ADMM-Based Audio Reconstruction for Low-Cost-Sound-Monitoring

Sandra Ramaswami, Yohei Kawaguchi, Ryoichi Takashima, Takashi Endo, and Masahito Togami Research and Development Group, Hitachi, Ltd.

1-280, Higashi-koigakubo Kokubunji-shi, Tokyo 185-8601, Japan

Email: yohei.kawaguchi.xk@hitachi.com

Abstract—For low-cost sound monitoring of machineries, we propose a novel audio reconstruction method superior in terms of accuracy and processing time. A conventional method based on the Orthogonal Matching Pursuit (OMP) has been proposed for audio recovery. However, the conventional method has a low performance for sounds of machineries because, sounds of machineries tend to be not highly sparse, and the reconstruction performance of OMP decreases extremely if the signal is not sufficiently sparse. To solve the problem of the conventional method, the proposed method is based on the Alternating Direction Method of Multipliers (ADMM) for Group Lasso combined with the Gabor dictionary. While OMP's performance decreases with the number of nonzero elements, the proposed method shows a better robustness to variations in sparsity and outputs a reasonable result in a few tens of iterations. Those features among others make the algorithm a reliable solution which offers a better trade-off between accuracy and processing time compared to the conventional method.

Index Terms—compressive sensing, alternating direction method of multipliers, sub-Nyquist sampling, coprime sampling, orthogonal matching pursuit

I. INTRODUCTION

One of the challenging tasks in audio signal processing is recovering the original sound from incomplete data. In this paper, we focus on the reconstruction of audio signals processed through a sub-Nyquist sampling method - random or coprime sampling - adapted to low-cost sound monitoring, a technique required for machine maintenance. This procedure presents some useful features as it allows the user to reduce the labor and sensing costs - equipment, energy consumption, etc. - by only taking a precisely known number of measurements from the original sound. The fewer data we collect, the better the cost saving is. Because the reconstruction becomes arduous when the unknown information about the input signal is significant, we need a method that overcomes this difficulty while providing an exploitable outcome.

In the audio field, as well as in image and video processing, compressive sensing [1] has become a promising technique to achieve the reconstruction of audio signals from partial information. Compressive sensing solves an undetermined linear problem where the number of unknown variables exceeds the number of equations available. Numerous versions of this method have been proposed in the past few years. Based on the assumption that the input signal possesses a sparse representation, compressive sensing is effective in a specific domain where this sparsity is found. Examples of such domains include the fast Fourier Transform (FFT), the Discrete Cosine Transform (DCT) which is widely used in image processing [2] and the Gabor dictionary [3]. Since the FFT does not easily handle the phase shifts involved in compressive sensing [4], the DCT and Gabor dictionaries are usually preferred [5][6][3]. Adler et al. [3] demonstrated that the Gabor dictionary is better than DCT for audio reconstruction. Also, what reconstruction algorithm is fit for audio reconstruction is an important issue. The conventional state-of-the-art reconstruction algorithm by Adler et al. [3] is based on OMP [7], which is a popular greedy algorithm for l_0 -minimization. However, as mentioned in [8], under the condition that the signal is not sufficiently sparse, OMP fails to provide the sparsest solution. On such condition, the reconstruction performance of OMP decreases extremely. For instance, medical imaging and wireless communications provide a very sparse representation, whereas the sounds of machineries tend to be not highly sparse. Thus, the conventional method [3] is not suitable for the reconstruction of the sounds of machineries.

To solve the problem of the conventional method [3], we propose a reconstruction method based on the Alternating Direction Method of Multipliers (ADMM) with the Gabor dictionary. The proposed method solves l_1 -minimization, and so it does not suffer from the extreme deterioration even in the case that the signal is not highly sparse. By splitting the l_1 norm problem in two distinct parts, ADMM has the ability to provide an acceptable solution within a few tens of iterations [9] that still outperforms OMP's result for a given mechanical sound. Moreover, in order to make use of the Gabor dictionary, which has a good property for audio reconstruction, we apply the ADMM for Group Lasso algorithm [9] instead of regular l_1 -minimization. Well suited for convex optimization, the algorithm presented in this paper is both flexible and easy to implement. Through our experimental results, we show that ADMM offers a better compromise between accuracy and processing time than OMP in the reconstruction for a signal that is not highly sparse. Moreover, we offer a precise analysis supported by a representative experiment of the application, whose results covering the reasons why ADMM is an efficient method proposed for this purpose.

II. PROBLEM STATEMENT

In this section, we introduce the concept of compressive sensing with some notations from [10] that we will use throughout this paper.

We define a real-valued, discrete-time audio signal represented by a column vector x of dimensions $N \times 1$. Based on the assumption that the signal possesses a sparse representation when expressed in a specific transform domain, x can be written as

$$\boldsymbol{x} = \sum_{i=1}^{N} \boldsymbol{s}_i \boldsymbol{\Psi}_i = \boldsymbol{\Psi} \boldsymbol{s}, \tag{1}$$

where s is the $N \times 1$ column vector whose elements correspond to $s_i = \Psi_i^T x$ with .^T being the transposition, and $\Psi = [\Psi_1 | \Psi_2 | ... | \Psi_N]$ is a $N \times N$ transform basis matrix with Ψ_i as columns. Considering the case when $K \ll N$, the signal x is K-sparse in the Ψ domain if the (N - K)entries in s are zero. To reconstruct the signal of interest, compressive sensing directly acquires m measurements such that m < N. Thus, we consider a $m \times 1$ measurement column vector y whose elements are the result of the inner products between x and $(\Phi_j)_{j=1}^m$. Consequently, compressive sensing's undetermined linear problem can be expressed as

$$y = \Phi x = \Phi \Psi s = \Theta s, \tag{2}$$

where $\mathbf{\Phi}$, the measurement matrix, and $\mathbf{\Theta}$ are both $m \times N$ matrices.

The solution to (2) resides in the nature of Θ and depends on the signal reconstruction algorithm employed. While Ψ is chosen according to the signal of interest, Φ depends on the method of acquisition. The measurement matrix is related to the *a priori* known elements from the audio signal. In our case, the audio signal is processed through a sub-Nyquist sampling method. Therefore, the measurement matrix is defined by a specific selection for the user. For instance, by setting 5 and 7 as our coprime values, we construct Φ so that it chooses only the elements of x located in positions corresponding to all the multiples of 5 and 7. In this paper, the discussion is focused on the choice of the reconstruction algorithm which is the most adequate to retrieve sparse signals resulting from such sampling methods. Our study on the sampling methods is reported in another paper [11].

III. CONVENTIONAL METHOD

As mentioned previously, *Audio Inpainting* [3] proposes the Orthogonal Matching Pursuit (OMP) as a solution to retrieve the original signal from few measurements. With the transform domain Ψ defined as the Gabor dictionary, the version of OMP found in [3] presents the following features: (a) as a heuristic algorithm, OMP answers to the l_0 -norm problem expressed as

$$\hat{s} = \operatorname{argmin} ||s||_0$$
 such that $\Theta s = y$, (3)

(b) since s is a sparse vector, every iteration, OMP selects the columns of Θ that correlate the most with the residual, set to y in the initialization step. Eventually, the algorithm produces a set of K columns that correspond to the location of the elements in the sparse signal, (c) Adler *et al.* 's version is particularly adapted to the Gabor dictionary. The latter allows the algorithm to select pairs of columns per iteration that are,

in our case, pairs of cosine and sine at the same frequency with a zero phase, and (d) the algorithm stops as soon as the residual energy becomes less than the threshold set by the user or when the number of iterations k reaches K^{OMP} , value linked to the maximum number of nonzero elements.

Although popular in compressive sensing, OMP loses its attractive performance when faced with a signal that is not highly sparse, which is similar to those found by low-cost sound monitoring methods. In this context, the nonzero elements are clustered making them difficult to locate. Using a greedy approach, OMP is more likely to show errors considering that the algorithm does not correct its approximation of the sparse solution at each iteration. While research has been conducted in order to determine the exact number of nonzero elements [12], K is generally unknown, thus presenting a handicap for OMP which relies on it. The experiments conducted in [7] show us that the larger the number of nonzeros K, the lower the percentage of correct reconstruction. For N = 256 and K = 30 which correspond to 12% of the original signal, at least 75% of measurements (m > 196)are needed to achieve 70% chance of correct reconstruction. Since the usefulness of compressive sensing lies in situations where m is reasonably below N, the number of measurements taken should be considered carefully for a given K. Following the same example, for K = 12, roughly 5% of N, 52 measurements only give us 5% chance of correct reconstruction. One possible option would be to increase m. However, by increasing the number of measurements, i.e. the length of y, the processing time using OMP becomes longer. Indeed, the first step and last step of the algorithm highly depend on the residual involved in the matrix multiplications operated every iteration.

IV. PROPOSED METHOD

The lack of flexibility on the compromise between accuracy of reconstruction and processing time makes OMP unfit for our application related to low-cost sound monitoring audio signals. Therefore, to solve (2), another approach is required. Instead of finding the answer to (3), we have tested a minimum l_1 norm reconstruction process where the proposed method tries to solve the following problem

$$\hat{s} = \operatorname{argmin} ||s||_1$$
 such that $\Theta s = y$. (4)

This convex optimization problem can be solved via numerous algorithms [13][14]. One particular example of such is *LASSO* [15] which goal is to directly find the sparse solution s. In this case, (4) can be rewritten as

minimize
$$\frac{1}{2} || \boldsymbol{\Theta} \boldsymbol{s} - \boldsymbol{y} ||_2^2 + \lambda || \boldsymbol{s} ||_1,$$
 (5)

where $\lambda > 0$ is a regularization parameter.

A general approach to convex optimization problems is the Alternating Direction Method of Multipliers (ADMM) [9]. As a powerful and flexible algorithm, it is applicable for l_1 -norm problems unlike usual methods such as *subgradient* or *Newton*'s. One of the advantages of ADMM is the separation

Algorithm 1 ADMM for Group Lasso *adapted to* Gabor dictionary

Input: observation vector \boldsymbol{u} and sensing matrix $\boldsymbol{\Theta}$ **Output**: approximation for signal, \hat{s} 1: Initialization: $s^0 = 0$, $z^0 = 0$, $u^0 = 0$, k = 1, $\lambda_0 = 0$, $j = 0, j_{end} = K_g$ 2: Initialize: $\lambda > 0$, $\rho > 0$, $k_{end} \ge 2$ 3: Compute $inv = (\boldsymbol{\Theta}^T \boldsymbol{\Theta} + \rho I)^{-1}$ and $prod = (\boldsymbol{\Theta}^T \boldsymbol{y})$ 4: for $k < k_{end}$ do $\boldsymbol{s}^{k} = inv(prod + \rho(\boldsymbol{z}^{k-1} - \boldsymbol{u}^{k-1}))$ 5: for $j < j_{end}$ do 6: $egin{aligned} & \mathbf{z}_j^k = R_{c_j\lambda/
ho}(\mathbf{s}_j^k + \mathbf{u}_j^{k-1}) \ & j = j+1 \end{aligned}$ 7: 8: end for 9: $\boldsymbol{u}^k = \boldsymbol{u}^{k-1} + \boldsymbol{s}^k - \boldsymbol{z}^k$ 10: k = k + 111: 12: end for 13: $\hat{s} = s^k$

of the solution s in two distinct parts s and z making parallel computing possible and simple. Since we are seeking a parsimonious model of the signal, a version of ADMM for *LASSO* [9] solves (4) expressed as

minimize
$$f(s) + g(z)$$
 subject to $s - z = 0$
with $f(s) = \frac{1}{2} ||\Theta s - y||_2^2$ and $g(z) = \lambda ||z||_1$, (6)

where ADMM handles the sparsity of the solution by processing z through a soft thresholding function every iteration.

In the frame of our application, we wish to apply the Gabor dictionary as our transform domain for the audio signal. Considering its properties, it is necessary to adapt the current form of the problem. By applying ADMM for *Group Lasso* [9] (cf. Algorithm 1) in our specific case, we are able to deal with the pairs of cosine and sine present in the sparse solution s thanks to a block soft thresholding function. Although, those groups are joined to a certain degree, they can be handled separately which makes the strength of this algorithm. Consequently, the problem can be rewritten as

minimize
$$f(s) + g(z)$$
 subject to $s - z = 0$
with $f(s) = \frac{1}{2} ||\Theta s - y||_2^2$ and $g(z) = \lambda \sum_{j=1}^{K_g} c_j ||z_j||_2$, (7)

where $K_g = \frac{N}{2}$ is the size of the Gabor dictionary, j is the index of frequency components, z_j is the two-element vector extracted from z corresponding to the j-th pair of cosine and sine, and c_j is the weight parameter for the j-th frequency component. In this study, c_j is set to $c_j = 1$, i.e., all frequencies are treated equally.

The proposed is solution presents several features to simplify its implementation and gain in processing speed. First, the sparsity of the output s is handled by the block soft thresholding function (cf. step 7 of Algorithm 1) corresponding to

$$R_L(\boldsymbol{a}) = \left(1 - \frac{L}{||\boldsymbol{a}||_2}\right)_+ \boldsymbol{a},\tag{8}$$

where the threshold value $\frac{\lambda}{\rho}$ is manually set once. Second, it is important to note that with $\rho > 0$, *inv* is always invertible. Finally, the computation of *inv* and *prod* before the start of the loop and the possibility to set manually k_{end} allow us to save considerably on the processing time.

ADMM focuses on minimizing s, then z and updating both of the vectors in a final step. Therefore, the algorithm refines the entire output s progressively throughout the iterations. As stated in [9], in practice, ADMM has a slow convergence rate when the expectation is a highly accurate result. However, in the framework of our application, such accuracy is not necessary, and one of the requirements is to provide an exploitable outcome allowing the user to spot the anomalies in the reconstructed signal in order to carry out an efficient maintenance of the machine. Unlike other convex optimization methods, ADMM is capable of both handling l_1 norm problems and present an acceptable result in a few tens of iterations.

When comparing OMP and ADMM regarding their method and output, for highly sparse signals, OMP is more likely to provide a better result than ADMM since the nonzero elements in the solution are well spaced thus reducing the chance of error when finding their indexes. However, for low sparse signals found in low-cost sound monitoring audio signals, ADMM appears to be a more promising approach as the different steps of the algorithm do not directly depend on the sparsity of the input, leading to a result that is acceptable and acquired in a reasonable amount of time. Besides, unlike OMP, ADMM is only slightly affected by the number of measurements, i.e. the length of y.

V. EXPERIMENTAL RESULTS

The aim of the following experiments is to show that with the proper reconstruction algorithm adapted to the sampling process used on the input signal, we can, rapidly and accurately, reconstruct an audio signal from few data, allowing the user to reduce the equipment expenses while providing a efficient product maintenance when needed. In order to better illustrate our application and evaluate the performance of the proposed method compared to the conventional one, we have conducted several tests on an audio signal generated by a printer whose spectrogram is shown below (cf. Fig.1). It is a 3.125 seconds 16-bit audio signal sampled at 16 kHz.

We apply the hamming window with a frame size set to 1024 and a frame shift of 128. Thus, for a full input of 50000 points, we repeat the reconstruction process 390 times. We decide to take our measurements via a coprime sampling technique - a method studied in the recent years for signal processing with promising results [16][17]. Initially, we set the coprime values to 5 and 7 in order to obtain 322 measurements - corresponding to roughly 30% of data from the original



Fig. 1. Spectrogram of the printer's audio signal.

signal for each frame since $1024 \times (\frac{1}{5} + \frac{1}{7} - \frac{1}{35}) = 322$. Therefore, the measurement matrix Φ has for dimensions 322×1024 . Besides, each part of the frame is represented in the Gabor dictionary Ψ whose dimensions are 1024×2048 . The reconstruction algorithm solves (2) with m = 322, N = 2048 and Θ of dimensions 322×2048 with the following parameters: $\rho = 1$ and $\lambda = b \times \max |\Theta^T y|$ [18] where the coefficient b = 0.0006.

The results that we will first display and comment in this paper are done accordingly to the above settings and will demonstrate two points from the three ones mentioned in this paper: the compromise between processing time and accuracy and the influence of the number of measurements taken. We determine the accuracy of reconstruction with the *signal-todistortion ratio* (SDR) which is calculated as follows

$$SDR_{dB} = 10 \log_{10} \left(\frac{P_{\text{original signal}}}{P_{\text{(original signal - reconstructed signal)}}} \right)$$
(9)

where P is the power of the signal. As for the processing time, it is mainly controlled by the number of iterations. Since we run the algorithms 390 times, in all the figures, we display the average processing time and the average accuracy of reconstruction. In the case of ADMM, we can neglect the computation time of *inv* (cf. Algorithm 1) as it can be calculated independently of the frames. Since we decided to set a frame shift of 128, we compute in a prior process 8 different *inv* which are repetitively used successively in the step 5 of Algorithm 1. Consequently, all the processing time results for ADMM do not include *inv*.

According to Fig.2, we can observe that from only a few iterations ADMM outperforms OMP in terms of accuracy for roughly the same processing time. ADMM's performance remains stable regardless of the number of iterations while OMP's accuracy declines progressively after hitting its optimum. This study validates our expectation where ADMM offers a better compromise between processing time and accuracy in the reconstruction of an audio signal.

As suggested previously, to improve OMP's accuracy, one option would be to increase the number of measurements



Fig. 2. Performance of OMP and ADMM regarding the processing time and the accuracy of reconstruction (SDR). The parameters changed are k_{end} and K^{OMP} . The optimum is reached when $k_{end} = 11$ for ADMM and $K^{OMP} = 30$ for OMP.



Fig. 3. Influence of the number of measurements m - quantity of *a priori* known data about the original signal - on the performance of OMP and ADMM in terms of processing time and accuracy of reconstruction for $N = 2048, K = 60, k_{end} = 11$ and $K^{OMP} = 30$

taken. For this experiment, we modify the coprime sampling settings in order to obtain 227 (8&9), 256 (7&8), 322 (5&7), 409 (4&5), 512 (3&4) and 683 (2&3) measurements. Fig.3 displays the result. Following the theoretical expectation, OMP is longer to achieve the signal reconstruction when the number of measurements becomes higher, unlike ADMM which stays at the speed of 0.54sec. When considering the accuracy of reconstruction, the latter logically increases with the number of measurements. In this regard, ADMM proves to be more effective than OMP as well. This observation supports ADMM's reliability where the quantity of *a priori* known information about the original signal does not harm the algorithm's efficiency.

The third and last point to highlight in this paper is the performance of those algorithms regarding the number of nonzero elements of the signal to reconstruct. A notable



Fig. 4. Influence of the number of nonzeros on the performance of OMP and ADMM in terms of accuracy of reconstruction for N = 100, m = 50, $k_{end} = 11$ and K^{OMP} ranging from 1 to 25 (Graphs for $K \in [2; 50]$ and [16; 50]).

remark is the challenge of correctly reconstructing a signal which is not exactly sparse since the problem solving for (2) increases in difficulty with the number of nonzeros. In order to determine the effect of this parameter, we change it in a simulation program where the reconstruction algorithm tries to reconstruct a random sparse signal s of N = 100, in the Gabor dictionary, from m = 50 measurements in y. We set $\rho = 1, b = 0.035$ and $k_{end} = 11$. We run the program 1000 times to generate different inputs to reconstruct. As shown by Fig.4, we test the number of nonzeros K from 2 to 50. For each one of them, we decide to plot the 25th, 50th (equivalent to the median) and 75th percentile. When the signal is highly sparse (K = 2 and K = 6), OMP reacts perfectly by guaranteeing an accurate result. Even though we work with $m = \frac{N}{2}$, the reconstruction performance inevitably decreases when \overline{K} increases. For K = 10, we can clearly observe that OMP can be an unstable algorithm since its accuracy of reconstruction ranges between extreme values. ADMM remains stable and robust to the increase of K. Moreover, for signals with a large number of nonzero elements, ADMM rapidly offers a better accuracy than OMP. This confirms that a greedy approach is less preferable in the case of sounds of machineries.

VI. CONCLUSION

With a proper sub-Nyquist sampling method, we propose a reconstruction method based on the ADMM for Group Lasso adapted for the Gabor dictionary as an accurate and rapid reconstruction algorithm for audio signals resulting from such processes. This proposed solution outperforms the conventional one based on OMP. By explaining and commenting the algorithm, we concluded on the reliability and flexibility of ADMM that solves the undetermined linear problem of compressive sensing while being robust to the quantity of known information about the original signal and the number of nonzero elements. Through our experiments, we confirmed that for a given audio signals processed through low-cost sound monitoring, the proposed method shows the best performance in terms of accuracy, stability and processing time.

REFERENCES

- D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] S.A. Khayam, "The discrete cosine transform (DCT): Theory and application," Department of Electrical and Computer Engineering Michigan State University, Mar. 2003.
- [3] A. Adler, V. Emiya, M.G. Jafari, M. Elad, R. Gribonval, and M.D. Plumbley, "Audio inpaiting," *IEEE Transactions on Audio, Speech,* and Language Processing, vol. 20, no. 3, Mar. 2012.
- [4] R.G. Moreno-Alvarado and M. Martinez-Garcia, "DCT-compressive sampling of frequency-sparse audio signals," in *Proc. The World Congress on Engineering*, July 2011, vol. II.
- [5] R.R. Karhe, P.B. Shinde, and J.N. Fasale, "Audio compression using DCT and CS," *International Journal of Engineering Research and Technology (IJERT)*, vol. 4, no. 1, pp. 263–266, Jan. 2015.
- [6] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [7] 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, Dec. 2007.
- [8] D.L. Donoho and Y. Tsaig, "Fast solution of l₁-norm minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
- [9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [10] R.G. Baraniuk, "Compressive sensing," IEEE Signal Processing Magazine, pp. 118–120, July 2007.
- [11] Y. Kawaguchi, S. Ramaswami, R. Takashima, T. Endo, and R. Ikeshita, "Sub-Nyquist non-uniform sampling for low-cost sound monitoring," in Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017, submitted.
- [12] V. Bioglio, T. Bianchi, and E. Magli, "On the fly estimation of the sparsity degree in compressed sensing using sparse sensing matrices," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 3801–3805.
- [13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [14] 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.
- [15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical*, vol. 58, no. 1, pp. 267–288, 1996.
- [16] H.K. Achanta, S. Biswas, S. Dasgupta, M. Jacob, B.N. Dasgupta, and R. Mudumbai, "Coprime conditions for fourier sampling for sparse recovery," in *Proc. The 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 533–536.
- [17] Q. Wu and Q. Liang, "Coprime sampling for nonstationary signal in radar signal processing," *EURASIP Journal on Wireless Communications* and Networking, vol. 2013, no. 1, pp. 58, Mar. 2013.
- [18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "MATLAB scripts for alternating direction method of multipliers," [Online] http: //stanford.edu/~boyd/papers/admm/.