J. Taghia, A. Leijon, M. E. Lutman, and S. Wang, "Sparsity level in a non-negative matrix factorization based speech strategy in cochlear implants," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 2432–2436.

[26] R. Hamon, L. Rencker, V. Emiya, W. Wang, and M. Plumbley, "Assessment of musical noise using localization of isolated peaks in time-frequency domain," in *42nd IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, 2017.

[27] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm and the Batch-OMP method," *Department of Computer Science, Technion, Israel, Tech. Rep*, 2008.

[28] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Learning overcomplete dictionaries based on atom-by-atom updating," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 883–891, 2014.

[29] S. Kitic, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. D. Vleeschouwer, "Consistent iterative hard thresholding for signal declipping," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 5939–5943.

[30] S. Kitic, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: a flexible non-convex approach," in *The 12th International Conference on Latent Variable Analysis and Signal Separation*, Liberec, Czech Republic, Aug. 2015, p. 8.

[31] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.