Sparse Sequential Generalization of K-means for Dictionary Training on Noisy Signals

Sujit Kumar Sahoo[a,*], Anamitra Makur[a]

[a]*School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore*

Abstract

Noise incursion is an inherent problem in dictionary training on noisy samples. Therefore, enforcing a structural constrain on the dictionary will be useful for a stable dictionary training. Recently, a sparse dictionary with predefined sparsity has been proposed as a structural constraint. However, a fixed sparsity can become too rigid to adapt to the training samples. In order to address this issue, this article proposes a better solution through sparse Sequential Generalization of K-means (SGK). The beauty of the sparse-SGK is that it does not enforce a predefined rigid structure on the dictionary. Instead, a flexible sparse structure automatically emerges out of the training samples depending on the amount of noise. In addition, a variation of sparse-SGK using an orthogonal base dictionary is proposed for a quicker training. The advantages of sparse-SGK are demonstrated via 3-D image denoising. The experimental results confirm that sparse-SGK has better denoising performance and it takes lesser training time.
Keywords: denoising, sparse representation, dictionary training, SGK, sparse dictionary, noise incursion, double sparsity, structured dictionary.

1. Introduction

Sparse representation is one of the advanced tools in signal processing. Taking sparsity as prior, numerous signal processing problems can be addressed,

*Corresponding author

Email addresses: sujit@pmail.ntu.edu.sg (Sujit Kumar Sahoo), eamakur@ntu.edu.sg (Anamitra Makur)

which ranges from compression, denoising, to interpolation. The main ingredient in the sparse representation is the dictionary or the signal transform. The trained dictionaries are the modern trend in signal processing. A dictionary trained on a small class of training samples outperforms the traditional bases. However, the performance of the trained dictionary drops when the training samples are contaminated by noise. It's because the noise get incurred into the trained dictionary.

1.1. Dictionary training on noisy samples

Let's denote the set of noisy samples for dictionary training as $\mathcal{Z} = \{\mathbf{y}_j\}_{j=1}^{M}$ [1]. Each noisy signal $\mathbf{y}_j = \mathbf{x}_j + \mathbf{v}_j$, where $\mathbf{x}_j$ is the original noise free signal and $\mathbf{v}_j$ is the associated noise. Starting with an initial dictionary $\mathbf{D}$, most of the dictionary training algorithms iterate between the *Sparse Signal Representation Stage* and *Dictionary Update Stage* [2].

- *Sparse Signal Representation Stage*: Using a pursuit algorithm, the sparse representation vector $\boldsymbol{\alpha}_j$ for each signal $\mathbf{y}_j$ is estimated, i.e.

$$\forall_j \boldsymbol{\alpha}_j = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ such that } \|\mathbf{y}_j - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \leq \epsilon^2. \tag{1}$$

  where $\epsilon = C\sqrt{n}\sigma$ is the error tolerance for the input noise variance σ^2 with a noise gain C.

- *Dictionary Update Stage*: In the case of sequential update algorithms, each column/atom ($\mathbf{d}_l \in \mathcal{R}^n$) for $l = 1, 2, \ldots, k$ in $\mathbf{D}$ is updated one after another as the following.
  - Find the signals that use the atom $\mathbf{d}_l$, i.e. $\boldsymbol{\omega}_l = \{j|\boldsymbol{\alpha}_j(l) \neq 0\}$.
  - Compute the representation error without the contribution of the atom $\mathbf{d}_l$ for each $j \in \boldsymbol{\omega}_l$, i.e.

$$\mathbf{e}_j^l = \mathbf{y}_j - \sum_{m \neq l} \mathbf{d}_m \boldsymbol{\alpha}_j(m). \tag{2}$$

[1] We kept the notations same as section III of [1]

– Stack all $\{\mathbf{e}_j^l\}_{j\in\boldsymbol{\omega}_l}$ to form a matrix $\mathbf{E}_l \in \mathcal{R}^{n\times|\boldsymbol{\omega}_l|}$, and also form a row vector $\mathbf{a}_l \in \mathcal{R}^{1\times|\boldsymbol{\omega}_l|}$ containing all corresponding $\{\boldsymbol{\alpha}_j(l)\}_{j\in\boldsymbol{\omega}_l}$.

– Update the dictionary column $\mathbf{d}_l$ to minimize

$$\|\mathbf{E}_l - \mathbf{d}_l\mathbf{a}_l\|_F^2 = \sum_{j\in\boldsymbol{\omega}_l} \mathbf{e}_j^l - \mathbf{d}_l\boldsymbol{\alpha}_j(l).$$

K-SVD [3] updates both $\mathbf{d}_l$ and $\mathbf{a}_l$ using SVD that solves

$$\{\mathbf{d}_l, \mathbf{a}_l\} = \arg\min_{\mathbf{d},\mathbf{a}} \|\mathbf{E}_l - \mathbf{d}\mathbf{a}\|_F^2, \tag{3}$$

whereas SGK [2] updates $\mathbf{d}_l$ alone as the solution of

$$\mathbf{d}_l = \arg\min_{\mathbf{d}} \|\mathbf{E}_l - \mathbf{d}\mathbf{a}_l\|_F^2. \tag{4}$$

As a result SGK executes faster than K-SVD with an equally efficient convergence [2].

Inherently, there exists a problem of noise incursion in all such update algorithms, because all the $\mathbf{e}_j^l$ have the noise component $\mathbf{v}_j$ in them. The expression of $\mathbf{e}_j^l$ in equation (2) can be rewritten as

$$\begin{aligned}\mathbf{e}_j^l =& \mathbf{y}_j - \sum_{m=1}^{k} \mathbf{d}_m\boldsymbol{\alpha}_j(m) + \mathbf{d}_l\boldsymbol{\alpha}_j(l) \\ =& \mathbf{v}_j + \mathbf{r}_j + \mathbf{d}_l\boldsymbol{\alpha}_j(l),\end{aligned} \tag{5}$$

where $\mathbf{r}_j = \mathbf{x}_j - \sum_{m=1}^{k} \mathbf{d}_m\boldsymbol{\alpha}_j(m)$ is the original signal residue. Hence, the $\mathbf{d}_l$ that minimizes $\|\mathbf{E}_l - \mathbf{d}_l\mathbf{a}_l\|_F^2$ will have some parts of $\mathbf{v}_j$ in it [1]. However, the noise is usually unstructured and random in nature. Therefore, by enforcing structural constraint on the dictionary atoms, the effect of noise can be eliminated.

1.2. Sparse-KSVD

In order to have implicit trained dictionaries, a sparse structure was proposed in [4]. It represents the dictionary $\mathbf{D} = \boldsymbol{\Phi\Psi}$, where $\boldsymbol{\Phi} \in \mathcal{R}^{n\times s}$ is a fixed base dictionary and $\boldsymbol{\Psi} \in \mathcal{R}^{s\times k}$ is a sparse dictionary. The sparse structure bridges the gap between the implicit and explicit dictionaries. The implicit dictionaries have efficient implementations yet lack adaptability, and explicit dictionaries are fully adaptable but non-efficient and costly to deploy. This sparse dictionary training is known as sparse-KSVD, which involves two steps. In the first, it solves equation (3) using approximate KSVD [5] to obtain an intermediate updated atom $\bar{\mathbf{d}}_l$. In the second step, the corresponding sparse dictionary atom $\boldsymbol{\psi}_l$ is obtained as the following,

$$\boldsymbol{\psi}_l = \arg\min_{\boldsymbol{\psi}} \left\|\bar{\mathbf{d}}_l - \boldsymbol{\Phi}\boldsymbol{\psi}\right\|_2^2 \text{ such that } \|\boldsymbol{\psi}\|_0 \leq p, \tag{6}$$

where p is a predefined sparsity imposed into the dictionary atom. Here sparsity refers to the number of base dictionary atoms used to approximate a trained dictionary atom.

One of the great advantages of sparse structure is its stability in presence of noise. However, the sparse structure enforced in equation (6) can be too rigid. As we don't have any prior knowledge of the underlying signal, a prefixed sparsity p may yield a suboptimal solution. Therefore, the value of p is heuristically fixed through a set of rigorous experiments for the case of the 3-D CT image [4]. To highlight the effect of this rigidity, an illustration is shown in Fig.1. It can be seen that the noise is being approximated in the updated atom by having a predefined sparsity. It is indeed a challenging task to identify the optimal value of p for a particular class of signal.

Therefore, taking the advantage of the noise incursion analysis of SGK in [1], we have proposed a flexible sparse dictionary training for better signal adaptation. We termed this algorithm as sparse-SGK. The enforced sparse structure is completely independent of the signal classes, and the sparsity is automatically decided by the underlying signal.

2. Sparse Sequential Generalization of K-means

The structural constraint enforced in $\mathbf{d}_l$ is that it has a sparse representation on a known base dictionary $\mathbf{\Phi}$.

$$\boldsymbol{\psi}_l = \arg\min_{\boldsymbol{\psi}} \|\boldsymbol{\psi}\|_0 \text{ such that } \mathbf{\Phi}\boldsymbol{\psi} = \mathbf{d}_l,$$

where $\boldsymbol{\psi}_l$ is the sparse representation of the dictionary atom $\mathbf{d}_l$. By adding this structural constraint to the optimization problem of SGK in equation (4), it can be stated in terms of the sparse $\boldsymbol{\psi}_l$ as

$$\boldsymbol{\psi}_l = \arg\min_{\psi} \|\mathbf{E}_l - \mathbf{\Phi}\boldsymbol{\psi}\mathbf{a}_l\|_F^2 + \lambda\|\boldsymbol{\psi}\|_0, \tag{7}$$

where λ is the Lagrangian parameter that determines the sparsity. In other words, λ prevents the noise from being added to the updated atom (i.e. $\lambda \propto n\sigma^2$, the noise variance). The above optimization problem, sparse-SGK, can equivalently be stated as the following equation.

$$\boldsymbol{\psi}_l = \arg\min_{\boldsymbol{\psi}} \left\| \frac{\mathbf{E}_l\mathbf{a}_l^T}{\|\mathbf{a}_l\|_2^2} - \mathbf{\Phi}\boldsymbol{\psi} \right\|_2^2 + \frac{\lambda}{\|\mathbf{a}_l\|_2^2}\|\boldsymbol{\psi}\|_0, \tag{8}$$

To establish the relationship between the problems (7) and (8), we use the following Lemma.

Lemma 1 *Let $\mathbf{E} \in \mathcal{R}^{n\times N}$ and $\mathbf{\Phi} \in \mathcal{R}^{n\times s}$ be matrices, $\boldsymbol{\psi} \in \mathcal{R}^{s\times 1}$ be a column vector, and $\mathbf{a} \in \mathcal{R}^{1\times N}$ be a row vector. Then the following holds:*

$$\|\mathbf{E} - \mathbf{\Phi}\boldsymbol{\psi}\mathbf{a}\|_F^2 = \left\| \frac{\mathbf{E}\mathbf{a}^T}{\|\mathbf{a}\|_2^2} - \mathbf{\Phi}\boldsymbol{\psi} \right\|_2^2 \|\mathbf{a}\|_2^2 + f\left(\mathbf{E}, \mathbf{a}\right)$$

Proof *The equality follows from the definition of Frobenius norm using trace* *function.*

$$
\begin{aligned}
&\|\mathbf{E}-\boldsymbol{\Phi}\boldsymbol{\psi}\mathbf{a}\|_F^2\\
=&tr\left((\mathbf{E}-\boldsymbol{\Phi}\boldsymbol{\psi}\mathbf{a})^T(\mathbf{E}-\boldsymbol{\Phi}\boldsymbol{\psi}\mathbf{a})\right)\\
=&tr\left(\mathbf{E}^T\mathbf{E}\right)-2tr\left(\mathbf{E}^T\boldsymbol{\Phi}\boldsymbol{\psi}\mathbf{a}\right)+tr\left(\mathbf{a}^T\boldsymbol{\psi}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\psi}\mathbf{a}\right)\\
=&tr\left(\mathbf{E}^T\mathbf{E}\right)-2tr\left(\mathbf{a}\mathbf{E}^T\boldsymbol{\Phi}\boldsymbol{\psi}\right)+tr\left(\mathbf{a}\mathbf{a}^T\boldsymbol{\psi}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\psi}\right)\\
=&tr\left(\mathbf{E}^T\mathbf{E}\right)-2tr\left(\mathbf{a}\mathbf{E}^T\boldsymbol{\Phi}\boldsymbol{\psi}\right)+tr\left(\boldsymbol{\psi}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\psi}\right)\|\mathbf{a}\|_2^2\\
&+\frac{tr\left(\mathbf{a}\mathbf{E}^T\mathbf{E}\mathbf{a}^T\right)}{\|\mathbf{a}\|_2^2}-\frac{tr\left(\mathbf{a}\mathbf{E}^T\mathbf{E}\mathbf{a}^T\right)}{\|\mathbf{a}\|_2^2}\\
=&\left(\frac{tr\left(\mathbf{a}\mathbf{E}^T\mathbf{E}\mathbf{a}^T\right)}{\|\mathbf{a}\|_2^2\|\mathbf{a}\|_2^2}-2\frac{tr\left(\mathbf{a}\mathbf{E}^T\boldsymbol{\Phi}\boldsymbol{\psi}\right)}{\|\mathbf{a}\|_2^2}+tr\left(\boldsymbol{\psi}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\psi}\right)\right)\|\mathbf{a}\|_2^2\\
&+tr\left(\mathbf{E}^T\mathbf{E}\right)-\frac{tr\left(\mathbf{a}\mathbf{E}^T\mathbf{E}\mathbf{a}^T\right)}{\|\mathbf{a}\|_2^2}\\
=&\left\|\frac{\mathbf{E}\mathbf{a}^T}{\|\mathbf{a}\|_2^2}-\boldsymbol{\Phi}\boldsymbol{\psi}\right\|_2^2\|\mathbf{a}\|_2^2+f\left(\mathbf{E},\mathbf{a}\right)
\end{aligned}
.
$$

The standard SGK [2] minimizes each $\|\mathbf{E}_l-\mathbf{d}_l\mathbf{a}_l\|_F^2$ by updating $\mathbf{d}_l$ as follows,

$$\bar{\mathbf{d}}_l=\mathbf{E}_l\mathbf{a}_l^T\left(\mathbf{a}_l\mathbf{a}_l^T\right)^{-1}=\frac{\mathbf{E}_l\mathbf{a}^T}{\|\mathbf{a}\|_2^2}=\frac{\sum_{j\in\boldsymbol{\omega}_l}\mathbf{e}_j^l\boldsymbol{\alpha}_j(l)}{\sum_{j\in\boldsymbol{\omega}_l}\boldsymbol{\alpha}_j^2(l)}. \tag{9}$$

Therefore, the sparse-SGK optimization problem in (8) can be restated as

$$\boldsymbol{\psi}_l=\arg\min_{\boldsymbol{\psi}}\left\|\bar{\mathbf{d}}_l-\boldsymbol{\Phi}\boldsymbol{\psi}\right\|_2^2+\frac{\lambda}{\|\mathbf{a}_l\|_2^2}\|\boldsymbol{\psi}\|_0. \tag{10}$$

Two techniques to obtain the solution for the above optimization are described in the next subsections. The steps of sparse-SGK are detailed in Algorithm 1.

2.1. Sparse-SGK with greedy pursuit

It is known from the noise incursion analysis in [1] that the SGK updated atom $\bar{\mathbf{d}}_l$ has the noise component ε_l. The expected energy of ε_l is as follows,

$$\mathbb{E}\left[\|\varepsilon_l\|_2^2\right]=\frac{n\sigma^2}{\sum_{j\in\boldsymbol{\omega}_l}\boldsymbol{\alpha}_j^2(l)}=\frac{n\sigma^2}{\|\mathbf{a}_l\|_2^2}. \tag{11}$$

As a result of the above known fact, the sparse-SGK optimization problem in (10) can be solved using greedy pursuits as the following.

$$\boldsymbol{\psi}_l=\arg\min_{\boldsymbol{\psi}}\|\boldsymbol{\psi}\|_0 \text{ such that } \left\|\bar{\mathbf{d}}_l-\boldsymbol{\Phi}\boldsymbol{\psi}\right\|_2^2\leq\frac{\epsilon^2}{\|\mathbf{a}_l\|_2^2}, \tag{12}$$

where ϵ carries the same value that was used in the sparse signal representation stage of the dictionary training.

Algorithm 1 (Sparse-SGK)

Input:

- *training samples* $\mathbf{Y} \in \mathbf{R}^{n \times M}$
- *base dictionary* $\mathbf{\Phi} \in \mathbf{R}^{n \times s}$
- *initial dictionary* $\mathbf{D} \in \mathbf{R}^{n \times k}$
- *noise tolerance* ϵ
- *number of iterations* t_{max}

Output:

- *sparse dictionary* $\mathbf{\Psi} \in \mathbf{R}^{s \times k}$

Procedure:

1. Initialize *iteration counter* $t = 0$;
2. *Increment* $t = t + 1$;
3. *obtain sparse signal representation for each* $\mathbf{y}_j \in \mathbf{Y}$ *using* $\mathbf{D}$ *and* ϵ;
4. *for each* $\mathbf{d}_l \in \mathbf{D}$
   - *obtain* $\bar{\mathbf{d}}_l$ *using SGK;*
   - *solve* $\boldsymbol{\psi}_l = \arg\min_{\boldsymbol{\psi}} \left\| \bar{\mathbf{d}}_l - \mathbf{\Phi}\boldsymbol{\psi} \right\|_2^2 + \lambda \|\boldsymbol{\psi}\|_0$, *where* λ *is determined by error tolerance* ϵ;
   - *assign* $\mathbf{d}_l = \mathbf{\Phi}\boldsymbol{\psi}_l$;
5. *Go to Step.2 if* $t < t_{\max}$, *else terminate;*
6. *The sparse dictionary* $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_k]$.

2.2. Sparse-SGK with orthogonal base dictionary

Let's take an orthonormal base dictionary $\mathbf{\Phi} \in \mathcal{R}^{n \times n}$ to train the sparse dictionary. Using the norm preserving property of the orthonormal basis, we can rewrite equation (10) as the following,

$$\boldsymbol{\psi}_l = \arg\min_{\boldsymbol{\psi}} \left\| \mathbf{\Phi}^T \bar{\mathbf{d}}_l - \boldsymbol{\psi} \right\|_2^2 + \frac{\lambda}{\|\mathbf{a}_l\|_2^2} \|\boldsymbol{\psi}\|_0. \tag{13}$$

Therefore, the solution can be obtained using hard thresholding as follows,

$$\boldsymbol{\psi}_l(m) = \begin{cases} \boldsymbol{\phi}_m^T \bar{\mathbf{d}}_l & \text{if } |\boldsymbol{\phi}_m^T \bar{\mathbf{d}}_l| \geq \frac{\sqrt{\lambda}}{\|\mathbf{a}_l\|_2} \\ 0 & \text{if } |\boldsymbol{\phi}_m^T \bar{\mathbf{d}}_l| < \frac{\sqrt{\lambda}}{\|\mathbf{a}_l\|_2} \end{cases},$$

where $\sqrt{\lambda} = C\sigma$ is the noise threshold for the input noise variance σ^2 with a noise gain C, and $\boldsymbol{\phi}_m$ is the m^{th} atom/column of the orthonormal base dictionary.

3. Practical Use

The sparse dictionary structure has several advantages which have been highlighted in [4]. Particularly, it enables training of larger dictionaries to be trained for high-dimensional data, and its applications range from image inpaining for larger occlusion to signal compression. In [4], the task of denoising of 3-D computed tomography (CT) imagery has been used to show the advantage of sparse-KSVD over standard KSVD. Therefore, the advantage of sparse-SGK over sparse-KSVD is illustrated through the same experiment. Moreover, the sparse-KSVD results are obtained at its optimal parameters set in [4].

An overcomplete dictionary is trained using blocks from the noisy image, and the trained dictionary is used to denoise the image blocks. Then the denoised blocks are averaged in the overlapping regions to obtain the final denoised image. We would like to mention that, the modern variants of the KSVD or sparse-KSVD based denoising schemes such as [6, 7], can also be improved with sparse-SGK. However, this work focuses on the original denoising scheme for simplicity. The experiments are performed on the Visible Male-Head and Visible Female-Ankle volumes [4]. Each volume is subjected to additive white Gaussian noise with varying standard deviations of $5 \leq \sigma \leq 75$. Both 2-D and 3-D denoising are tested. In 2-D case, each CT slice is processed separately, whereas the volume is processed as a whole in the case of 3-D.

The local block sizes are taken as 8×8 for the 2-D case and $8 \times 8 \times 8$ for the 3-D case. The base dictionary is taken as overcomplete DCT of size 64×100 for the 2-D case and 512×1000 for the 3-D case. This makes the sparse dictionary size 100×100 for the 2-D case and 1000×1000 for the 3-D case. The orthogonal matching pursuit (OMP [8]) is used as the greedy pursuit in the experiment. The sparse signal representation stages of all the dictionary training (1) are performed keeping noise gain value $C = 1.15$ for the 2-D case and $C = 1.04$ for the 3-D case. The same parameters are used for sparse-SGK with greedy pursuit in (12). The sparse atoms as described in (6) are obtained using OMP with fixed atom sparsity $p = 6$ for the 2-D case and $p = 16$ for the 3-D case, which is empirically found to be optimal [4]. Sparse-SGK with orthogonal base dictionary is performed taking 2-D and 3-D DCTs, which makes the sparse dictionary size 64×100 for the 2-D case and 512×1000 for the 3-D case. The sparse atoms are obtained using the hard-thresholding solution of (13) with noise gain value $C = 3.45$ for the 2-D case and $C = 3.12$ for the 3-D case. All the dictionary training algorithms are iterated for 15 iterations.

The sparsity statistics of the trained dictionary atoms at various input noise levels are shown in Fig. 2. The plots show that the flexibility of sparse-SGKs allow atoms to have diversified sparsity. This also shows that the sparsity level in sparse-SGKs emerges automatically from the training samples depending on the noise. As the noise level becomes higher, a more sparser approximation is used for the updated atoms. The denoising results are summarized in Table 1. Some actual denoising results are shown in Fig. 3. These results highlight that the flexibility brought in by sparse-SGK provides better scope for adaptation in low noise level. This also confirms the observation of the first experiment, that at low noise level sparse-SGK is better adapting to the signal by allowing more number of base dictionary atoms. Similarly, it is automatically restricting the adaptation at higher noise levels by using less number of base dictionary atoms, which produces a coarse approximation of the signal equivalent to sparse-KSVD. Table 2 shows the execution times obtained by MATLAB 2014b running on a Intel Core i5 CPU clocked at 3.10 GHz for all the algorithms. It can be clearly seen that the sparse-SGKs perform better than sparse-KSVD both in terms of denoising and training time.

3.1. Reproducible Research

The complete Sparse-KSVD and Sparse-SGK codes reproducing the results in this paper, along with the original CT volumes used, will be made available for download at https://sites.google.com/site/sujitkusahoo/codes. This code uses the Sparse-KSVD code that was made available by the authors of [4] in support of reproducible research.

4. Discussion

Apart from bridging the gap between the implicit and explicit dictionaries, the sparse dictionary provides greater stability in presence of noise. However, the sparse dictionary in its known form, particularly the sparse-KSVD, is restrictive in nature. It enforces the train dictionary atom to have a fixed sparsity representation on the base dictionary, as a result of which the adaptability of the dictionary becomes dependent on this predefined sparsity. The optimal value of sparsity can only be fixed heuristically, which will also vary with different class of signals. This drawback of sparse dictionary is addressed by introducing sparse-SGK, where the sparsity level atomically emerges from the training samples. The improvement in dictionary adaptation along with increase in computational speed is illustrated through denoising experiment.

References

[1] S. Sahoo, A. Makur, Enhancing image denoising by controlling noise incursion in learned dictionaries, Signal Processing Letters, IEEE 22 (8) (2015) 1123–1126.

[2] S. Sahoo, A. Makur, Dictionary training for sparse representation as generalization of k-means clustering, Signal Processing Letters, IEEE 20 (6) (2013) 587–590.

[3] M. Aharon, M. Elad, A. Bruckstein, K -svd: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Processing 54 (11) (2006) 4311–4322.

[4] R. Rubinstein, M. Zibulevsky, M. Elad, Double sparsity: Learning sparse dictionaries for sparse signal approximation, IEEE Trans. Signal Processing 58 (3) (2010) 1553–1564.

[5] R. Rubinstein, M. Zibulevsky, M. Elad, Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit, Tech. rep. (apr 2008).

[6] J. Mairal, G. Sapiro, M. Elad, Learning multiscale sparse representations for image and video restoration, Multiscale Modeling & Simulation 7 (1) (2008) 214–241.

[7] M. Protter, M. Elad, Image sequence denoising via sparse and redundant representations, Image Processing, IEEE Transactions on 18 (1) (2009) 27–35.

[8] S. Sahoo, A. Makur, Signal recovery from random measurements via extended orthogonal matching pursuit, IEEE Trans. Signal Processing 63 (10) (2015) 2572–2581.

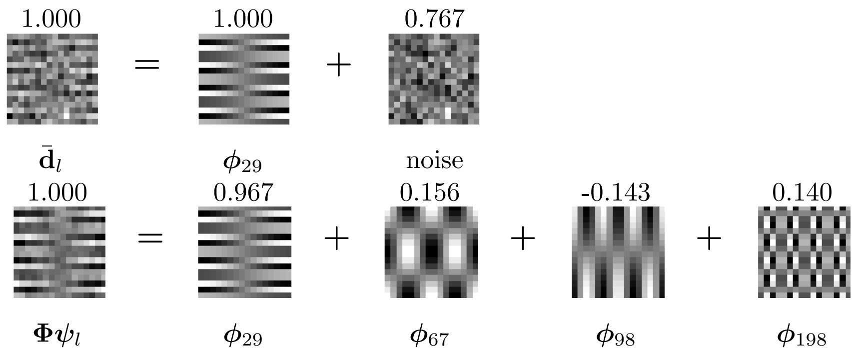


Figure 1: Illustrating the effect of fixed sparsity (e.g. $p \leq 4$), where the dictionary $\mathbf{D}$ is trained on 16×16 noisy image patches, and the base dictionary $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_{256}]$ is 2-D Discrete Cosine Transform (DCT). Each displayed atom/image is of unit norm, and the number on top represents the coefficient. Let the actual updated atom $\bar{\mathbf{d}}_l$ be $\boldsymbol{\phi}_{29}$ contaminated with noise of standard deviation 0.767. The atom $\bar{\mathbf{d}}_l$ is then approximated by $\boldsymbol{\phi}_{29}$, $\boldsymbol{\phi}_{67}$, $\boldsymbol{\phi}_{98}$ and $\boldsymbol{\phi}_{198}$ using (6). However, the best approximation could have been $\boldsymbol{\phi}_{29}$ alone, i.e. $p = 1$.

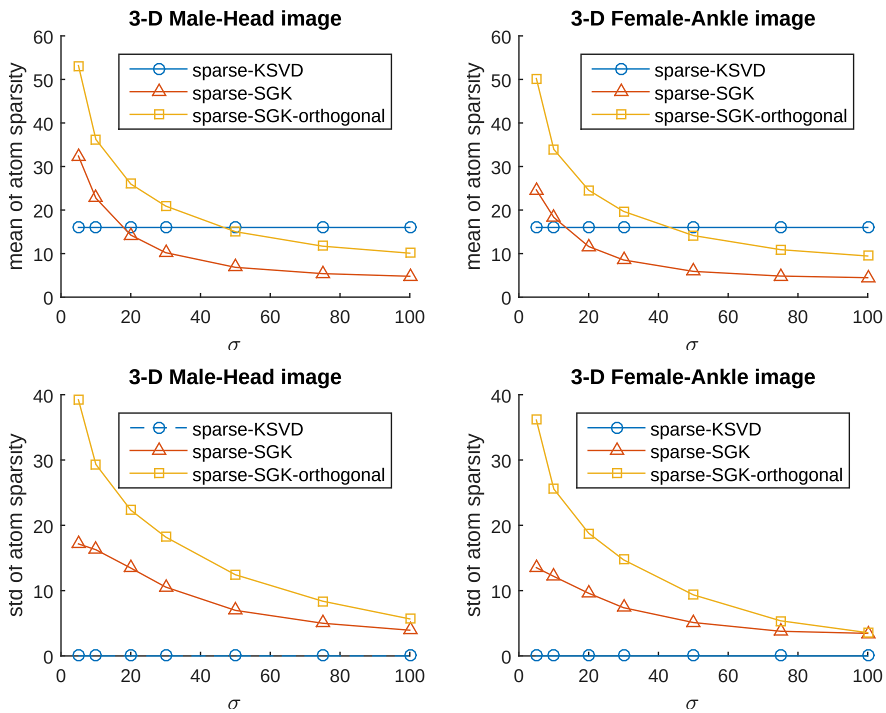
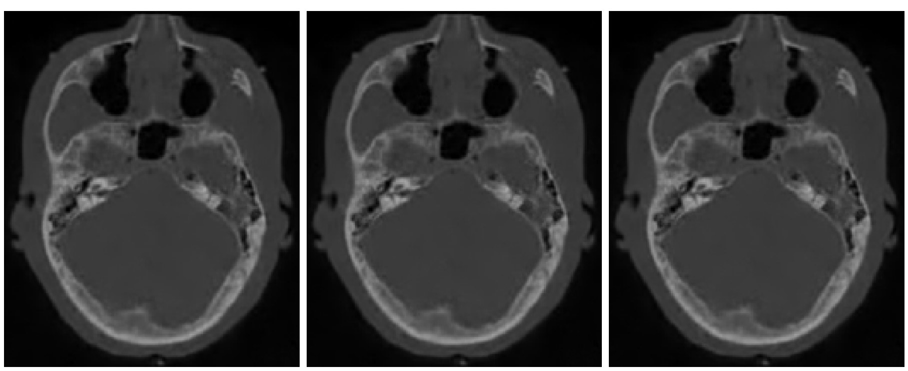


Figure 2: The sparsity statistics of the trained dictionary atoms at various noise level σ. The top plots show the mean sparsity, and the bottom plots show the standard deviation of sparsity. The shown results are average of 10 trials.



Sparse-KSVD	Sparse-SGK	Sparse-SGK orthogonal
(PSNR 38.12 dB)	(PSNR 38.23 dB)	(PSNR 38.23 dB)

Figure 3: 3-D denoising results for *Male-Head*, slice #98 ($\sigma = 20$). Images are provided for qualitative evaluation, and are best viewed by zooming-in using a computer display.

Table 1: Comparison of the CT denoising PSNR results in decibel (dB). In each cell three denoising results are reported. Top: using standard Sparse-KSVD learned dictionary [4]. Middle: using Sparse-SGK. Bottom: using Sparse-SGK with orthonormal basis. All numbers are an average over 10 trials. The last two columns present the average result over two images. Bold numerals denote the improved results in each test with at least **0.1 dB** difference.

	F. Ankle		M. Head		Average	
σ/PSNR	3D	2D	3D	2D	3D	2D
5/34.15	44.66	43.64	45.17	41.38	44.91	42.51
	44.77	**43.74**	**45.30**	**41.54**	**45.03**	**42.64**
	44.71	43.73	45.23	**41.55**	44.97	**42.64**
10/28.13	41.22	39.68	41.57	37.22	41.39	38.45
	41.36	**39.84**	**41.73**	**37.48**	**41.55**	**38.66**
	41.32	**39.85**	41.66	**37.50**	**41.49**	**38.68**
20/22.11	38.04	35.79	38.11	33.42	38.08	34.60
	38.18	**35.95**	**38.22**	**33.54**	**38.20**	**34.74**
	38.18	**35.96**	**38.23**	**33.56**	**38.20**	**34.76**
30/18.59	36.21	33.54	36.17	31.05	36.19	32.29
	36.30	33.60	36.20	31.06	36.25	32.33
	36.34	33.63	36.24	31.09	**36.29**	32.36
50/14.15	33.84	29.76	33.56	27.99	33.70	28.88
	33.90	29.72	33.53	27.99	33.71	28.86
	33.93	29.79	33.56	27.98	33.75	28.88
75/10.63	32.01	26.90	31.09	26.29	31.55	26.60
	32.03	26.89	31.12	26.29	31.57	26.59
	32.05	26.92	31.11	26.28	31.58	26.60

Table 2: Comparison of dictionary training time in seconds. In each cell three times are reported. Top: using standard Sparse-KSVD learned dictionary [4]. Middle: using Sparse-SGK. Bottom: using Sparse-SGK with orthonormal basis. Bold numerals denote the best results in each test.

	F. Ankle		M. Head		Average	
σ/PSNR	3D	2D	3D	2D	3D	2D
5/34.15	115.09 106.01 **94.27**	2.04 1.88 **1.74**	108.87 98.19 **89.83**	2.83 2.56 **2.43**	111.98 102.10 **92.05**	2.43 2.22 **2.09**
10/28.13	61.09 56.93 **48.56**	1.50 1.38 **1.26**	61.55 55.65 **49.13**	1.91 1.77 **1.63**	61.32 56.29 **48.85**	1.71 1.57 **1.45**
20/22.11	39.64 36.22 **30.63**	1.26 1.15 **1.03**	41.27 36.94 **32.04**	1.45 1.36 **1.23**	40.45 36.58 **31.34**	1.36 1.26 **1.13**
30/18.59	33.88 30.38 **25.99**	1.16 1.08 **0.96**	35.58 31.58 **27.33**	1.25 1.15 **1.04**	34.73 30.98 **26.66**	1.21 1.11 **1.00**
50/14.15	29.39 25.94 **22.35**	1.05 0.97 **0.86**	30.64 26.83 **23.33**	1.06 0.97 **0.86**	30.02 26.38 **22.84**	1.06 0.97 **0.86**
75/10.63	26.84 23.21 **20.27**	0.98 0.90 **0.79**	26.79 23.06 **20.28**	0.98 0.90 **0.79**	26.81 23.14 **20.28**	0.98 0.90 **0.79**