Speech Enhancement using Greedy Dictionary Learning and Sparse Recovery

K. N. H. Srinivas *

*Corresponding Author, Research Scholar, ECE Department, JNTUK, Kakinada, India. E-mail: knhsrinivas@yahoo.com

I. Santhi Prabha

Professor, ECE Department, JNTUK, Kakinada, India. E-mail: santiprabha@yahoo.com

M. Venugopala Rao

Professor, ECE Department, K. L. University, Guntur, India. E-mail: mvgr03@kluniversity.in

Abstract

Most real-time speech signals are frequently disrupted by noise such as traffic, babbling, and background noises, among other things. The goal of speech denoising is to extract the clean speech signal from as many distorted components as possible. For speech denoising, many researchers worked on sparse representation and dictionary learning algorithms. These algorithms, however, have many disadvantages, including being overcomplete, computationally expensive, and susceptible to orthogonality restrictions, as well as a lack of arithmetic precision due to the usage of double-precision. We propose a greedy technique for dictionary learning with sparse representation to overcome these concerns. In this technique, the input signal's singular value decomposition is used to exploit orthogonality, and here the $\ell 1$-$\ell 2$ norm is employed to obtain sparsity to learn the dictionary. It improves dictionary learning by overcoming the orthogonality constraint, the three-sigma rule-based number of iterations, and the overcomplete nature. And this technique has resulted in improved performance as well as reduced computing complexity. With a bit-precision of Q7 fixed-point arithmetic, this approach is also used in resource-constrained embedded systems, and the performance is considerably better than other algorithms. The greedy approach outperforms the other two in terms of SNR, Short-Time Objective Intelligibility, and computing time.

Keywords: Sparse Representation, Greedy Dictionary Learning, Singular Value Decomposition, Orthogonal Matching Pursuit, Quantization.

Introduction

The explosive amounts of massive data became the norm of the modern-day in many engineering disciplines and scientific areas. And this became a daunting challenge for computation as well as learning. To rise to this challenge, Sparse Dictionary Learning (SDL) framework provides a potential for representation learning by exploiting the consecration of dimensionality. Real data tends to lie in or near some low-dimensional subspaces or manifolds, even though the ambient dimension is often extremely large (Wang et al., 2020). The data matrix is given as 'Y' where $Y \in R^{nxp}$ containing 'n' by 'p' samples and the aim is to find the dictionary $D \in R^{nxm}$ and its associated sparse representation 'W' where $W \in R^{mxp}$, and the corresponding equation satisfying is given in equation (1) as

$$Y=D*W \qquad\qquad (1)$$

The matrix data 'Y' represents a wide variety of signals like audio, images, genetics, etc., applications (Zhai et al., 2020), and the SDL is applied in neuroscience, image processing, audio or speech processing, etc.

Sparse representation or sparse coding is representing a signal as a linear combination of elements/atoms a few, from the dictionary matrix containing overcomplete atoms. The sparse coefficient vector contains the useful atoms obtained from the dictionary to linearly combined with the original signal. The value of each access of the coefficient matrix is read as a weight that defines the linear combination fraction of each chosen atom (Tang et al., 2021). The two-stage sparse algorithm is proposed by (Aharon et al., 2006), the sparse coding step is said to be the first stage and is based on the pursuit method. The second stage is a dictionary update step in which at an instant each column updation of the dictionary is done. Later, methods related to dictionary learning were online-based adaptive learning algorithms. The Greedy Adaptive Dictionary (GAD) transform takes a similar approach, except that the atoms with the highest ℓ2-norm are removed first, and thus the algorithm produces signal approximations strongly as the number of atoms is reduced (Jafari & Plumbley, 2011). The sparse recovery is proposed by (Rubinstein et al., 2008) and it uses the orthogonal matching pursuit method, later an adaptive recovery algorithm is proposed (Beheshti et al., 2018). An offline and online-based K-means Singular Value Decomposition (KSVD) dictionary learning and the recovery algorithm for speech denoising using fixed-point arithmetic is presented in (Srinivas et al., 2020) using Orthogonal Matching Pursuit (OMP). The dictionary algorithms are analyzed (Srinivas et al., 2019), and in the variants of Singular Value Decomposition (SVD) like KSVD, Label Consistency-KSVD (LC-KSVD), and Rotate-KSVD (R-SVD) it is found that the best learning method is KSVD for metrics like Signal to Noise Ratio (SNR).

The ℓ4-norm trait has long been recognized and employed in the search for (orthonormal) functions having similar properties. The 'highest weight' function maximizes the ℓ4-norm locally. It is the 'most concentrated' in measure of all the Laplacian eigenfunctions on the sphere. "Matching, Stretching, and Projection" also called ℓ4-norm is used for dictionary learning (Yuexiang Zhai et al., 2020; Zhai et al., 2020).

The main contribution of the work is speech enhancement using the ℓ1-ℓ2 norm approach, resulting in a sparsity augmented dictionary. Further, it can improve signal reconstruction significantly. It also reduced the computational time to learn the dictionary with sparse-based techniques. It is extended to fixed-point arithmetic and with this fixed-point Online dictionary learning, the memory utilization is reduced and also the load on the processor due to its non-overcomplete nature. We compared the existing techniques with the proposed using the metrics like the signal-to-noise ratio, peak signal-to-noise ratio, Short-Time Objective Intelligibility, and Improve Signal-to-noise ratio and found that the proposed algorithm gave better results.

The organization of the paper is as follows in Section 2, the speech denoising methods are introduced, and the greedy dictionary learning algorithm is elaborated in Section 3. In Section 4, experiments are conducted and presented with results and in Section 5 the conclusions are made.

Literature Review

The potential steps involved in Speech denoising are dictionary learning and sparse recovery. When a signal is transformable then such signal is sparse and it is possible to learn a dictionary, a wavelet or a cosine transform for example can be chosen as such source. An initial dictionary is chosen in a random or transform domain and the K-SVD algorithm (Rubinstein et al., 2008) is used to train this initial dictionary. The learned dictionary is used in speech enhancement. 'D0' is the initial dictionary and learned to form an overcomplete dictionary in 'k' iterations. The dictionary 'D' and the sparse representation 'Γ' are simultaneously learned using KSVD and OMP algorithms, using the sensed input signal 'X'. By ℓ2-norm minimization of (X-D*Γ) as revealed in equation (2), for every i'th iteration the dictionary 'Di' is learned along with the sparse representation matrix 'Γi'.

$$\text{MIN}_{D_i,S} \left|\left| X - D_i\Gamma \right|\right|^2 \text{ Subject To } \forall i \left|\right| \Gamma_i \left|\right|_0 \leq k \tag{2}$$

The column by column is considered to train the dictionary by using the KSVD algorithm. The signal 'X' columns are represented given the initial or the current dictionary sparsely to evaluate 'Γ', the sparse representation matrix, and then the dictionary atoms are constructed given the generated sparse matrix. At each step, the cost function is optimized atom by atom while maintaining the sparse representation for each atom. The column vector in the sparse matrix representation 'Γ' will have elements based on the sparsity 'K' requirement. The dictionary is built column by column after the sparse matrix is obtained. The indices 'I' represent the non-zero items received from the sparse matrix and are utilized to find the error 'E'. The indices for 'I' are derived from the input signal 'X', which employs the jth atom, and the updating is accomplished by maximizing the cost function as given in equation (3)

$$\left|\left| X_I - D\Gamma_I \right|\right|_F^2 \tag{3}$$

In 'ΓΙ' over both the atom and its associated row coefficient. The problem resulting is a simple approximation task given in equation (4).

$$\{d; g\} := \underset{d; g}{ARGMIN} \left|\left| E - dg \right|\right|_F^2 \quad \text{Subject To } ||d||_2 \leq 1 \tag{4}$$

where $\left|\left| X_I - \sum_{i \neq j} d_I \ \Gamma_{i,I} \right|\right|$ is error matrix excluding the jth atom, the atom of 'd' is updated, and gT is the updated coefficient of 'ΓΙ'. A Singular Value Decomposition is used to address the problem.

Greedy adaptive dictionary (GAD) learning (Jafari & Plumbley, 2011) is based on orthogonal transform. The sparsity index of a vector is defined as the ratio of ℓ1-norm and ℓ2-norm of the vector or signal 'X'. The smaller the value of this index the vector is sparser. The dictionary is updated by extracting the atoms from the signal or vector 'X'. At each iteration, the residual matrix changes concerning the vector corresponding to the pth column of Rp. Initially, the residue is initialized to 'X'. The dictionary is constructed by selecting the residual vector r_k^p with the lowest sparsity index (Bai et al., 2019; Chi et al., 2019; Geng & Wright, 2014; Huang et al., 2020; Ron Rubinstein, Alfred M. Bruckstein, 2010). Analytical and numerical examples introduced sequentially in convex algorithms illustrate that the ratio of ℓ1/ℓ2 penalty when computed produces both stable and sparse solutions (Repetti et al., 2015; Yin et al., 2014). The residual column of Rp with the lowest ℓ1 to ℓ2 norm ratio is given as equation (5)

$$k^p = \arg \min_{k \notin I^2} \left\{ \left\| r_k^p \right\|_1 / \left\| r_k^p \right\|_2 \right\} \tag{5}$$

And the p-th atom is normalized $\hat{r}_{k^p}^p$ as given in equation (6)

$$\psi^p = \hat{r}_{k^p}^p / \left\| \hat{r}_{k^p}^p \right\|_2 \tag{6}$$

And the new residual is calculated from equation (7) and once it is calculated the 'pth' column is made dummy to reduce further processing.

$$r_k^{p+1} = r_k^p - \psi^p \langle \psi^p, r_k^p \rangle \text{ for all columns k} \tag{7}$$

Where, $\hat{x}^p(t)$ is the speech signal approximation given as x(t) obtained at the pth iteration from equation (8)

$$\hat{\hat{x}}^p = D^p * (D^p)^T * X \tag{8}$$

The framing process is reversed and D^p is the learned dictionary iteratively.

Zhai et al. (Yuexiang Zhai et al., 2020) proposed a global strategy in the complete dictionary learning setting, which presents a formulation that can efficiently recover the sparse signal matrix 'X' once and for all known as Matching, Stretching, and Projection (MSP or ℓ4 -norm). An orthogonal dictionary is efficiently learned solving via optimization using ℓ4 -norm given in equation (9) over the orthogonal group O(n; ℝ) entirely.

$$\max_{D} \|D^*X\|_4^4, \quad \text{subject to } D \in O(n; \mathbb{R}) \tag{9}$$

where the sum of 4th powers are the $\ell 4$ -norm of a matrix means: $\forall A \in \mathbb{R}^{n \times m}, \| A \|_4^4 = \sum_{i,j} a_{i,j}^4$. Equation (10) shows the:

$$\max_{\Gamma', D} \| \Gamma \|_4^4, \quad \text{subject to } X = D\Gamma, D \in O(n; \mathbb{R}) \tag{10}$$

where take full advantage of the $\ell 4$-norm of $'\Gamma'$ promotes (Zhai et al., 2020) "sparsity" or "spikiness" (over a sphere). The sparsest points can be seen on the unit $\ell 2$-sphere. Equation (11) and equation (12) indicate the 4th norm of the dictionary and the random measurement ∂A_t.

$$\partial A_t \doteq 4(A_t D_o)^{\circ 3} D_o^* \tag{11}$$

$$\nabla_A \|A D_o\|_4^4 = 4(A D_o)^{\circ 3} D_o^* \tag{12}$$

The singular value decomposition of the ∂A_t is obtained to get the

$$U \Sigma V^* = SVD(\partial A_t) \tag{13}$$

Project A_{t+1} onto orthogonal group

$$A_{t+1} = UV^* \tag{14}$$

There are two critical issues involved in the implementation of speech enhancement using an embedded system and they are time and resources. One such resource is memory i.e., ROM. So, an Offline quantized learned dictionary is introduced to reduce memory utilization and it also reduces the computational load on the processor or microcontroller & time consumption. The quantized offline dictionary learning supports the fixed-point arithmetic as given in equation (15). As most embedded systems like hearing aids use fixed-point arithmetic so, a fixed-point offline KSVD dictionary learning is implemented (Srinivas et al., 2020).

Arithmetic designs with lower bit depth decrease the memory usage and the computation are also faster and are known that some controllers don't support floating-point arithmetic. The decomposition of the matrix is done by altering the double-precision floating-point variables (x) to the integer variables (x^{Qb}) with 'Qb' format in which 'b' is the target bit-precision shown in equation (15) and the matrix is updated until the error reduces.

$$x^{Qb} = \frac{(\text{INT}(x \cdot \text{Qmax}))}{L} \tag{15}$$

Where, $\text{Qmax} = 2b - 1$, $L = 2b$ and $b = $ Target bit precision. This is used for quantization.

The Batch-Orthogonal Matching Pursuit (OMP) (Elvira et al., 2021; Rubinstein et al., 2008) algorithm is intended to represent a large number of sensed signals sparsely. The progressive Cholesky is used to minimize the amount of work required when inverting a

matrix. In each step, the greedy OMP algorithm chooses an atom that is highly associated with the current residue.

To compute the residue Once a strongly correlated atom has been chosen, the residue is computed by projecting orthogonally on that atom. The greedy step is performed on kth column as $\bar{k} := \text{Arg Max } k\{|\alpha_k|\}$, where $\alpha_k = D^T X$ with D as the dictionary and X as the input signal the sparsity of the signal and the orthogonalization are performed using the Cholesky factorization (Rubinstein et al., 2008; Zhang et al., 2020). The equation (16) of Cholesky factorization i.e., $A^\sim = L^\sim L^{\sim T} \in R\,(n-1) \times (n-1)$, is given as

$$A = \begin{pmatrix} A^\sim & v \\ u & c \end{pmatrix} \text{ , with}$$

$$L = \begin{pmatrix} L^\sim & 0 \\ w^T & \sqrt{c - w^T w} \end{pmatrix}, \text{ where } \quad w = L^{\sim-1}v \tag{16}$$

The other part of the OMP involves using fixed-point arithmetic in the Cholesky decomposition. The sparse representation calculations using the Cholesky factorization, instead of double-precision as needed by (Aharon et al., 2006) and (Rubinstein et al., 2008) are implemented in fixed-point arithmetic (Srinivas et al., 2020).

In this paper, an "Online Greedy Sparse" Dictionary Learning algorithm is proposed which is fast and effective. It takes less computational time and resources when compared to the sparse KSVD dictionary learning algorithm. Here dictionary learning is applied using the double-precision floating-point and also using fixed-point arithmetic. The fixed-point arithmetic is based on quantization introduced into the algorithms i.e., in both learning and recovery are proposed in a way that, can be potentially used in real-time processing applications. With simply integer Arithmetic and Logical Units, fixed-point arithmetic can be achieved in hardware (ALUs). And Section 3 describes the proposed greedy algorithm.

Online SVD Greedy Dictionary Learning

The invertibility of a matrix is convenient if the matrix is orthogonal. If the matrices are orthogonal (AAT = 1), computing the inverse is as simple as using the matrix transpose. Under the Bernoulli-Gaussian assumption, the problem of learning an arbitrary complete dictionary can be reduced to that of learning an orthogonal dictionary. So, Singular Value Decomposition (SVD) is used here to obtain orthogonal matrices. So an online SVD Greedy dictionary learning algorithm is proposed in which the SVD is performed before the learning process.

$$U\Sigma V^* = SVD(X) \tag{17}$$

The residue is taken as the product of the most significant value of Σ singular values and the left and right orthonormal eigenvectors as shown in equation (18).

$$R = U * \Sigma(1) * V^* \tag{18}$$

The input noisy speech signal is formed as frames and it is decomposed into eigenvectors by using SVD. As the Singular Value Decomposition gives orthonormal columns i.e., left and right orthonormal eigenvectors, which is used as Residue. This Residue is used to formulate the Dictionary. This decomposed signal has now the basis vector orthogonal as shown in the Algorithm.

Algorithm -- SVD Greedy Algorithm
1.   Input: Signal - X, Initial dictionary-D.
2.   Output: Dictionary-D.
3.   Initialize: Set D: = [ ]
4.   Find the SVD of X.
5.   From the Eigenvectors obtain the Residue - R.
6.   Find the size of columns (col) of the R.
7.   for length 1 to col repeat
8.   Using the lowest $\ell 1$ - $\ell 2$ norm the residual column of R^p:
9. $k^p = \arg \min\limits_{k \notin I^2} \left\{ \|r_k^p\|_1 - \|r_k^p\|_2 \right\}$
10.   Set the pth atom as the normalized $\hat{r}_{k^p}^p$
11. $\psi^p = \hat{r}_{k^p}^p / \|\hat{r}_{k^p}^p\|_2$
12.   Add to the dictionary:
13. $D^p = [D^{p-1} \mid \psi^p], \mathcal{I}^p = \mathcal{I}^{p-1} \cup \{k^p\}$
14.   Compute the new residual
15. $r_k^{p+1} = r_k^p - \psi^p \langle \psi^p, r_k^p \rangle$ for all columns k
16. $r_k^p = [\ ]$
17.   End

The dictionary is learned from each column of the residue by the difference of $\ell 1$-norm and $\ell 2$-norm as in equation (19). The $\ell 1$- $\ell 2$-norm produces a sparse solution of a signal. residual column of Rp with lowest $\ell 1$ - $\ell 2$ norm:

$$k^p = \arg \min_{k \notin I^2} \left\{ \|r_k^p\|_1 - \|r_k^p\|_2 \right\} \tag{19}$$

The residue 'R' is obtained in each iteration at the 'p^{th}' column from

$$\bar{R} = D^p * (D^p)^T * R \tag{20}$$

where, D^p is the learned dictionary from equation 13 of the algorithm. The new residue is obtained by subtracting the prior residue with the $\bar{R}$ using equation (20) and thereby the residue is minimized for each column. The residue is updated and the 'pth' atom once learned is removed from the columns during further processing to reduce the computational complexity. The remaining columns are considered in calculating the $\ell 1$- $\ell 2$-norm and the minimum value column is taken to update the dictionary.

Results and Discussion

The KSVD-based dictionary learning results are compared to that of our algorithm. Table.1 shows the results with respective the signal to noise ratio (SNR) and other metrics in which the input noise is street noise of 0dB and the clean speech signal remains "The birch canoe slid on the smooth planks" called as "sp01" with the sampling rate of the signals 8KHz. Here two methods are compared concerning both floating-point and fixed-point forms. The noisy speech corpus (NOIZEUS) was used for evaluating the algorithms (Hu & Loizou, 2007). The enhanced speech signal was obtained by using SVD greedy dictionary algorithm (SVD-GDL) and compared to the KSVD algorithm and GAD algorithm for a street noise of 0dB. The OMP algorithm is used for sparse recovery. The SVD greedy algorithm with OMP recovery gave a better SNR (Signal to Noise Ratio).

The computational complexity of the KSVD algorithm is running the iterations for K columns and with SVD having $O(NK^2max)$, so in total having $O(NK^3max)$. In the GAD algorithm, the computational complexity is over the signal length 'n', based on the difference between $\ell 1$ to $\ell 2$ norm and by 'K' maximum columns, so the computation is given by $O(n^2Kmax)$. In the MSP algorithm, with SVD having $O(NK^2max)$ computational complexity and is iterated for 'p' times then the overall computational complexity will be $O(pNK^2max)$. Now, for the proposed greedy algorithm, the computational complexity is with one-time SVD and with 'N' columns iterated for 'Kmax' terms giving rise to $O(NK^2max) + O(N^2Kmax)$.

Table 1 . Metrics Comparison for Sp01 Signal - 0dB Noise

Sp04 Signal - street Noise – 0dB				
Method	**SNR (dB)**	**PSNR (dB)**	**MSE (10^{-6})**	**MAE (10^{-3})**
Online Floating-Point GAD	1.67	19.21	1366	28.69
Online Floating-Point Sparse KSVD	4.53	24.41	412	12.15
Online L4 – MSP Dictionary Learning	5.45	25.30	336	10.90
Online SVD Greedy Dictionary Learning (SVD-GDL)	**5.60**	**25.50**	**323**	**10.70**

Table 2 . Metrics Comparison for Sp01 Signal - 0dB Noise – Fixed-Point

Sp04 Signal - street Noise – 5dB				
Method	**SNR (dB)**	**PSNR (dB)**	**MSE (10^{-6})**	**MAE (10^{-3})**
Online Greedy Adaptive Dictionary Learning – with *8-bit* fixed-point	1.64	19.13	1326	28.69
Online Sparse KSVD Learned Dictionary – with *8-bit* fixed-point	4.51	24.34	419	12.23
Online L4 – MSP Dictionary Learning – with *8-bit* fixed-point	5.40	25.28	337	10.90
Online SVD Greedy Dictionary Learning (SVD-GDL) – with *8-bit* fixed-point	**5.58**	**25.40**	**327**	**10.73**

Table 3 . Comparison of Computational Time

Method	Time Taken
Online Floating-Point GAD	60 ms
Online Floating-Point Sparse KSVD	8 sec
L4 – MSP Dictionary Learning	100 ms
Online SVD Greedy Dictionary Learning (SVD-GDL)	**30 ms**

Table 4 . Comparison Speech Quality & Intelligibility

Method	STOI	PESQ	ISNR
Online Floating-Point GAD	0.401	1.04	1.000
Online Floating-Point Sparse KSVD	0.581	1.19	5.197
L4 – MSP Dictionary Learning	0.586	1.42	6.128
Online SVD Greedy Dictionary Learning (SVD-GDL)	**0.589**	**1.44**	**6.262**
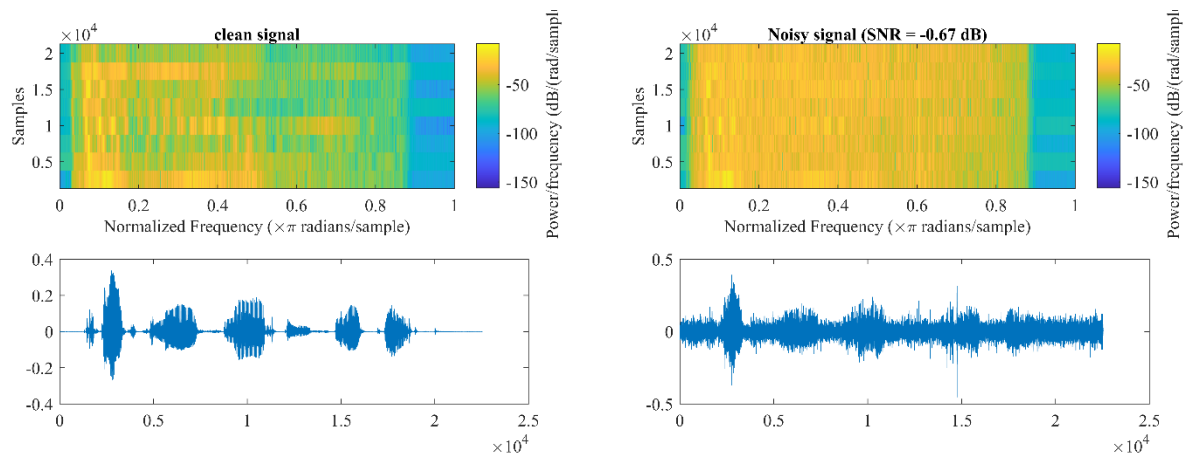
Fig.1. a. Speech Clean Signal (left), b. Speech Noisy Signal (right)
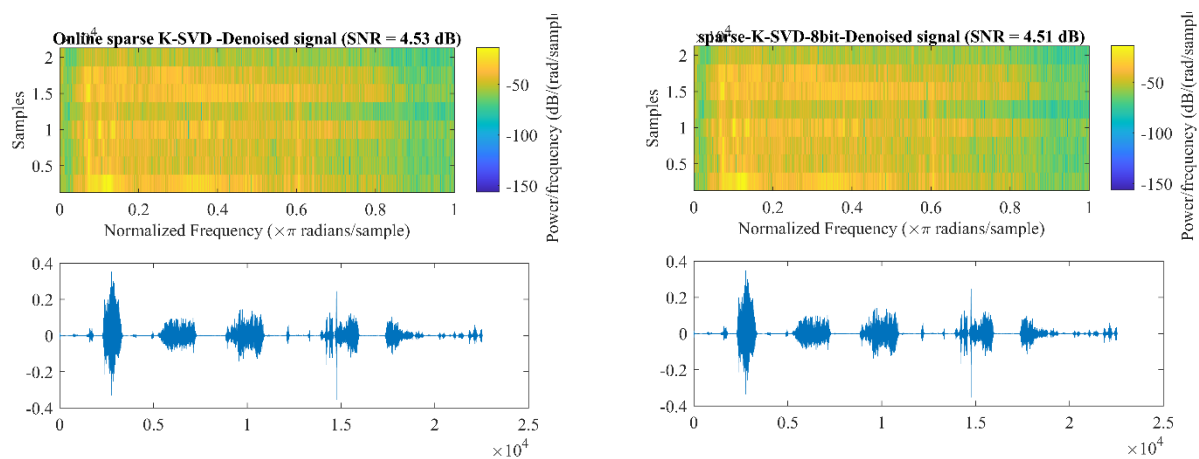

Fig. 2. Denoised using sparse KSVD Speech Signals – a. Online & Floating-point (left), b. Online Fixed-point with 8-bit (right)
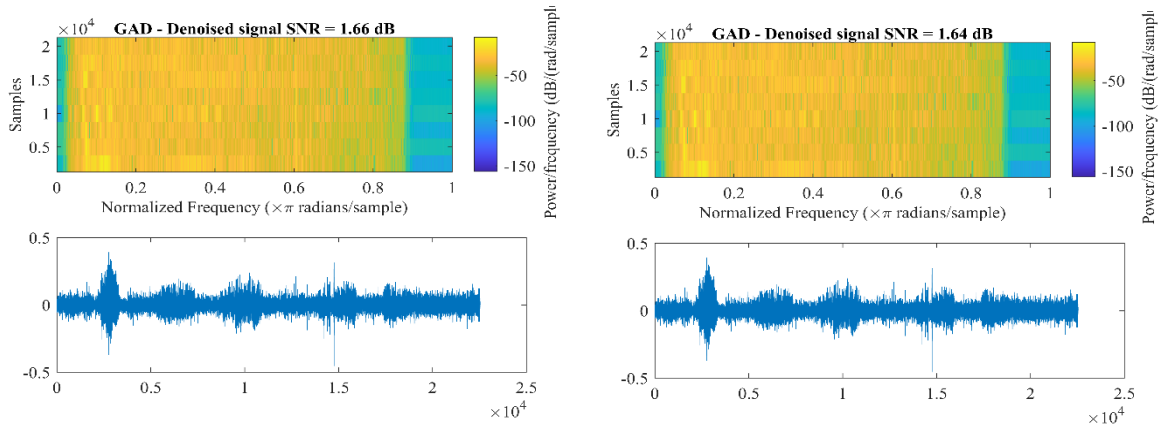
Fig. 3. Denoised using GAD Speech Signals – a. Fixed-point 8-bit (left), b. Online & Floating-point (right)
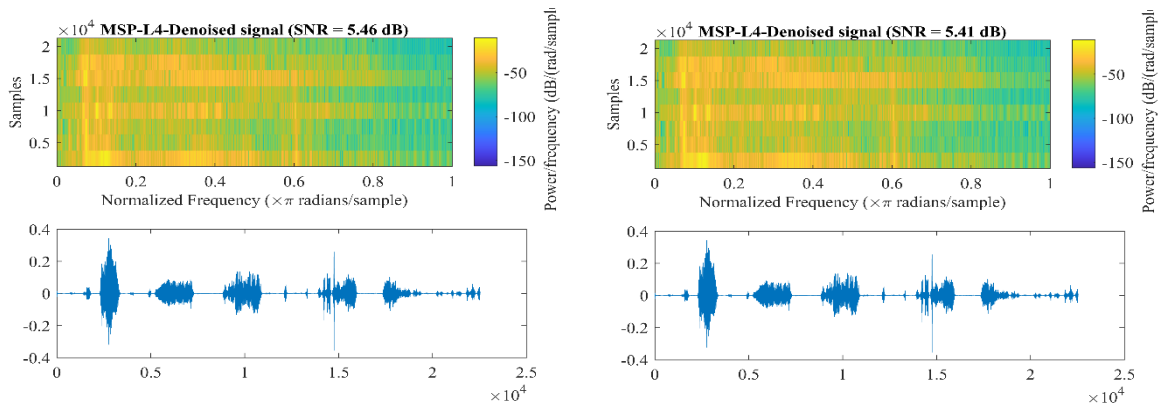

Fig. 4. Denoised using MSP or L4 Speech Signals – a. Fixed-point 8-bit (left), b. Online & Floating-point (right)
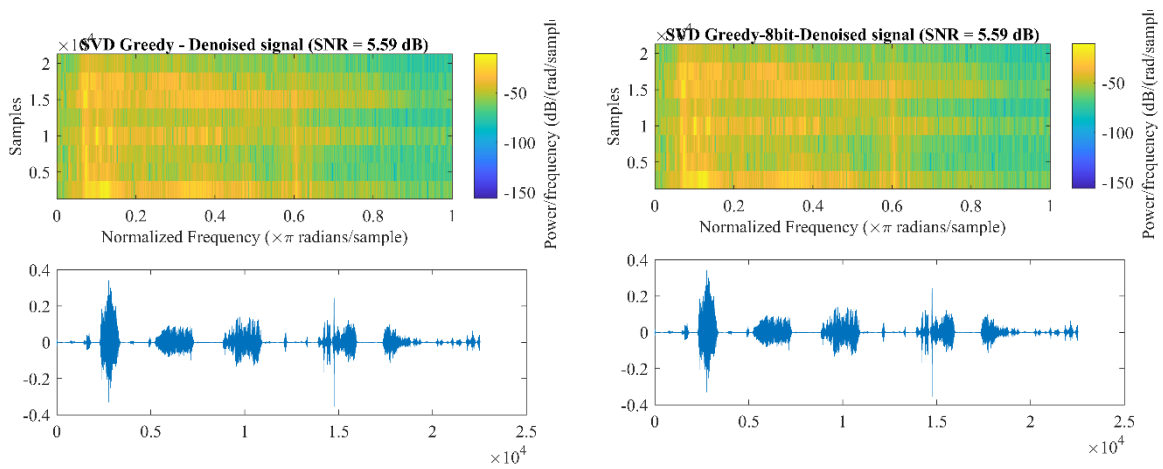

Fig. 5 –Denoised using SVD-Greedy Speech Signals – a. Fixed-point 8-bit (left), b. Online & Floating-point (right)

Fig. 1 (a) is a clean speech signal and Fig. 1 (b) is a noisy speech signal with a street noise of 0dB at a sampling frequency of 8KHz. Fig. 2 (a) is denoised speech signals obtained after using a sparse KSVD-based Online dictionary learning algorithm and Fig. 2 (b) is denoised speech signals obtained after using a sparse KSVD-based 8-bit fixed-point dictionary learning algorithm. Fig. 3 (a) is denoised speech signals obtained after using a GAD-based Online dictionary learning algorithm and Fig. 3 (b) is denoised speech signals obtained after using an online GAD 8-bit fixed-point dictionary learning algorithm. Fig. 4 (a) is obtained after processing using 8-bit fixed-point MSP -based dictionary learning and Fig. 4 (b) is obtained after processing using MSP-based dictionary learning. Fig. 5 is proposed method i.e., (a) is obtained after processing using 8-bit fixed-point SVD-Greedy-based dictionary learning and Fig. 5 (b) is obtained after processing using SVD-GDL-based dictionary learning.

In the resource-constrained embedded system also, the greedy algorithm when compared to the existing algorithms have an improved SNR for bit-precision of 8-bit. The fixed-point KSVD is approximately 1dB less than the online sparse KSVD and the online GAD algorithm with fixed-point is further less as shown in Tables 1 and 2. The online SVD greedy algorithm is slightly compared to that of the online sparse KSVD, GAD & MSP algorithms and the metrics show that the performance is better. Table.3 shows the computational time of dictionary learning algorithms, and it shows that our algorithm took less time when compared to existing algorithms when run on an Intel I7, 2.8GHz processor with 16GB RAM. The speech quality & intelligibility of denoised speech signal using the proposed & existing methods are shown in Table 4. The metrics used in Table 4 are perceptual quality measured by PESQ, Improved Signal to noise ratio (ISNR) and short-time objective intelligibility (STOI) which indicate the proposed is better.

Conclusion

In this paper, the online greedy algorithm is compared with the existing dictionary learning algorithms and the results show that our algorithm is performing better. The proposed algorithm shows that a complete sparsifying dictionary is learned effectively and efficiently in the enhancement of speech signals. The existing algorithms are computationally complex but the online greedy dictionary learning algorithm is less computationally complex. The fixed-point arithmetic for the resource-constrained devices, the online SVD greedy algorithm gave better performance when compared to the existing ones. Overall, dictionary learning using the proposed method provided better enhancement when compared to the existing algorithms.

Conflict of interest
The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

References

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, *54*(11), 4311–4322. https://doi.org/10.1109/TSP.2006.881199

Bai, Y., Jiang, Q., & Sun, J. (2019). Subgradient descent learns orthogonal dictionaries. *7th International Conference on Learning Representations, ICLR 2019*.

Beheshti, H., Daei, S., & Haddadi, F. (2018). Adaptive Recovery of Dictionary-sparse Signals using Binary Measurements. *ArXiv*, *4*, 1–5. http://arxiv.org/abs/1805.07784

Chi, Y., Lu, Y. M., & Chen, Y. (2019). Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview. *IEEE Transactions on Signal Processing*, *67*(20), 5239–5269. https://doi.org/10.1109/TSP.2019.2937282

Elvira, C., Gribonval, R., Soussen, C., & Herzet, C. (2021). When does OMP achieve exact recovery with continuous dictionaries? *Applied and Computational Harmonic Analysis*, *51*, 374–413. https://doi.org/10.1016/j.acha.2020.12.002

Geng, Q., & Wright, J. (2014). On the local correctness of ℓ1-minimization for dictionary learning. *IEEE International Symposium on Information Theory - Proceedings*, 3180–3184. https://doi.org/10.1109/ISIT.2014.6875421

Hu, Y., & Loizou, P. C. (2007). Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, *49*(7–8), 588–601. https://doi.org/10.1016/j.specom.2006.12.006

Huang, L., Xu, Z., Zhang, Z., & He, Y. (2020). A ratio model of L1/L2 norm for sound source identification. *Sensors (Switzerland)*, *20*(18), 1–17. https://doi.org/10.3390/s20185290

Jafari, M. G., & Plumbley, M. D. (2011). Fast dictionary learning for sparse representations of speech signals. *IEEE Journal on Selected Topics in Signal Processing*, *5*(5), 1025–1031. https://doi.org/10.1109/JSTSP.2011.2157892

Repetti, A., Pham, M. Q., Duval, L., Chouzenoux, É., & Pesquet, J. C. (2015). Euclid in a Taxicab: Sparse blind deconvolution with smoothed ℓ1/ℓ2 regularization. *IEEE Signal Processing Letters*, *22*(5), 539–543. https://doi.org/10.1109/LSP.2014.2362861

Ron Rubinstein, Alfred M. Bruckstein, M. E. (2010). Dictionaries for Sparse Representation Modeling. *IEEE*, *98*(6), 1045–1057. https://doi.org/10.1109/JPROC.2010.2040551

Rubinstein, R., Zibulevsky, M., & Elad, M. (2008). Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *CS Technion*, *January 2008*, 1–15. http://cs.technion.ac.il/users/wwwb/cgi-bin/tr-get.cgi/2008/CS/CS-2008-08.revised.pdf

Srinivas, K. N. H., Prabha, I. S., & Rao, M. V. (2020). Speech Enhancement Based on Offline Dictionary Learning and Fixed-Point Recovery. *SERSC*, *29*(5), 11498–11509. http://sersc.org/journals/index.php/IJAST/article/view/25259

Srinivas, K. N. H., Santhi Prabha, I., & Venugopala Rao, M. (2019). Speech enhancement based on dictionary learning and sparse representation. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(8), 20–30.

Tang, H., Liu, H., Xiao, W., & Sebe, N. (2021). When Dictionary Learning Meets Deep Learning: Deep Dictionary Learning and Coding Network for Image Recognition with Limited Data. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(5), 2129–2141. https://doi.org/10.1109/TNNLS.2020.2997289

Wang, C., Yan, M., Rahimi, Y., & Lou, Y. (2020). Accelerated schemes for the L1/L2 minimization. *IEEE Transactions on Signal Processing*, *68*(1), 2660–2669. https://doi.org/10.1109/TSP.2020.2985298

Yin, P., Esser, E., & Xin, J. (2014). Ratio and difference of L1 and L2 norms and sparse representation with coherent dictionaries. *Communications in Information and Systems*, *14*(2), 87–109. https://doi.org/10.4310/cis.2014.v14.n2.a2

Yuexiang Zhai,Hermish Mehta, Zhengyuan Zhou, Y. M. (2020). Understanding L4-Based Dictionary Learning: Interpretation, Stability, And Robustness. *Iclr*, *7*, 4–6.

Zhai, Y., Yang, Z., Liao, Z., Wright, J., & Ma, Y. (2020). Complete dictionary learning via l4-Norm maximization over the orthogonal group. *Journal of Machine Learning Research*, *21*, 1–68.

Zhang, Y., Kuo, H. W., & Wright, J. (2020). Structured Local Optima in Sparse Blind Deconvolution. *IEEE Transactions on Information Theory*, *66*(1), 419–452. https://doi.org/10.1109/TIT.2019.2940657

Bibliographic information of this paper for citing:

Srinivas, K. N. H.; Santhi Prabha, I. & Venugopala Rao, M. (2023). Speech Enhancement using Greedy Dictionary Learning and Sparse Recovery. *Journal of Information Technology Management,* 15 (Special Issue), 120-132. https://doi.org/10.22059/jitm.2022.89415