Unit-Norm Tight Frame-based Sparse Representation With Application to Speech Inpainting

Huang Bai[1], Chuanrong Hong[2], Sheng Li[3], Yimin D. Zhang[4], Xiumei Li[1*]

[1]*School of Information Science and Technology, Hangzhou Normal University, Hangzhou, 311121, Zhejiang, China*
[2]*Alibaba Group, Hangzhou, 311121, Zhejiang, China*
[3]*College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310023, Zhejiang, China*
[4]*Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19312, USA*

Abstract

This paper deals with the unit-norm tight frames (UNTFs) that better represent and solve sparse representation problems. Under this sparse representation framework, a novel model termed tight dictionary learning (TDL) is proposed. Unlike those dictionaries that only focus on the signals of interest, TDL intends to learn a more comprehensive dictionary that better represents the signals with the UNTF properties to facilitate sparse recovery. To better describe the physical meanings, the normalization constraints on the dictionary columns are imposed. A gradient-based approach is developed to obtain the tight dictionary where the normalization constraints are embedded into the cost function by a parametrization method. The learned tight dictionary is applied to the speech inpainting problem to restore speech signals corrupted by data missing. In order to compensate the degradation effect of the data missing, a UNTF-based preconditioning technique is developed to reform the frame properties of the degraded system. Parametrization and gradient-based approaches similar to those applied in the TDL are adopted to design a tight preconditioner. Extensive experiments on speech denoising and speech segment restoration are conducted to demonstrate the superiority of the proposed schemes.

Keywords:

Sparse representation, dictionary learning, unit-norm tight frame, speech inpainting, preconditioning.

1. Introduction

Signal representation intends to transform interested signals to a different domain via a basis, where certain signal characteristics become more readily apparent in the transform coefficients for facilitating various signal processing tasks. The notion of basis in finite-dimensional spaces implies that the number of representative vectors is the same as the dimension of the space, which means that the representation is nonredundant and, as a result, corruption or loss of transform coefficients may cause serious consequence. In order to provide a robust transformation, redundant signal representation is usually adopted. The redundant counterpart of a basis is called a *frame* [1], [2].

Frames give stable signal representations and allow modelling for noisy environments [1]. Signal representation using a frame expresses a signal vector $\boldsymbol{x} \in \Re^{N\times 1}$ in the form of

$$\boldsymbol{x} \;\approx\; \sum_{k=1}^{K} \boldsymbol{s}(k)\boldsymbol{D}(:,k) \;\triangleq\; \boldsymbol{D}\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{D} \in \Re^{N\times K}$ $(N < K)$ is a frame matrix with its columns $\{\boldsymbol{D}(:,k)\}$ referred to as atoms, and $\boldsymbol{s} \in \Re^{K\times 1}$ is the corresponding coefficient vector. If $\boldsymbol{s}$ is sparse, i.e., there are few non-zero elements in $\boldsymbol{s}$, equation (1) is specified as a *sparse representation* (SR) of $\boldsymbol{x}$, and the frame matrix $\boldsymbol{D}$ is named as a *dictionary* [3], [4].

Signal SR theory is an evolving theory with state-of-the-art results applied in many signal processing tasks, such as denoising, video encryption, speech recognition, and compressed sensing [5]-[8]. Frames are less constrained than bases and thus enable more flexible representations of corresponding signals [1], [2]. Hence, investigation of frame-based SR is of great importance.

A fundamental requirement in employing the SR theory is the proper choice of the dictionary and this leads to the well-known *dictionary learning* (DL) problem [9]. A great volume of works have been performed to learn a data-adaptive dictionary so that a particular class of signals can be sparsely represented in this dictionary with a low approximation error [4], [10], [11].

Let $\{\boldsymbol{x}_l\}_{l=1}^{L}$ be a set of training signals. The basic problem of DL is to find a dictionary $\boldsymbol{D}$ such that for each $\boldsymbol{x}_l$ there exists a sparse representation vector $\boldsymbol{s}_l$. Denote

$$\boldsymbol{X} \;\triangleq\; \left[\begin{array}{cccc} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_L \end{array}\right] \in \Re^{N\times L}, \qquad \boldsymbol{S} \;\triangleq\; \left[\begin{array}{cccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \cdots & \boldsymbol{s}_L \end{array}\right] \in \Re^{K\times L}.$$

The general DL problem is formulated as

$$\min_{\boldsymbol{D},\,\boldsymbol{S}} \; \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}\|_F^2 + f(\boldsymbol{D}) + f(\boldsymbol{S}), \tag{2}$$

where $\|\cdot\|_F$ denotes the *Frobenius* norm, and $f(\boldsymbol{D})$ and $f(\boldsymbol{S})$ are some regularization functions of $\boldsymbol{D}$ and $\boldsymbol{S}$, respectively. In general, $f(\boldsymbol{D})$ may be subject to the matrix column-normalization constraint, whereas $f(\boldsymbol{S})$ is set as the sparsity constraint [4]. It is a common assumption that $f(\boldsymbol{D})$ has nothing to do with $\boldsymbol{S}$ and $f(\boldsymbol{S})$ is also independent of $\boldsymbol{D}$.

Assume $\boldsymbol{v}(k)$ as the k-th element of a vector $\boldsymbol{v} \in \Re^{K\times 1}$. The ℓ_p-norm of vector $\boldsymbol{v}$ is defined as

$$\|\boldsymbol{v}\|_p \;\triangleq\; \left(\sum_{k=1}^{K} |\boldsymbol{v}(k)|^p\right)^{1/p}, \quad p \geq 1.$$

For convenience, $\|\boldsymbol{v}\|_0$ is used to denote the number of non-zero elements in $\boldsymbol{v}$.

A common choice of $f(\boldsymbol{S})$ in (2) is $f(\boldsymbol{S}) = \sum_{l=1}^{L} \tau_l \|\boldsymbol{s}_l\|_0$ (throughout this paper, this $f(\boldsymbol{S})$ will be retained) with $\{\tau_l\}$ denoting proper weighting constants [10], [11]. This leads to

$$\min_{\boldsymbol{D},\,\{\boldsymbol{s}_l\}_{l=1}^{L}} \; \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}\|_F^2 + f(\boldsymbol{D}) + \sum_{l=1}^{L} \tau_l \|\boldsymbol{s}_l\|_0. \tag{3}$$

Such a problem is difficult to solve as it is non-convex in $\boldsymbol{D}$ and $\boldsymbol{S}$, and $\|\cdot\|_0$ is non-smooth and highly unstable. A popularly used approach is based on the alternating minimization strategy [4], [10], [11]. A two-stage procedure is usually carried out for solving the above problem and also avoiding the selections of $\{\tau_l\}$. The first stage is referred to as *sparse coding* or *sparse decomposition*, also known as *sparse recovery* in signal reconstruction, aiming at finding the (column) sparse matrix $\boldsymbol{S}$ with a given dictionary matrix $\boldsymbol{D}$ through the following inverse problem

$$\hat{\boldsymbol{s}}_l \triangleq \arg\min_{\boldsymbol{s}_l} \|\boldsymbol{x}_l - \boldsymbol{D}\boldsymbol{s}_l\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{s}_l\|_0 \leq \kappa, \ \ \forall\, l, \tag{4}$$

with κ denoting the sparsity level. The roles of the penalty and the constraints in (4) may be reversed [4]. We can choose to constrain the SR error and obtain the sparsest representation by

$$\hat{\boldsymbol{s}}_l \triangleq \arg\min_{\boldsymbol{s}_l} \|\boldsymbol{s}_l\|_0 \quad \text{s.t.} \quad \|\boldsymbol{x}_l - \boldsymbol{D}\boldsymbol{s}_l\|_2^2 \leq \epsilon, \ \ \forall\, l, \tag{5}$$

where ϵ denotes the error threshold. Such a problem can be solved using the orthogonal matching pursuit (OMP) based methods [12]-[14]. Furthermore, a popular strategy is to replace the ℓ_0-norm by the ℓ_1-norm, which is its convex approximation in a natural sense, leading to

$$\hat{\boldsymbol{s}}_l \triangleq \arg\min_{\boldsymbol{s}_l} \|\boldsymbol{s}_l\|_1 \quad \text{s.t.} \quad \|\boldsymbol{x}_l - \boldsymbol{D}\boldsymbol{s}_l\|_2^2 \leq \epsilon, \ \ \forall\, l. \tag{6}$$

Equation (6) is convex and can be solved using algorithms such as the basis pursuit (BP) [15] and the ℓ_1-based optimization techniques [16].

Many methods that solve (3) differentiate each other mainly in the second stage, namely, *dictionary updating*. In other words, we need to determine what kind of frames should be selected for coding the signals of interest more sparsely and precisely. In classical DL problems with normalization constraints on dictionary columns $\{\boldsymbol{D}(:,k)\}$, i.e., $f(\boldsymbol{D}) = \sum_{k=1}^{K} \left(\|\boldsymbol{D}(:,k)\|_2^2 - 1\right)^2$ in (3), the dictionary is learned adaptively based on the training signals [11]. With such a strategy, some inherent properties of the dictionary, e.g., the tightness property that is concerned in this paper, may lose. It is desirable to learn a dictionary with a desired representation capability for signals as well as possessing certain excellent frame properties to facilitate sparse coding.

For a linear system satisfying expression (1), matrix $\boldsymbol{D}$ is called the *system matrix*. References [17] and [18] investigated the optimized design of the system matrix that minimizes the mean squared error (MSE) of the oracle estimator [19]. The oracle MSE coincides with the unbiased Cramér-Rao bound for sparse deterministic vectors and represents the best achievable performance for any unbiased estimator [20]. Besides, the oracle MSE performance acts as a benchmark to the performance of various sparse recovery algorithms. For example, it has been demonstrated both theoretically and numerically that the Dantzig selector, the OMP and thresholding algorithms all achieve performance that is proportional to the oracle MSE [20]. Since minimizing the MSE of the oracle estimator with respect to the system matrix is a non-convex problem, Chen *et al.* adopted convex relaxation techniques and proved that a *unit-norm*

tight frame (UNTF) is the closest design in the Frobenius norm sense to the solution of the relaxed problem [17], [18]. UNTFs offer the advantage of redundancy in signal representations and numerical stability of reconstruction and, therefore, have received great interests in recent years in the signal processing community [21]-[25].

A UNTF is viewed as a generalization of an orthonormal basis which brings redundancy and stability [22]. Generally speaking, the properties of a UNTF refer to the *unit-norm* (i.e., normalization) property of the atoms and the *tightness* property of the frame. The unit-norm property avoids the explosion or degeneration of the frame elements and the sparse coefficients, and the tightness property is valuable to ensure fast convergence of signal sparse decompositions [22], [26]. Besides, according to [23], the properties of the UNTF have been proved valid against additive noise and erasures.

The main objective of this paper is to develop a *tight dictionary learning* (TDL) method to learn a dictionary that minimizes the SR error and approximates to a UNTF. In the sequel, tight dictionary implies that both the unit-norm property and the tightness property hold. As speech signals satisfy the SR model with a high probability [27]-[30] and UNTFs show great superiority in sparse signal processing [21]-[25], the learned tight dictionary can be applied to speech inpainting applications. A *preconditioning* technique [31]-[33] based on UNTF is further presented to improve the inpainting performance. Utilizing the properties of the UNTF to design the system matrix to facilitate sparse recovery is termed as *tight sparse representation* (TSR).

The contributions of this paper are four-fold:

- A novel model is proposed for learning a tight dictionary. The inherent properties of the dictionary matrix, including the unit-norm property and the tightness property, are taken into consideration when minimizing the SR error. Because of the UNTF properties that facilitate the sparse recovery, the learned dictionary can well represent the signals of interest.
- An alternating minimization strategy is developed to solve the TDL problem. When updating the dictionary, a parametrization method is carried out to convert the constrained optimization problem to an unconstrained one. A gradient-based approach is developed to solve the parameterized problem and obtain the tight dictionary. Such an updating procedure results in a precise solution of the TDL problem.
- When applied to speech inpainting problems, the impairment of data missing is effectively mitigated. The data missing process is formulated as a degradation matrix that removes samples from the speech signals. Noting the sparse structure of the speech signals in the SR model, the learned tight dictionary is applied to effectively restore the missing data.
- In order to mitigating the degradation effect of data missing, the preconditioning technique based on UNTF is employed. Similar parametrization and gradient-based approaches as in the TDL are adopted to design the tight preconditioner. Such a preconditioner highly reforms the UNTF properties of the degraded system to facilitate the sparse recovery, thereby improving the accuracy of the speech inpainting results.

In addition, extensive experiments on speech data and real speech segments are carried out to demonstrate the superior performance of the proposed TSR schemes.

The remainder of this paper is arranged as follows. In Section 2, some preliminaries are provided and two related works are introduced as the baselines to evaluate our method. A novel model for learning the tight dictionary with desirable representation capability is described in Section 3. In Section 4, application of the SR theory and the preconditioning technique to speech inpainting is discussed in detail. Experiments on speech signals are carried out in Section 5 to examine the performance of the proposed schemes and to compare with existing techniques. Concluding remarks are given in Section 6.

In this paper, we use italic characters to denote scalars, whereas lowercase and uppercase bold italic characters indicate vectors and matrices, respectively. All parameters are real-valued. Besides, MATLAB notations are used to denote the selections of elements in a vector or a matrix.

2. Preliminaries and related works

In this section, we review some preliminaries on frames, summarize two important works related to TDL, and point out a problem that is usually encountered in DL.

2.1. Brief review on frames

A finite frame $\boldsymbol{D} \in \mathfrak{R}^{N\times K}$ in the Hilbert space $\mathfrak{R}^{N\times 1}$ is a set of $K \geq N$ vectors $\{\boldsymbol{D}(:,k)\}_{k=1}^{K}$ that satisfies the following generalized Parseval condition:

$$\alpha\|\boldsymbol{v}\|_2^2 \;\leq\; \sum_{k=1}^{K} |\langle \boldsymbol{D}(:,k), \boldsymbol{v}\rangle|^2 \;\leq\; \beta\|\boldsymbol{v}\|_2^2 \tag{7}$$

for all $\boldsymbol{v} \in \mathfrak{R}^{N\times 1}$, where α and β are two positive constants referred to as the lower and upper bounds, respectively [1], [2]. If it is possible to have $\alpha = \beta$ in (7), then we achieve an α-tight frame with α being referred to as the tightness constant in this case. A frame $\{\boldsymbol{D}(:,k)\}_{k=1}^{K}$ is α-tight if and only if the corresponding matrix $\boldsymbol{D}$ has a singular value decomposition (SVD) in the following form

$$\boldsymbol{D} = \boldsymbol{U}\begin{bmatrix} \sqrt{\alpha}\boldsymbol{I}_N & \boldsymbol{0} \end{bmatrix}\boldsymbol{V}^{\mathcal{T}}, \tag{8}$$

where superscript $\mathcal{T}$ denotes the transpose operator, $\boldsymbol{I}_N$ is the $N\times N$ identity matrix, and both $\boldsymbol{U} \in \mathfrak{R}^{N\times N}$ and $\boldsymbol{V} \in \mathfrak{R}^{K\times K}$ are orthonormal matrices [34], [35].

When designing a frame for a specific application, it is important to control over the spectrum of the corresponding frame matrix [21], [22]. This spectral characteristic is usually studied via the condition number of the frame matrix. The closer the condition number is to 1, the better conditioned the frame matrix will be [36]. A linear system with a

well-conditioned system matrix is said to be stable. It is clear that the condition number of a matrix expressed in (8) is 1, which means that a tight frame forms a well-conditioned frame matrix.

Where all the atoms have the same unit ℓ_2-norm, those frames are called unit-norm frames, i.e., $\|\boldsymbol{D}(:,k)\|_2^2 = 1$ for all k. A frame $\boldsymbol{D}$ is named UNTF when both tightness property and unit-norm property hold [22]-[26]. According to the definitions of the tight frame and the unit-norm frame, it is straightforward to obtain

$$\alpha N \;=\; \mathrm{tr}\Big[\boldsymbol{D}\boldsymbol{D}^{\mathcal{T}}\Big] \;=\; \mathrm{tr}\Big[\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}\Big] \;=\; K \qquad \Rightarrow \qquad \alpha \;=\; \frac{K}{N}, \tag{9}$$

by appealing to the cyclic property of the trace operation tr[·] [36]. It means that the tightness constant of any UNTF is K/N, i.e., the redundancy of corresponding frame.

2.2. Related works of TDL

The two DL schemes proposed in [37] and [38] are closely related to the work presented in this paper, and we briefly introduce them in this part.

Based on the general DL problem (3) and the alternating minimization strategy to solve it, we reformulate (3) as:

$$\min_{\boldsymbol{D},\,\{\boldsymbol{s}_l\}_{l=1}^{L}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{S}\|_F^2 + f(\boldsymbol{D}) \quad \text{s.t.} \quad \|\boldsymbol{s}_l\|_0 \le \kappa, \;\; \forall\, l. \tag{10}$$

In [37], the regularization term $f(\boldsymbol{D})$ in (10) is set as

$$f(\boldsymbol{D}) = \omega\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2, \tag{11}$$

with ω being the regularization parameter. Substituting (11) into (10) renders the following DL model:

$$\min_{\boldsymbol{D},\,\{\boldsymbol{s}_l\}_{l=1}^{L}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{S}\|_F^2 + \omega\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{s}_l\|_0 \le \kappa, \;\; \forall\, l. \tag{12}$$

By employing the alternating minimization strategy, $\{\boldsymbol{s}_l\}_{l=1}^{L}$ are computed by certain sparse recovery algorithm like OMP, and (12) turns to

$$\min_{\boldsymbol{D}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{S}\|_F^2 + \omega\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2. \tag{13}$$

A gradient-based approach is then carried out for solving (13) [37].

In order to avoid the choice of the regularization parameter ω, [38] modifies the problem (13) to

$$\min_{\boldsymbol{D}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{S}\|_F^2 \quad \text{s.t.} \quad \min_{\boldsymbol{D}} \|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2. \tag{14}$$

That is, they set the regularization term $f(\boldsymbol{D})$ indicated in (11) as the constraint. The closed-form solution set of $\{\min_{\boldsymbol{D}} f(\boldsymbol{D})\}$ was derived in [34] as:

$$\boldsymbol{D} \;=\; \boldsymbol{U}\Big[\;\boldsymbol{I}_N \quad \boldsymbol{0}\;\Big]\boldsymbol{V}^{\mathcal{T}}, \tag{15}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices of dimensions $N\times N$ and $K\times K$, respectively. These two matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are then alternately optimized to minimize the SR error and improve the representation capability of the dictionary [38].

Comparing (15) and (8), it is clear that the result of $\{\min_{\boldsymbol{D}} f(\boldsymbol{D})\}$ with $f(\boldsymbol{D})$ indicated in (11) is exactly the form of the 1-tight frame which is closely related to TDL.

2.3. A shortcoming in general DL

As introduced in Section 1, classical DL problems take the normalization constraints on dictionary columns as the regularization term $f(\boldsymbol{D})$ because these constraints can avoid the explosion or degeneration of the matrix elements and the sparse coefficients. This is a common choice almost in all DL literatures. For works like [37] and [38], more constraints on the dictionary are made and these models, including TDL, are sometimes called constrained DL.

Most works take two steps to update the dictionary with normalization constraints. First, compute a dictionary without considering the normalization constraints, e.g., (13) and (14), and then execute an extra normalization step to the half-finished dictionary [8]-[10], [37], [38]. However, such a separated strategy compromises the optimality of the solution as those two steps are not compatible. It is important to maintain the normalization throughout the learning of the dictionary.

From the perspective of designing frames, these facts motivate us to develop a novel model with tightness and unit-norm properties taken into consideration simultaneously.

3. Tight dictionary learning model and algorithm

In this section, a novel model for TDL is proposed which minimizes the SR error and, at the same time, approximates a UNTF that maintains the tightness and unit-norm properties. Parametrization and gradient-based approaches are developed to solve the corresponding design problem.

3.1. Tight dictionary learning model

According to *Proposition 2* in [17], any $N \times K$ UNTF is a solution to the optimization problem

$$\min_{\boldsymbol{D}} \ \|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D} - \boldsymbol{I}_K\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{D}(:,k)\|_2^2 = 1, \ \ \forall\, k. \tag{16}$$

As indicated in *Theorem 1* of [34], the closed-form solution set of (16) can be derived as

$$\boldsymbol{D} \ = \ \boldsymbol{U}\left[\ \sqrt{\tfrac{K}{N}}\boldsymbol{I}_N \quad \boldsymbol{0}\ \right]\boldsymbol{V}^{\mathcal{T}}, \tag{17}$$

with the orthonormal matrix $\boldsymbol{V}$ properly designed to ensure the normalization constraints in (16) to hold. As shown in [34] and [39], such a $\boldsymbol{V}$ can always be achieved. The result (17), i.e., the solution of (16), is a K/N-tight frame with unit-norm atoms, which is the exact definition of an $N \times K$ UNTF [22], [26]. This is consistent with that indicated in

(8) and (9). For convenience, the term tight dictionary is used to represent a dictionary possessing the properties of the UNTF that ensure tightness and unit-norm.

Recalling the definition of DL, the capability of accurately representing the dictionary to signals of interest must be enforced. Therefore, in addition to the inherent properties of the dictionary, a signal-adapted penalty term is also needed. Based on the discussions above, we propose the following TDL model:

$$\min_{\boldsymbol{D},\,\{\boldsymbol{s}_l\}_{l=1}^{L}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{S}\|_F^2+\omega\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{s}_l\|_0\le\kappa,\ \forall\, l, \quad \|\boldsymbol{D}(:,k)\|_2^2=1,\ \forall\, k, \tag{18}$$

where ω trades off the significance between the representation accuracy and the frame inherent properties. As the identity matrix has the smallest condition number, model (18) will lead to a well-conditioned dictionary with a flat spectrum. Such a dictionary results in a stable representation system [31], [36], [37].

Remark 1.

- The main difference between (18) and (12) lies in the dictionary column-normalization constraints. These constraints are set for two reasons. On one hand, it gives better description of the physical meaning to hold the unit-norm. As shown in [34], without the normalization constraints, the solution to

  $$\arg\min_{\boldsymbol{D}}\ \|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2 \ =\ \boldsymbol{U}\begin{bmatrix}\boldsymbol{I}_N & \boldsymbol{0}\end{bmatrix}\boldsymbol{V}^{\mathcal{T}},$$

  is a 1-tight frame but is not necessarily a UNTF. On the other hand, it avoids the degeneration of the sparse coefficients. In equation (1), if some of the elements in $\boldsymbol{D}$ become very large, small valid values in $\boldsymbol{s}$ may be easily neglected.

- Frame potential is a famous concept related to the UNTFs and has been defined explicitly in [22] as:

  $$\mathrm{FP}(\boldsymbol{D})\ \triangleq\ \sum_{i=1}^{K}\sum_{j=1}^{K}|\langle\boldsymbol{D}(:,i),\boldsymbol{D}(:,j)\rangle|^2\ =\ \|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}\|_F^2.$$

  A number of methods were developed to construct UNTFs by minimizing the frame potential [25], [40], [41]. It can be observed that the regularization term $\{\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2\}$ is closely associated with the frame potential if the unit-norm property holds, i.e., $\mathrm{FP}(\boldsymbol{D})=\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D}-\boldsymbol{I}_K\|_F^2+K$ when $\forall\, k,\ \|\boldsymbol{D}(:,k)\|_2^2=1$. This fact once again verifies the importance of the normalization constraints.

- Equation (17) is a closed-form expression of the UNTF when a proper $\boldsymbol{V}$ is set [34]. We do not constrain the form of the dictionary as (17) because there only exists one orthonormal matrix $\boldsymbol{U}$ to improve the representation capability. The impact of the sole orthonormal matrix on the SR capability is limited. The weighting model described in (18) is flexible to learn a desired dictionary by selecting a proper value of ω.

3.2. Tight dictionary learning algorithm

This part is devoted to the detailed algorithm description for solving the TDL problem. Noting that, because (18) is still non-convex in $\boldsymbol{D}$ and $\boldsymbol{S}$, and $\|\cdot\|_0$ is non-smooth and highly unstable, the alternating minimization strategy is employed. The pseudo-code is given in **Algorithm 1**.

Algorithm 1: Tight dictionary learning algorithm

Initialization:

$\boldsymbol{D}^{[0]} \in \mathfrak{R}^{N\times K}$: initial column-normalized dictionary;

$\boldsymbol{X} = \{\boldsymbol{x}_l\}_{l=1}^{L} \in \mathfrak{R}^{N\times L}$: training signal set;

κ: sparsity level;

N_{iter}: number of iterations for DL.

Set $i = 1$.

Begin: For $i = 1 : N_{\text{iter}}$, repeat the following two stages:

- *Sparse coding:* With $\boldsymbol{D}^{[i-1]}$, update the sparse coefficient set $\boldsymbol{S}^{[i]} = \{\hat{\boldsymbol{s}}_l\}_{l=1}^{L}$ by

$$\hat{\boldsymbol{s}}_l \;=\; \arg\min_{\boldsymbol{s}_l} \|\boldsymbol{x}_l - \boldsymbol{D}^{[i-1]}\boldsymbol{s}_l\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{s}_l\|_0 \le \kappa, \;\; \forall\, l \tag{19}$$

  using a certain OMP-based algorithm.

- *Dictionary updating:* For $\boldsymbol{S}^{[i]}$ fixed, update the dictionary with

$$\boldsymbol{D}^{[i]} \;=\; \arg\min_{\boldsymbol{D}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}^{[i]}\|_F^2 + \omega\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D} - \boldsymbol{I}_K\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{D}(:,k)\|_2^2 = 1, \;\; \forall\, k. \tag{20}$$

  Because of the normalization constraints in (20), the closed-form solution is hard to be derived. A parametrization method will be adopted to embed the normalization constraints on the dictionary columns into the cost function, and then a gradient-based approach is developed to solve the parameterized problem for the dictionary. This will be described in detail below.

End: End the algorithm, and output the tight dictionary.

As (19) is solved by an OMP-based algorithm, the remaining issue is to update the dictionary, i.e., solving (20), which is described below. For notational convenience, the iteration index $[i]$ will be omitted in the following.

Denote $\varrho(\boldsymbol{D}) \triangleq \|\boldsymbol{E}_1\|_F^2 + \omega\|\boldsymbol{E}_2\|_F^2$ with

$$\boldsymbol{E}_1 \;\triangleq\; \boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}, \quad \boldsymbol{E}_2 \;\triangleq\; \boldsymbol{D}^{\mathcal{T}}\boldsymbol{D} - \boldsymbol{I}_K. \tag{21}$$

The minimization problem in (20) turns to

$$\min_{\boldsymbol{D}} \varrho(\boldsymbol{D}) \quad \text{s.t.} \quad \|\boldsymbol{D}(:,k)\|_2^2 = 1, \;\; \forall\, k. \tag{22}$$

It is difficult to derive the closed-form solution of (22) due to the normalization constraints. Referring to [42], we present a gradient-based approach to solve (22). A key step of this approach is the following parametrization of the dictionary

$$\boldsymbol{D} \;\triangleq\; \boldsymbol{Q}\boldsymbol{Q}_d, \tag{23}$$

where $\boldsymbol{Q}$ is an $N \times K$ matrix with a mild condition that it contains no zero columns, and $\boldsymbol{Q}_d \in \Re^{K\times K}$ is a diagonal matrix formed as

$$\boldsymbol{Q}_d \;=\; \mathrm{diag}\left(\|\boldsymbol{Q}(:,1)\|_2^{-1},\; \|\boldsymbol{Q}(:,2)\|_2^{-1},\; \cdots,\; \|\boldsymbol{Q}(:,K)\|_2^{-1}\right). \tag{24}$$

It is clear that the columns of $\boldsymbol{D}$ defined in (23) are inherently ℓ_2-normalized as $\boldsymbol{Q}$ has no zero columns.

With the definitions of (23) and (24), the dictionary updating problem (22) is converted to the following unconstrained minimization problem

$$\min_{\boldsymbol{Q}}\; \varrho(\boldsymbol{Q}\boldsymbol{Q}_d). \tag{25}$$

(25) can be solved using the following gradient descent procedure:

$$\boldsymbol{Q}_n \;=\; \boldsymbol{Q}_{n-1} - \gamma \frac{\partial \varrho}{\partial \boldsymbol{Q}}\bigg|_{\boldsymbol{Q}=\boldsymbol{Q}_{n-1}}, \tag{26}$$

where γ is the step size, and $\frac{\partial \varrho}{\partial \boldsymbol{Q}}$ is the gradient of $\varrho(\boldsymbol{Q}\boldsymbol{Q}_d)$ with respect to $\boldsymbol{Q}$.

Define

$$\varrho_1(\boldsymbol{Q}) \;\triangleq\; \|\boldsymbol{E}_1\|_F^2, \quad \varrho_2(\boldsymbol{Q}) \;\triangleq\; \|\boldsymbol{E}_2\|_F^2 \tag{27}$$

with $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ defined in (21). Then, one has

$$\varrho \;=\; \varrho_1(\boldsymbol{Q}) + \omega\varrho_2(\boldsymbol{Q}). \tag{28}$$

Hence

$$\frac{\partial \varrho}{\partial \boldsymbol{Q}} \;=\; \frac{\partial \varrho_1(\boldsymbol{Q})}{\partial \boldsymbol{Q}} + \omega \frac{\partial \varrho_2(\boldsymbol{Q})}{\partial \boldsymbol{Q}}. \tag{29}$$

The expressions of $\frac{\partial \varrho_1(\boldsymbol{Q})}{\partial \boldsymbol{Q}}$ and $\frac{\partial \varrho_2(\boldsymbol{Q})}{\partial \boldsymbol{Q}}$ can be obtained by the following Theorem 1 and Theorem 2, respectively.

Theorem 1. *Let $\boldsymbol{Q} \in \Re^{N\times K}$ be a matrix with no zero columns and $\boldsymbol{Q}_d$ is defined in (24). Denote*

$$\boldsymbol{E}_1 \;\triangleq\; \boldsymbol{X} - (\boldsymbol{Q}\boldsymbol{Q}_d)\boldsymbol{S}. \tag{30}$$

The gradient of $\|\boldsymbol{E}_1\|_F^2$ with respect to $\boldsymbol{Q}$ is obtained as

$$\frac{\partial(\|\boldsymbol{E}_1\|_F^2)}{\partial \boldsymbol{Q}} \;=\; 2\boldsymbol{Q}\boldsymbol{Q}_d^3\boldsymbol{R}_1 - 2\boldsymbol{E}_1\boldsymbol{S}^{\mathcal{T}}\boldsymbol{Q}_d, \tag{31}$$

where

$$\boldsymbol{R}_1 \;\triangleq\; \mathrm{diag}\left(\boldsymbol{H}_1(1,1),\; \boldsymbol{H}_1(2,2),\; \cdots,\; \boldsymbol{H}_1(K,K)\right), \qquad \boldsymbol{H}_1 \;\triangleq\; \boldsymbol{S}\boldsymbol{E}_1^{\mathcal{T}}\boldsymbol{Q}. \tag{32}$$

Theorem 2. *Let $\boldsymbol{Q} \in \Re^{N\times K}$ be a matrix with no zero columns and $\boldsymbol{Q}_d$ is defined in (24). Denote*

$$\boldsymbol{E}_2 \;\triangleq\; (\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d) - \boldsymbol{I}_K. \tag{33}$$

The gradient of $\|\boldsymbol{E}_2\|_F^2$ with respect to $\boldsymbol{Q}$ is obtained as

$$\frac{\partial(\|\boldsymbol{E}_2\|_F^2)}{\partial \boldsymbol{Q}} \;=\; 2(\boldsymbol{Q}\boldsymbol{Q}_d)(\boldsymbol{E}_2^{\mathcal{T}} + \boldsymbol{E}_2)\boldsymbol{Q}_d - 2\boldsymbol{Q}\boldsymbol{Q}_d^3\boldsymbol{R}_2, \tag{34}$$

where

$$\boldsymbol{R}_2 \;\triangleq\; \mathrm{diag}\left(\boldsymbol{H}_2(1,1),\; \boldsymbol{H}_2(2,2),\; \cdots,\; \boldsymbol{H}_2(K,K)\right), \qquad \boldsymbol{H}_2 \;\triangleq\; \boldsymbol{Q}^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)(\boldsymbol{E}_2^{\mathcal{T}} + \boldsymbol{E}_2). \tag{35}$$

The proofs of the above two theorems can be found in Appendices A and B, respectively.

Based on Theorems 1 and 2, the expression of $\frac{\partial \varrho}{\partial \boldsymbol{Q}}$ is obtained as

$$\frac{\partial \varrho}{\partial \boldsymbol{Q}} \;=\; \frac{\partial \varrho_1(\boldsymbol{Q})}{\partial \boldsymbol{Q}} + \omega \frac{\partial \varrho_2(\boldsymbol{Q})}{\partial \boldsymbol{Q}} \;=\; 2\boldsymbol{Q}\boldsymbol{Q}_d^3\boldsymbol{R}_1 - 2\boldsymbol{E}_1\boldsymbol{S}^{\mathcal{T}}\boldsymbol{Q}_d + \omega\left[2(\boldsymbol{Q}\boldsymbol{Q}_d)(\boldsymbol{E}_2^{\mathcal{T}} + \boldsymbol{E}_2)\boldsymbol{Q}_d - 2\boldsymbol{Q}\boldsymbol{Q}_d^3\boldsymbol{R}_2\right]. \tag{36}$$

Using the result of (36), $\boldsymbol{Q}$ is updated via (26) by the gradient descent procedure, and the solution of (22) is obtained as $\boldsymbol{D} = \boldsymbol{Q}\boldsymbol{Q}_d$.

Remark 2.

- When $\omega \to +\infty$, the dictionary updating turns to the UNTF construction problem which is an attractive topic [25], [40], [41]. With the Lagrange multiplier, the problem (20) becomes

$$\min_{\boldsymbol{D}} \; \|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D} - \boldsymbol{I}_K\|_F^2 + \sum_{k=1}^{K} \lambda_k(\|\boldsymbol{D}(:,k)\|_2^2 - 1).$$

  The above expression is also known as the famous Paulsen problem [26] with the two penalty terms concerning the tightness property and the unit-norm property, respectively. When dealing with multiple objects, some works alternately optimize these properties until the cost function converges (sometimes the alternating optimization method may not converge) [32], while we search for the tight frame with its atoms inherently normalized.

- For the proposed parametrization and gradient-based strategy, other penalty terms that subject to the unit-norm constraints are convenient to be added to the cost function, e.g., the signal-adapted penalty term $\{\|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}\|_F^2\}$ considered in this work. The computations of the gradients such as (31) and (34) are not complex. This strategy can be conveniently extended to design high dimensional matrices because it does not involve computationally expensive operations such as matrix inversion or SVD.

- It can be seen from the pseudo-code that **Algorithm 1** is carried out with respect to $\boldsymbol{D}$ and $\boldsymbol{S}$. The convergence of our algorithm can be easily guaranteed. For sparse coding, the popularly adopted OMP is viewed as an efficient method to solve (19) [12]-[14]. Therefore, in the i-th iteration,

$$\|\boldsymbol{x}_l - \boldsymbol{D}^{[i-1]}\hat{\boldsymbol{s}}_l\|_2^2 \;\le\; \|\boldsymbol{x}_l - \boldsymbol{D}^{[i-1]}\boldsymbol{s}_l\|_2^2, \;\; \forall\, l = 1, 2, \cdots, L,$$

  for any $\{\boldsymbol{s}_l\}$. Hence, with $\boldsymbol{S}^{[i]} = \{\hat{\boldsymbol{s}}_l\}_{l=1}^{L}$ and $\boldsymbol{S}^{[i-1]}$ being the result obtained in the previous iteration,

$$\|\boldsymbol{X} - \boldsymbol{D}^{[i-1]}\boldsymbol{S}^{[i]}\|_F^2 \;\le\; \|\boldsymbol{X} - \boldsymbol{D}^{[i-1]}\boldsymbol{S}^{[i-1]}\|_F^2$$

  can be ensured. This leads to

$$\|\boldsymbol{X} - \boldsymbol{D}^{[i-1]}\boldsymbol{S}^{[i]}\|_F^2 + \omega\|(\boldsymbol{D}^{[i-1]})^{\mathcal{T}}(\boldsymbol{D}^{[i-1]}) - \boldsymbol{I}_K\|_F^2 \;\le\; \|\boldsymbol{X} - \boldsymbol{D}^{[i-1]}\boldsymbol{S}^{[i-1]}\|_F^2 + \omega\|(\boldsymbol{D}^{[i-1]})^{\mathcal{T}}(\boldsymbol{D}^{[i-1]}) - \boldsymbol{I}_K\|_F^2.$$

  Considering dictionary updating, i.e., solving (20), as long as the step size γ in (26) is properly set, the corresponding cost function is non-increasing. Therefore, $\boldsymbol{D}^{[i]}$ resulted from the gradient-based approach guarantees

$$\|\boldsymbol{X} - \boldsymbol{D}^{[i]}\boldsymbol{S}^{[i]}\|_F^2 + \omega\|(\boldsymbol{D}^{[i]})^{\mathcal{T}}(\boldsymbol{D}^{[i]}) - \boldsymbol{I}_K\|_F^2 \;\le\; \|\boldsymbol{X} - \boldsymbol{D}^{[i-1]}\boldsymbol{S}^{[i]}\|_F^2 + \omega\|(\boldsymbol{D}^{[i-1]})^{\mathcal{T}}(\boldsymbol{D}^{[i-1]}) - \boldsymbol{I}_K\|_F^2.$$

  Subsequently,

$$\|\boldsymbol{X} - \boldsymbol{D}^{[i]}\boldsymbol{S}^{[i]}\|_F^2 + \omega\|(\boldsymbol{D}^{[i]})^{\mathcal{T}}(\boldsymbol{D}^{[i]}) - \boldsymbol{I}_K\|_F^2 \;\le\; \|\boldsymbol{X} - \boldsymbol{D}^{[i-1]}\boldsymbol{S}^{[i-1]}\|_F^2 + \omega\|(\boldsymbol{D}^{[i-1]})^{\mathcal{T}}(\boldsymbol{D}^{[i-1]}) - \boldsymbol{I}_K\|_F^2$$

  is guaranteed throughout the iteration indexed by i.

4. Tight sparse representation-based speech inpainting

Speech inpainting, leveraging the concept of image inpainting, refers to filling in missing data whose locations are assumed to be known *a priori* in the speech signals [4], [28]-[30]. In this section, the learned tight dictionary described in the previous section is applied to speech inpainting problems where the speech signals are corrupted by data missing. We formulate the missing process as a degradation matrix that removes samples from the speech signals and employ preconditioning technique to promote the speech restoration.

4.1. Formulation of data missing

Data missing is defined as a general problem that a partial set of reliable speech data is observed while the remaining unreliable data are totally missing. Speech inpainting aims to estimate the unreliable (missing) data from the reliable data portion. This process is also known as missing data imputation [28]-[30].

According to [27], speech signals satisfy the SR model with a high probability. Assume that $\boldsymbol{x} \in \mathfrak{R}^{N\times 1}$ is a speech signal, meaning that there exists a sparse coefficient vector $\boldsymbol{s} \in \mathfrak{R}^{K\times 1}$ with $\|\boldsymbol{s}\|_0 \leq \kappa$ ($\kappa \ll N$) satisfying $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{s}$ with $\boldsymbol{D} \in \mathfrak{R}^{N\times K}$ denoting a pre-specified dictionary.

Further assume that M samples in $\boldsymbol{x}$ are observed with *a priori* known positions. The degradation matrix $\boldsymbol{L} \in \mathfrak{R}^{M\times N}$ that records the missing process is a matrix which has only a single element of one and other elements are zeros in each row. Such a matrix can be expressed as

$$\boldsymbol{L}(i,j) \;=\; \begin{cases} 1, & \text{when } j = \mathcal{J}(i), \\ 0, & \text{otherwise}, \end{cases} \tag{37}$$

where i is the element index number of the observed signal, and $1 \leq \mathcal{J}(i) \leq N$ is the index number of corresponding element in the original signal $\boldsymbol{x}$. Matrix $\boldsymbol{L}$ removes $N - M$ samples from $\boldsymbol{x}$ and this process is viewed as data missing. This matrix can also be described as taking the $N \times N$ identity matrix and removing the $N - M$ rows corresponding to the missing samples.

Denote the observed signal as $\boldsymbol{y}$. The speech data missing problem is formulated as

$$\boldsymbol{y} \;\triangleq\; \boldsymbol{L}\boldsymbol{x} \;\approx\; \boldsymbol{L}\boldsymbol{D}\boldsymbol{s}. \tag{38}$$

When Gaussian noise is added to the observations, speech inpainting should be considered by solving the following inverse problem

$$\hat{\boldsymbol{s}} \;\triangleq\; \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_0 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{L}\boldsymbol{D}\boldsymbol{s}\|_2^2 \leq \epsilon, \tag{39}$$

with ϵ being the error threshold determined by the noise level. Then the restored signal would become $\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{s}}$.

Remark 3.

- The above strategy would result in successful missing data imputation based on a presupposed condition that $\boldsymbol{x}$ can be sparsely represented using dictionary $\boldsymbol{D}$. This condition can be satisfied by the previously proposed TDL algorithm. As will be seen in the experiment results, the algorithm learns a tight dictionary with promising SR capability. In addition, when the degradation matrix $\boldsymbol{L} = \boldsymbol{I}_N$ (i.e., $M = N$, implying that no sample is missing), (39) can be viewed as a speech denoising problem and will also be considered in the experimental studies to demonstrate the performance of the designed tight dictionary.
- Equation (38) is similar to a compressed sensing model, where $\boldsymbol{L} \in \Re^{M\times N}$ is the sensing matrix [8]. In many works, with the dictionary $\boldsymbol{D}$ fixed, the sensing matrix $\boldsymbol{L}$ is designed to make the product $\boldsymbol{LD}$ possess certain properties such as the tightness property to facilitate the sparse recovery process, thereby improving the accuracy of solution indicated in (39) [17], [18], [24], [25]. This is infeasible in our degradation model because $\boldsymbol{L}$ is determined by the data missing process and cannot be optimized. Depending on the value of $\boldsymbol{L}$, the property of $\boldsymbol{LD}$ may be destroyed even when $\boldsymbol{D}$ is well designed. This fact will significantly compromise the solution of (39). Motivated by this, we introduce the preconditioning technique to modify (39) and ensure the UNTF properties.

4.2. Construction of tight preconditioner

Preconditioning is one of the most critical ingredients in the development of efficient solvers for challenging problems in scientific computation [31]-[33]. The term preconditioning refers to transforming a model like (38) into another form with more favorable properties. A preconditioner is a matrix that affects such a transformation. Generally speaking, preconditioning attempts to improve the spectral properties of the degraded system matrix $\boldsymbol{LD}$. Hopefully, the transformed (preconditioned) system matrix will have a smaller spectral condition number. A well-conditioned system matrix will lead to a stable sparse coding procedure [31].

Assume $\boldsymbol{LD}$ to be an $M \times K$ matrix that does not satisfy the necessary conditions required for sparse recovery. Suppose that there exists a nonsingular matrix $\boldsymbol{P} \in \Re^{M\times M}$ such that the product $\boldsymbol{PLD}$ possesses certain desired properties to facilitate the sparse recovery process. In this case, $\boldsymbol{P}$ is referred to as a preconditioner. Multiplying both sides of (38) by $\boldsymbol{P}$, we define

$$\boldsymbol{z} \triangleq \boldsymbol{Py} = \boldsymbol{PLx} \approx \boldsymbol{As}, \tag{40}$$

where $\boldsymbol{A} \triangleq \boldsymbol{PLD}$ is the preconditioned (system) matrix. As $\boldsymbol{P}$ is invertible, the solution $\boldsymbol{s}$ of (40) is the same as that of (38). The difference between (38) and (40) in calculating $\boldsymbol{s}$ lies in the properties of $\boldsymbol{A}$ and $\boldsymbol{LD}$. The effectiveness of applying algorithms such as OMP to solve (40) for sparse coefficient $\boldsymbol{s}$ depends strongly on the property of $\boldsymbol{A}$. The question is how to construct a nonsingular $\boldsymbol{P}$ such that $\boldsymbol{s}$ can be precisely obtained via

$$\hat{\boldsymbol{s}} \triangleq \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_0 \quad \text{s.t.} \quad \|\boldsymbol{z} - \boldsymbol{As}\|_2^2 \le \epsilon, \tag{41}$$

with $\boldsymbol{z}$ and $\boldsymbol{A}$ both known, then $\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{s}}$ is recovered.

Equation (41) is similar to the sparse coding procedure in DL. Therefore, the UNTF properties are preferred. The preconditioner construction problem is formulated as

$$\min_{\boldsymbol{P}\in\Re^{M\times M}} \|\boldsymbol{A}^{\mathcal{T}}\boldsymbol{A} - \boldsymbol{I}_K\|_F^2 \quad \text{s.t.} \quad \boldsymbol{A} = \boldsymbol{PLD} \ \text{ and } \ \|\boldsymbol{A}(:,k)\|_2^2 = 1, \ \ \forall\, k. \tag{42}$$

Note that the column-normalization constraints are included because our goal is to approximate a UNTF in a strict way. The resulting $\boldsymbol{P}$ obtained from (42) is named as a tight preconditioner.

Once again, due to the normalization constraints in (42), it is difficult to derive a closed-form solution. Denote

$$\tilde{\boldsymbol{A}} \triangleq \boldsymbol{L}\boldsymbol{D} \in \mathfrak{R}^{M\times K}, \tag{43}$$

and

$$\tilde{\boldsymbol{Q}} \triangleq \boldsymbol{P}\tilde{\boldsymbol{A}} \in \mathfrak{R}^{M\times K} \tag{44}$$

is determined by $\boldsymbol{P}$ with a mild condition that it contains no zero columns. The gradient-based approach will be employed and the key step is the following parametrization

$$\boldsymbol{A} \triangleq \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d \tag{45}$$

with $\tilde{\boldsymbol{Q}}_d \in \mathfrak{R}^{K\times K}$ being a diagonal matrix formed as

$$\tilde{\boldsymbol{Q}}_d = \operatorname{diag}\left(\|\tilde{\boldsymbol{Q}}(:,1)\|_2^{-1},\ \|\tilde{\boldsymbol{Q}}(:,2)\|_2^{-1},\ \cdots,\ \|\tilde{\boldsymbol{Q}}(:,K)\|_2^{-1}\right). \tag{46}$$

It is clear that the columns of $\boldsymbol{A}$ in (45) are inherently normalized, yielding the following unconstrained problem:

$$\min_{\boldsymbol{P}\in\mathfrak{R}^{M\times M}} \left\{\|(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d) - \boldsymbol{I}_K\|_F^2 \triangleq \tilde{\varrho}(\boldsymbol{P})\right\} \quad \text{s.t.} \quad (44), (46). \tag{47}$$

Similar to (25), equation (47) can be solved using the gradient descent procedure:

$$\boldsymbol{P}_n = \boldsymbol{P}_{n-1} - \eta \frac{\partial\tilde{\varrho}(\boldsymbol{P})}{\partial\boldsymbol{P}}\bigg|_{\boldsymbol{P}=\boldsymbol{P}_{n-1}}, \tag{48}$$

where η denotes the step size, and $\frac{\partial\tilde{\varrho}(\boldsymbol{P})}{\partial\boldsymbol{P}}$ is the gradient of $\tilde{\varrho}(\boldsymbol{P})$ with respect to $\boldsymbol{P}$ and can be computed based on Theorem 3 below.

Theorem 3. *Given a known matrix $\tilde{\boldsymbol{A}} \in \mathfrak{R}^{M\times K}$, let $\boldsymbol{P} \in \mathfrak{R}^{M\times M}$ be a nonsingular matrix such that $\tilde{\boldsymbol{Q}} = \boldsymbol{P}\tilde{\boldsymbol{A}}$ has no zero columns. Denote*

$$\tilde{\boldsymbol{E}} \triangleq (\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d) - \boldsymbol{I}_K, \tag{49}$$

with $\tilde{\boldsymbol{Q}}_d$ defined in (46). The gradient of $\|\tilde{\boldsymbol{E}}\|_F^2$ with respect to $\boldsymbol{P}$ is calculated as

$$\frac{\partial(\|\tilde{\boldsymbol{E}}\|_F^2)}{\partial\boldsymbol{P}} = 2(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)(\tilde{\boldsymbol{E}}^{\mathcal{T}} + \tilde{\boldsymbol{E}})\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{A}}^{\mathcal{T}} - 2\boldsymbol{P}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{R}}\tilde{\boldsymbol{A}}^{\mathcal{T}}, \tag{50}$$

where

$$\tilde{\boldsymbol{R}} \triangleq \operatorname{diag}\left(\tilde{\boldsymbol{H}}(1,1),\ \tilde{\boldsymbol{H}}(2,2),\ \cdots,\ \tilde{\boldsymbol{H}}(K,K)\right), \qquad \tilde{\boldsymbol{H}} \triangleq (\tilde{\boldsymbol{E}}^{\mathcal{T}} + \tilde{\boldsymbol{E}})(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d^3. \tag{51}$$

The proof of Theorem 3 can be found in Appendix C.

With the result of Theorem 3, $\boldsymbol{P}$ is updated via (48) by the gradient descent procedure, and the designed tight preconditioner can be obtained as $\lim_{n\to+\infty}\boldsymbol{P}_n$.

Remark 4.

- In our interested scenarios, the dictionary $\boldsymbol{D}$ is learned in advance and the degradation matrix $\boldsymbol{L}$ is decided according to the data missing process. Therefore, the property of $\boldsymbol{LD}$ is unpredictable and it may not satisfy the necessary conditions for sparse recovery. That is the reason why we introduce the preconditioning.
- In general, a good preconditioner should lead to a model which can be easily handled. The UNTF properties can make the inverse problem easier to be solved and facilitate sparse recovery [17], [18]. We design the preconditioner to make the preconditioned matrix closely approximate the UNTF by enclosing the normalization constraints.
- Looking from a different perspective, a well-designed preconditioned matrix will result in effective sparse coding [31], [37]. Choosing the identity matrix as the target in (42) makes the physical meanings of the model clearer because the identity matrix has the smallest condition number. This is a key point as the purpose of the preconditioning technique is to improve the spectral properties of the system matrix [31].
- A similar preconditioner construction problem has been investigated in [32]. The problem is described as

$$\min_{\boldsymbol{P}\in\Re^{M\times M}} \|\boldsymbol{A}^{\mathcal{T}}\boldsymbol{A}-\boldsymbol{G}_t\|_F^2 \quad \text{s.t.} \quad \boldsymbol{A}=\boldsymbol{P}\tilde{\boldsymbol{A}},$$

  where $\boldsymbol{G}_t$ is the target Gram matrix possessing some desired properties. The authors of [32] first computed a matrix $\boldsymbol{A}$ for the above problem and then solved a least squares problem to obtain $\boldsymbol{P} = \arg\min_{\boldsymbol{P}}\{\|\boldsymbol{A}-\boldsymbol{P}\tilde{\boldsymbol{A}}\|_F^2\} = \boldsymbol{A}\tilde{\boldsymbol{A}}^{\mathcal{T}}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{A}}^{\mathcal{T}})^{-1}$. A similar strategy is adopted in [25] to design the sensing matrix for a compressed sensing system. Such a two-step strategy compromises the optimality of the solution. Using equation (48) and Theorem 3, we can directly construct the preconditioner for problem (42).
- The proposed preconditioner construction method shares some of the ideas in our recent paper [33]. Different from [33], however, the data missing process is taken into consideration in this paper, and the parametrization and gradient-based approaches are described in detail.

5. Experiment results

In this section, we evaluate the performance of the proposed models and algorithms with speech signals that are popularly used for testing the effectiveness of the constrained DL [37], [38].

The speech data are extracted from the GRID corpus [43]. Every recording in the corpus is decomposed into 50% overlapping time-domain frames with a frame length of $N = 256$. A data matrix consisting of $100,000$ columns is formed by randomly selecting 100 frames from each of the $1,000$ recordings in the corpus. A twice overcomplete dictionary $\boldsymbol{D}^{[0]}$ (i.e., $K = 512$) is initialized by choosing K columns of the data matrix followed by a column-normalization procedure. The training data matrix $\boldsymbol{X} = \{\boldsymbol{x}_l\}_{l=1}^{L}$ is obtained as a subset of the data matrix with $L = 10,000$ columns to learn the dictionaries. The testing data matrix $\tilde{\boldsymbol{X}} = \{\tilde{\boldsymbol{x}}_l\}_{l=1}^{\tilde{L}}$ is obtained in the same way as $\boldsymbol{X}$ but with $\tilde{L} = 1,000$ columns. It should be noted that the initial dictionary, the training data matrix, and the testing data matrix are all selected randomly without intersection. When learning the dictionaries, the OMP algorithm is employed for sparse coding with sparsity $\kappa = 12$ which corresponds to about 5% of active elements when compared with the length N.

The standard deviation of the Gaussian noise added to the testing data $\tilde{\boldsymbol{X}}$ is denoted as σ. The recovery accuracy is quantified with the output signal-to-noise ratio (SNR) [28]-[30] defined as

$$\rho_{snr} \triangleq 20\log_{10}\frac{\|\tilde{\boldsymbol{X}}\|_F}{\|\tilde{\boldsymbol{X}}-\boldsymbol{D}\tilde{\boldsymbol{S}}\|_F}, \tag{52}$$

where $\boldsymbol{D}$ is the output dictionary of each learning method and $\tilde{\boldsymbol{S}} = \{\tilde{\boldsymbol{s}}_l\}_{l=1}^{\tilde{L}}$ is the sparse coefficient matrix calculated from

$$\tilde{\boldsymbol{s}}_l = \arg\min_{\boldsymbol{s}_l}\|\boldsymbol{s}_l\|_1 \quad \text{s.t.} \quad \|\tilde{\boldsymbol{x}}_l - \boldsymbol{D}\boldsymbol{s}_l\|_2^2 \le \epsilon, \ \ \forall\, l$$

by the ℓ_1-Homotopy algorithm [16] with $\epsilon = 0.5N\sigma^2$. This process can be viewed as the speech denoising problem.

5.1. Performance of the tight dictionary

We now carry out experiments to illustrate the performance of the dictionaries learned with different algorithms. Algorithms developed in [11], [37] and [38] are compared. For convenience, the learning systems are denoted as $\mathbf{Dict}_{\rm NEW}$, $\mathbf{Dict}_{\rm KSVD}$ [11], $\mathbf{Dict}_{\rm SDB}$ [37], and $\mathbf{Dict}_{\rm BLJ}$ [38] for the proposed TDL and the algorithms in the respective references. The number of iterations for DL is set to $N_{\rm iter} = 100$ for all the comparisons. For $\mathbf{Dict}_{\rm NEW}$, the gradient descent is executed for 50 times with step size $\gamma = 0.1$. The extra parameter settings for $\mathbf{Dict}_{\rm KSVD}$, $\mathbf{Dict}_{\rm SDB}$ and $\mathbf{Dict}_{\rm BLJ}$ are kept the same as those in the corresponding references.

5.1.1. Choice of ω in (12) and (18)

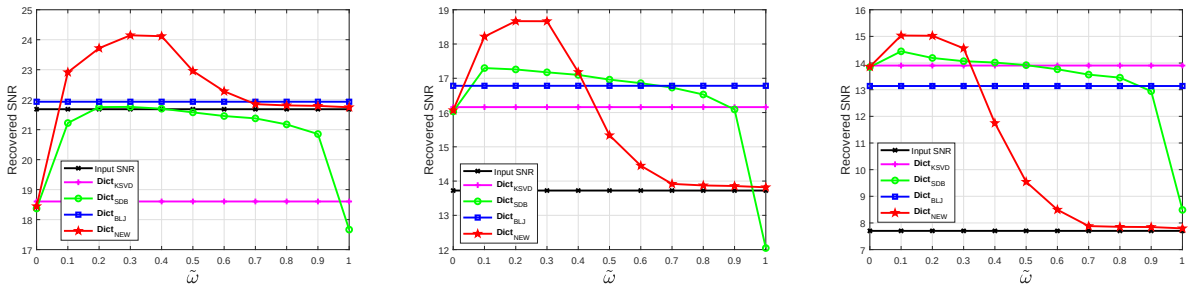
Let us consider the effect of factor ω. In order to test the tradeoff between the representation capability and the tightness property roundly, we denote

$$\tilde{\omega} \triangleq \frac{\omega}{1+\omega}.$$

When ω increases from $0 \to +\infty$, $\tilde{\omega}$ changes monotonically from $0 \to 1$.

Gaussian noises with different standard deviations are added to the testing data $\tilde{\boldsymbol{X}}$. Figures 1 (a), (b) and (c) show the results of the recovered SNR with respect to $\tilde{\omega}$. The input SNR is the SNR of the noisy speech determined by the noise standard deviation σ.

For Figures 1 (a), (b) and (c), the significance of the regularization term $\|\boldsymbol{D}^{\mathcal{T}}\boldsymbol{D} - \boldsymbol{I}_K\|_F^2$ is judged by comparing the cases for $\tilde{\omega} = 0$ with the others, while the significance of $\|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{S}\|_F^2$ is examined by the cases with $\tilde{\omega} = 1$. It can be clearly seen that the best results of $\mathbf{Dict}_{\rm NEW}$ and $\mathbf{Dict}_{\rm SDB}$ are always achieved when $0 < \tilde{\omega} < 1$, which indicates that the weighting models (12) and (18) are valid. Besides, comparing the results of $\mathbf{Dict}_{\rm NEW}$ with those of $\mathbf{Dict}_{\rm SDB}$, the importance of the column-normalization constraints is revealed, as the satisfactory results of $\mathbf{Dict}_{\rm NEW}$ are much better than those of $\mathbf{Dict}_{\rm SDB}$. Though the optimum $\tilde{\omega}$ values are data-dependent and differ for various σ, hereinafter, $\tilde{\omega} = 0.2$, hence $\omega = 0.25$ is set for $\mathbf{Dict}_{\rm NEW}$ and $\mathbf{Dict}_{\rm SDB}$ according to its satisfactory performance for all the tests.



(a) $\sigma = 0.004$ (b) $\sigma = 0.010$ (c) $\sigma = 0.020$

Figure 1: Recovered SNR versus $\tilde{\omega}$.

5.1.2. Theoretical performance of different dictionaries

Before testing the denoising performance, we take a brief look at the singular value distribution of each learned dictionary. As indicated in Section 2, for a specific application, it is important to control over the spectrum of the corresponding system matrix and this spectrum is closely related to the singular value distribution of the system matrix. The more compact the singular value distribution is, the flatter the matrix spectrum will be, and this also implies smaller condition number [21], [22], [36], [37]. So a compact singular value distribution means a well-conditioned dictionary, leading to efficient sparse coding.

A UNTF of dimension $N \times K$ has all its singular values equal to $\sqrt{K/N}$ [34], [37]. Taking this as the benchmark, the singular value distributions of the learned dictionaries are shown in Figure 2.

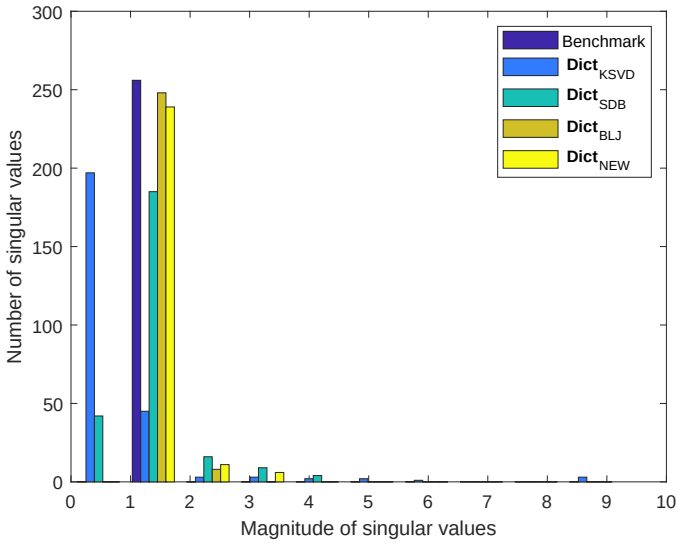


Figure 2: Singular value distributions of the learned dictionaries.

As can be seen, the distribution of $\mathbf{Dict}_{\mathrm{BLJ}}$ is closest to the benchmark $\sqrt{K/N}$ because $\mathbf{Dict}_{\mathrm{BLJ}}$ is designed under the constraint (15), which is exactly the expression of 1-tight frame. Though extra normalization is enforced, $\mathbf{Dict}_{\mathrm{BLJ}}$ is still the best-conditioned one. However, the constraint (15) limits the SR capability of the dictionary to signals since only two orthonormal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are optimized to minimize the SR error. Defined as the ratio of the maximum singular value and the minimum one, the condition number of each dictionary is calculated as cond($\mathbf{Dict}_{\mathrm{KSVD}}$)=2235.10, cond($\mathbf{Dict}_{\mathrm{SDB}}$)=17.12, cond($\mathbf{Dict}_{\mathrm{BLJ}}$)=1.89, cond($\mathbf{Dict}_{\mathrm{NEW}}$)=2.57. As indicated in [36], the closer the condition number to 1, the better the matrix stability will be. We can say that $\mathbf{Dict}_{\mathrm{NEW}}$ is also well conditioned. It can be expected that, if a large ω is set, cond($\mathbf{Dict}_{\mathrm{NEW}}$) will be much lower at the cost of the SR capability. So proper tradeoff is necessary in the proposed model (18).

5.1.3. Denoising performance on testing data

Now we test the denoising performance of the compared dictionaries with the testing data. Figure 3 depicts the output SNRs of the recovered signals when the Gaussian noises with standard deviation σ varying between 0.001 and 0.030 are added to the testing data.

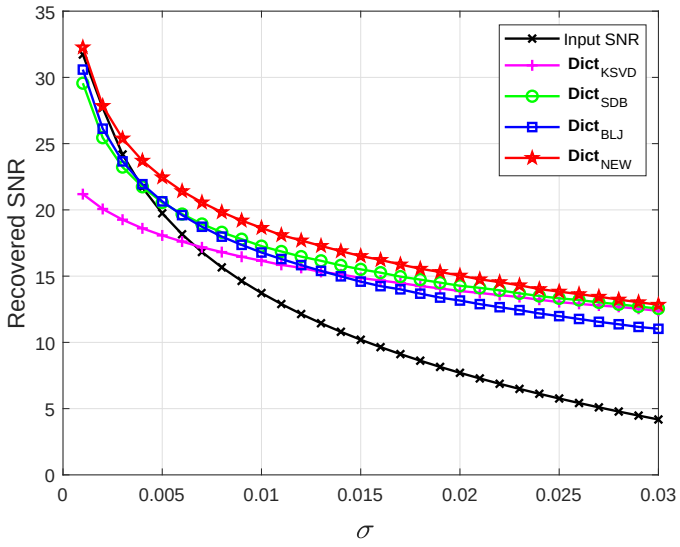


Figure 3: Recovered SNR results versus noise standard deviation σ.

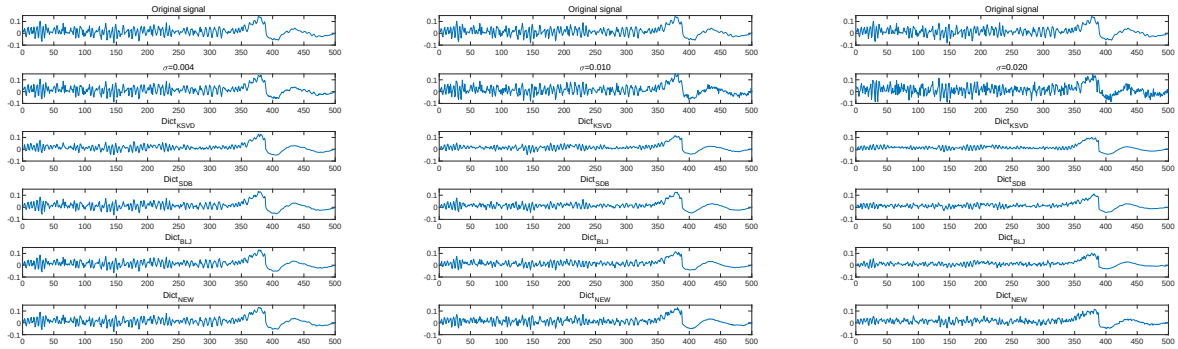
As shown in Figure 3, the recovered SNR values all decrease as σ increases, and $\mathbf{Dict}_{\mathrm{NEW}}$ achieves the best denoising results among these tests. The superiority of $\mathbf{Dict}_{\mathrm{NEW}}$ over $\mathbf{Dict}_{\mathrm{KSVD}}$ is clearly observed in low noise cases. When $\sigma \leq 0.012$, the gap is more than 2 dB. For high noise cases, e.g., $\sigma \geq 0.020$, the performance of $\mathbf{Dict}_{\mathrm{BLJ}}$ is poorer than the other three. So we conclude that neither $\mathbf{Dict}_{\mathrm{KSVD}}$ nor $\mathbf{Dict}_{\mathrm{BLJ}}$ is robust enough. The results of $\mathbf{Dict}_{\mathrm{SDB}}$ are close to those of $\mathbf{Dict}_{\mathrm{NEW}}$. This verifies the effectiveness of the weighting model which emphasizes the

tightness property of the dictionary. The remaining differences between these two can be explained by the column-normalization constraints that make the physical meanings of our model much clearer.

5.1.4. Denoising performance on speech segment

Besides the speech data, we carry out experiments on real speech segment for testing the denoising performance of different dictionaries. Phrase "place blue at P zero please" is randomly selected for testing. The speech segment is first divided into 143 non-overlapping time-domain frames with each frame of length $N = 256$. Denoising is executed in every frame independently by each of the learned dictionaries, and the restored frames synthesize a speech segment.

Figures 4 (a), (b) and (c) show the time-domain waveforms of the speech segments restored by different dictionaries when the standard deviations of the Gaussian noises are 0.004, 0.010 and 0.020, respectively. In order to clearly see the details, only 500 samples (starting from 5, 501st in 36, 608 samples) of each waveform are drawn for comparison.

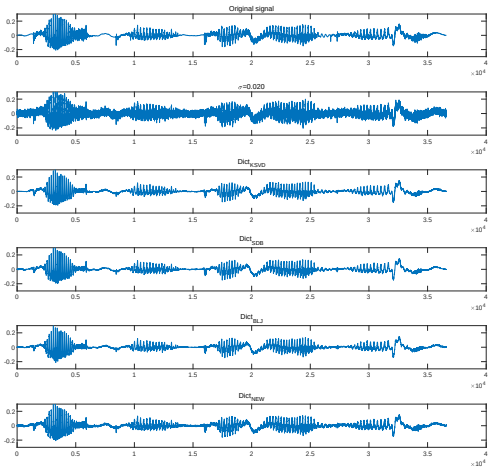


(a) $\sigma = 0.004$ (b) $\sigma = 0.010$ (c) $\sigma = 0.020$

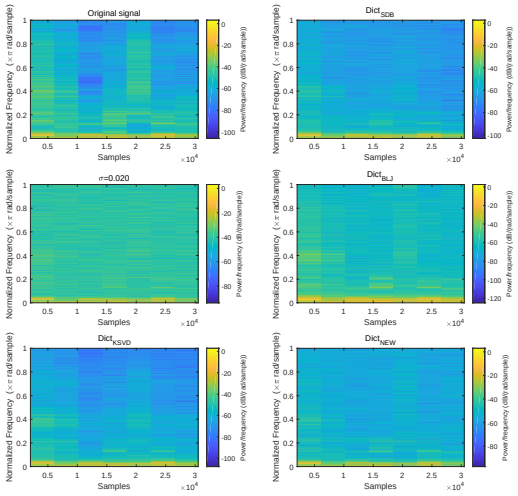
Figure 4: Restored speech waveforms by different dictionaries.

Taking a close look at these waveforms, one notes that $\mathbf{Dict}_{\mathrm{KSVD}}$ loses the details easily when the signals fluctuate acutely, even in the case with a small $\sigma = 0.004$, whereas the other three dictionaries perform favourably. When $\sigma = 0.020$, it can be observed that the waveform resulted from $\mathbf{Dict}_{\mathrm{BLJ}}$ is still heavily disturbed by noise. These results confirm that neither $\mathbf{Dict}_{\mathrm{KSVD}}$ nor $\mathbf{Dict}_{\mathrm{BLJ}}$ is robust enough.

The spectrograms of the restored speech segments are compared when $\sigma = 0.020$. Figures 5 (a) and (b) depict the complete speech waveforms and the corresponding spectrograms (drawn by MATLAB with default settings), respectively. The dark color parts (closing to blue) and the light color parts (closing to yellow) in Figure 5 (b) refer to low frequency and high frequency components, respectively. As can be observed, $\mathbf{Dict}_{\mathrm{KSVD}}$ loses a lot of high frequency information and $\mathbf{Dict}_{\mathrm{BLJ}}$ cannot eliminate noise well. These conclusions are consistent with those obtained from Figure 4.



(a) Restored speech waveforms with $\sigma = 0.020$



(b) Spectrograms of the speech segments

Figure 5: Restored speech waveforms and spectrograms with $\sigma = 0.020$.

We summarize the SNR results of all the restored speech segments in Table 1. The output SNR results also verify the above observations. Although the results of $\mathbf{Dict}_{\mathrm{SDB}}$ are acceptable, it should be emphasized that both the waveforms and the recovered SNR results clearly verify that $\mathbf{Dict}_{\mathrm{NEW}}$ outperforms other methods in all these tests.

Table 1: Recovered SNRs of speech segments by different dictionaries (dB)

.

	$\sigma = 0.004$	$\sigma = 0.010$	$\sigma = 0.020$
Input SNR	21.93	13.97	7.95
$\mathbf{Dict}_{\mathrm{KSVD}}$	19.59	17.02	14.86
$\mathbf{Dict}_{\mathrm{SDB}}$	21.83	17.65	14.89
$\mathbf{Dict}_{\mathrm{BLJ}}$	21.91	16.99	13.40
$\mathbf{Dict}_{\mathrm{NEW}}$	23.61	18.85	15.53

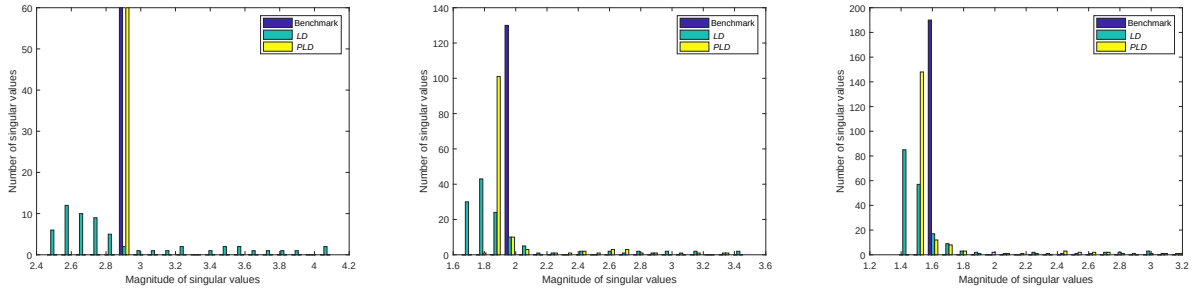
5.2. Performance of TSR-based speech inpainting

In this part, the learned tight dictionary $\boldsymbol{D}$ is applied to speech inpainting cases where the speech signals are corrupted by data missing and Gaussian noise. In order to understand the inpainting performance intuitively, real speech segments are considered. The inpainting process is carried out on non-overlapping time-domain frames with frame length $N = 256$. For each frame, data missing happens at random and the degradation matrix $\boldsymbol{L} \in \Re^{M\times N}$ that

records the data missing process is structured as (37). It means that $N - M$ samples are lost in every frame. When designing the tight preconditioner $\boldsymbol{P}$, the gradient descent is executed for $1,000$ times with step size $\eta = 0.01$.

5.2.1. Theoretical performance of tight preconditioner

First, the theoretical performance of the constructed preconditioner $\boldsymbol{P}$ is illustrated in Figure 6 by the singular value distribution of the preconditioned matrix $\boldsymbol{PLD}$. For comparison, the distribution of $\boldsymbol{LD}$ is also attached. Taking $\sqrt{K/M}$ (i.e., the spectrum of the $M \times K$ UNTF [34], [37]) as the benchmark, the singular value distributions are shown in Figures 6 (a), (b) and (c) with $M = 60$, $M = 130$ and $M = 190$, respectively.



(a) $M = 60$ (b) $M = 130$ (c) $M = 190$

Figure 6: Singular value distributions of corresponding system matrices.

In Figure 6 (a), the preconditioner $\boldsymbol{P}$ makes the preconditioned matrix $\boldsymbol{PLD}$ an exact UNTF when $M = 60$. This verifies the validity of our parametrization process (45) and the gradient-based approach for solving (42). Though similar phenomenon is not observed in Figures 6 (b) and (c), a more compact singular value distribution is expected if the step size η in (48) is chosen more carefully and intelligently. This would be explored in the future work.

For these three values of M, the condition numbers of the corresponding system matrices are summarized in Table 2. The results further confirm the effectiveness of the designed preconditioner. For these three values of M, the condition numbers of $\boldsymbol{PLD}$ are much closer to 1 than those of $\boldsymbol{LD}$. They lead to stable preconditioned matrices that facilitate the sparse recovery process, i.e., solving the inverse problem (41).

Table 2: Condition numbers of corresponding system matrices

.

	$M = 60$	$M = 130$	$M = 190$
cond($\boldsymbol{LD}$)	1.68	2.11	2.34
cond($\boldsymbol{PLD}$)	1.00	1.89	2.18

5.2.2. Inpainting performance evaluation

In this part, the speech inpainting problem is investigated and three processes are compared. The speech inpainting process executed by (39) is a TSR-based strategy that is denoted as $\mathbf{SI}_{\mathrm{TSR}}$. (41) is a preconditioned sparse recovery process and we name it as $\mathbf{SI}_{\mathrm{PTSR}}$. In addition, $\mathbf{Dict}_{\mathrm{NEW}}$ is applied to eliminate noise without data missing just as in the denoising part in Section 5.1.4 for comparison. In each of the tests, the input SNR is calculated prior to data missing.

The same phrase "place blue at P zero please" is chosen for testing as in the denoising cases. When $M = 130$, the recovered SNR values are depicted in Figure 7 for σ varying between 0.001 and 0.030. The results verify the validity of the TSR-based inpainting. Figure 7 shows that, when $\sigma > 0.008$, the results of $\mathbf{SI}_{\mathrm{TSR}}$ are higher than the input SNR (without data missing taken into account) values. It means that the inpainting strategy not only restores missing data, but also eliminates additive noises.

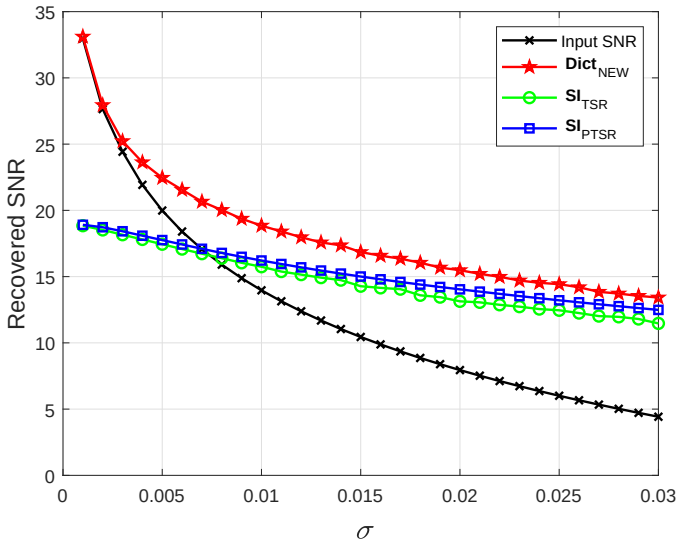


Figure 7: Recovered SNR results versus σ with $M = 130$.

Next, we evaluate the effects of the number of observations. For $\sigma = 0.010$ fixed, when M changes between 60 and 190, the SNR results are shown in Figure 8. As can be seen, even when only $M = 80$ observations are reserved, $\mathbf{SI}_{\mathrm{TSR}}$ restores the signals and leads to an SNR result that is comparable with the input SNR. More observations will result in remarkable improvements.

With two more speech segments tested, the statistics of the output SNR for these three phrases in different σ and M cases are summarized in Table 3. In all the tests, $\mathbf{SI}_{\mathrm{PTSR}}$ achieves higher SNR results than those of $\mathbf{SI}_{\mathrm{TSR}}$. This indicates that the preconditioning technique employed to facilitate the sparse recovery process is effective, and the tight preconditioner constructed by our method is superior.

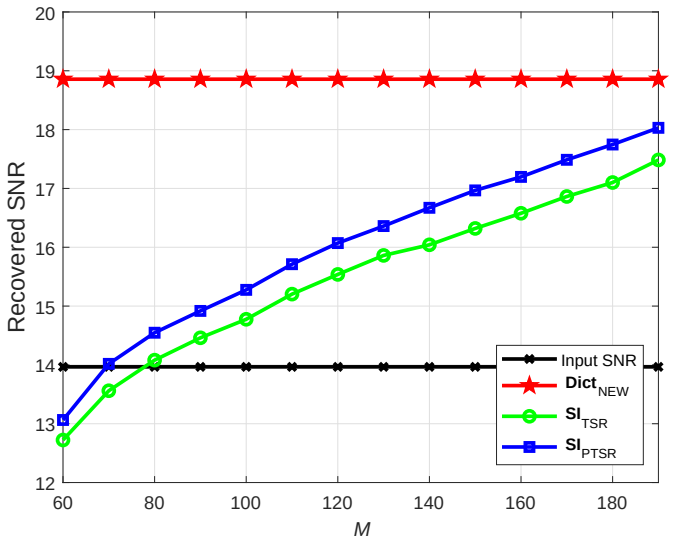


Figure 8: Recovered SNR results versus M with $\sigma = 0.010$.

Table 3: Statistics of SNR for three phrases in different σ and M cases (dB)

Phrases	Processes	$\sigma = 0.004$			$\sigma = 0.010$			$\sigma = 0.020$		
		$M = 60$	$M = 130$	$M = 190$	$M = 60$	$M = 130$	$M = 190$	$M = 60$	$M = 130$	$M = 190$
place blue at P zero please	Input SNR	21.93			13.97			7.95		
	Dict$_{\text{NEW}}$	23.66			18.95			15.60		
	SI$_{\text{TSR}}$	13.76	17.59	20.39	12.73	15.89	17.48	10.32	13.48	14.74
	SI$_{\text{PTSR}}$	14.07	17.93	20.86	13.08	16.38	18.04	10.70	14.16	15.59
bin blue at L two please	Input SNR	21.94			13.98			7.96		
	Dict$_{\text{NEW}}$	24.39			19.54			16.03		
	SI$_{\text{TSR}}$	16.04	19.08	21.07	13.95	16.59	18.15	10.47	14.12	15.26
	SI$_{\text{PTSR}}$	16.34	19.39	21.43	14.25	17.08	18.73	10.82	14.81	16.34
set white with I nine soon	Input SNR	21.92			13.97			7.95		
	Dict$_{\text{NEW}}$	23.73			18.31			14.53		
	SI$_{\text{TSR}}$	12.31	16.68	19.83	8.78	14.26	16.35	7.46	11.52	13.19
	SI$_{\text{PTSR}}$	12.57	17.07	20.38	8.99	14.87	17.20	7.80	12.23	14.15

6. Conclusion

In this paper, we have investigated the TSR problem with application to speech inpainting. A novel DL model, named TDL, was proposed. Under this model, the dictionary is learned to minimize the SR error adapted to the training signals and, at the same time, approximate the UNTF to gain the desired frame properties. The parametrization

method was adopted to embed the normalization constraints on the dictionary columns into the cost function. A gradient-based approach was developed to solve the parameterized problem for the tight dictionary. Moreover, we have applied the tight dictionary to the speech inpainting problem where data missing is considered. Finally, in order to maintain the frame properties of the system, the preconditioning technique based on UNTF was employed, and the parametrization and gradient-based approaches were carried out once again to solve the normalization constrained tight preconditioner design problem. Extensive experiments have been conducted to demonstrate the validity and effectiveness of the proposed TSR schemes on speech denoising and speech segments restoration.

Acknowledgement

This work is supported in part by the Natural Science Foundation of China under Grants 61801159 and 61571174, in part by the Key R&D Program Projects in Zhejiang Province under Grant 2020C03074, in part by the Start-up Research Program from Hangzhou Normal University under Grant 2018QDL018, and in part by the China-Montenegro Bilateral Science & Technology Cooperation Project.

Declaration of interest

The authors declare that they do not have any financial or non-financial conflict of interests.

Appendix A. Proof of Theorem 1

First, one notes that

$$\|\boldsymbol{E}_1\|_F^2 \;=\; \mathrm{tr}\left[\boldsymbol{E}_1\boldsymbol{E}_1^{\mathcal{T}}\right]$$

with tr[·] denoting the trace operator.

Denote $q_{ij} = \boldsymbol{Q}(i,j)$, then

$$\frac{\partial(\|\boldsymbol{E}_1\|_F^2)}{\partial q_{ij}} \;=\; 2\mathrm{tr}\left[\frac{\partial \boldsymbol{E}_1}{\partial q_{ij}}\boldsymbol{E}_1^{\mathcal{T}}\right],$$

and

$$\frac{\partial \boldsymbol{E}_1}{\partial q_{ij}} \;=\; -\frac{\partial(\boldsymbol{Q}\boldsymbol{Q}_d)}{\partial q_{ij}}\boldsymbol{S}.$$

Given the definition of $\boldsymbol{Q}_d$ in (24), it can be shown that

$$\frac{\partial(\boldsymbol{Q}\boldsymbol{Q}_d)}{\partial q_{ij}} \;=\; \boldsymbol{e}_i\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}_d - \boldsymbol{Q}q_{ij}d_j^3\boldsymbol{e}_j\boldsymbol{e}_j^{\mathcal{T}},$$

where $\boldsymbol{e}_k$ denotes the k-th elementary vector with a proper dimension, whose elements are all zeros except the k-th one that is 1. For example, in the above expression, the dimensions of $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are respectively $N \times 1$ and $K \times 1$. $d_j \triangleq (\sum_i q_{ij}^2)^{-1/2} = \|\boldsymbol{Q}(:,j)\|_2^{-1}$ is the j-th diagonal element of $\boldsymbol{Q}_d$. Then, we get

$$\frac{\partial \boldsymbol{E}_1}{\partial q_{ij}} \;=\; -\boldsymbol{e}_i\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}_d\boldsymbol{S} + q_{ij}d_j^3\boldsymbol{Q}\boldsymbol{e}_j\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{S}$$

and

$$\frac{\partial(\|\boldsymbol{E}_1\|_F^2)}{\partial q_{ij}} \;=\; 2(-\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}_d\boldsymbol{S}\boldsymbol{E}_1^{\mathcal{T}}\boldsymbol{e}_i + q_{ij}d_j^3\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{S}\boldsymbol{E}_1^{\mathcal{T}}\boldsymbol{Q}\boldsymbol{e}_j).$$

As $q_{ij} = \boldsymbol{Q}(i,j)$, i.e., $q_{ij} = \boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{Q}\boldsymbol{e}_j$, we finally obtain

$$\frac{\partial(\|\boldsymbol{E}_1\|_F^2)}{\partial \boldsymbol{Q}} \;=\; 2\boldsymbol{Q}\boldsymbol{Q}_d^3\boldsymbol{R}_1 - 2\boldsymbol{E}_1\boldsymbol{S}^{\mathcal{T}}\boldsymbol{Q}_d$$

with $\boldsymbol{R}_1$ defined in (32). This completes the proof.

Appendix B. Proof of Theorem 2

Similar to Appendix A, we have

$$\frac{\partial(\|\boldsymbol{E}_2\|_F^2)}{\partial q_{ij}} \;=\; 2\mathrm{tr}\left[\frac{\partial \boldsymbol{E}_2}{\partial q_{ij}}\boldsymbol{E}_2^{\mathcal{T}}\right]$$

and

$$\frac{\partial \boldsymbol{E}_2}{\partial q_{ij}} = \frac{\partial(\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}}{\partial q_{ij}}(\boldsymbol{Q}\boldsymbol{Q}_d)+(\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}\frac{\partial(\boldsymbol{Q}\boldsymbol{Q}_d)}{\partial q_{ij}} = \boldsymbol{Q}_d\boldsymbol{e}_j\boldsymbol{e}_i^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)-q_{ij}d_j^3\boldsymbol{e}_j\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)+(\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}\boldsymbol{e}_i\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}_d-q_{ij}d_j^3(\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}\boldsymbol{Q}\boldsymbol{e}_j\boldsymbol{e}_j^{\mathcal{T}}$$

as $\boldsymbol{Q}_d$ is diagonal. Hence

$$\frac{\partial(\|\boldsymbol{E}_2\|_F^2)}{\partial q_{ij}} \;=\; 2\left[\boldsymbol{e}_i^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)\boldsymbol{E}_2^{\mathcal{T}}\boldsymbol{Q}_d\boldsymbol{e}_j - q_{ij}d_j^3\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)\boldsymbol{E}_2^{\mathcal{T}}\boldsymbol{e}_j + \boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{Q}_d\boldsymbol{E}_2^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}\boldsymbol{e}_i - q_{ij}d_j^3\boldsymbol{e}_j^{\mathcal{T}}\boldsymbol{E}_2^{\mathcal{T}}(\boldsymbol{Q}\boldsymbol{Q}_d)^{\mathcal{T}}\boldsymbol{Q}\boldsymbol{e}_j\right].$$

Then, we obtain

$$\frac{\partial(\|\boldsymbol{E}_2\|_F^2)}{\partial \boldsymbol{Q}} \;=\; 2(\boldsymbol{Q}\boldsymbol{Q}_d)(\boldsymbol{E}_2^{\mathcal{T}} + \boldsymbol{E}_2)\boldsymbol{Q}_d - 2\boldsymbol{Q}\boldsymbol{Q}_d^3\boldsymbol{R}_2$$

with $\boldsymbol{R}_2$ defined in (35). This completes the proof.

Appendix C. Proof of Theorem 3

Because

$$\|\tilde{\boldsymbol{E}}\|_F^2 \;=\; \mathrm{tr}\left[\tilde{\boldsymbol{E}}\tilde{\boldsymbol{E}}^{\mathcal{T}}\right],$$

we obtain

$$\frac{\partial(\|\tilde{\boldsymbol{E}}\|_F^2)}{\partial p_{ij}} \;=\; 2\mathrm{tr}\left[\frac{\partial\tilde{\boldsymbol{E}}}{\partial p_{ij}}\tilde{\boldsymbol{E}}^{\mathcal{T}}\right]$$

and

$$\frac{\partial\tilde{\boldsymbol{E}}}{\partial p_{ij}} \;=\; \frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}}{\partial p_{ij}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d) + (\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)}{\partial p_{ij}}.$$

Let $K\times K$ matrix $\boldsymbol{\Xi}_k \triangleq \boldsymbol{e}_k\boldsymbol{e}_k^{\mathcal{T}}$. Denote

$$\tilde{d}_k \;=\; \|\tilde{\boldsymbol{Q}}(:,k)\|_2^{-1} \;=\; \|\boldsymbol{P}\tilde{\boldsymbol{A}}(:,k)\|_2^{-1}$$

as the k-th diagonal element of $\tilde{\boldsymbol{Q}}_d$. Then $\tilde{\boldsymbol{Q}}_d$ can be expressed as

$$\tilde{\boldsymbol{Q}}_d \;=\; \sum_{k=1}^{K}\tilde{d}_k\boldsymbol{\Xi}_k.$$

Denote $\boldsymbol{a}_k = \tilde{\boldsymbol{A}}(:,k)$ as the k-th column of $\tilde{\boldsymbol{A}}$. One can derive that

$$\frac{\partial\tilde{\boldsymbol{Q}}_d}{\partial p_{ij}} \;=\; -\sum_{k=1}^{K}\tilde{d}_k^3\boldsymbol{a}_k^{\mathcal{T}}\boldsymbol{e}_j\boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{a}_k\boldsymbol{\Xi}_k$$

with $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ both elementary vectors of dimension $M\times 1$. Hence

$$\frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)}{\partial p_{ij}} \;=\; \frac{\partial\tilde{\boldsymbol{Q}}}{\partial p_{ij}}\tilde{\boldsymbol{Q}}_d + \tilde{\boldsymbol{Q}}\frac{\partial\tilde{\boldsymbol{Q}}_d}{\partial p_{ij}} \;=\; \boldsymbol{e}_i\boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d - \tilde{\boldsymbol{Q}}\sum_{k=1}^{K}\tilde{d}_k^3\boldsymbol{a}_k^{\mathcal{T}}\boldsymbol{e}_j\boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{a}_k\boldsymbol{\Xi}_k.$$

Therefore,

$$\begin{aligned}
&\mathrm{tr}\left[(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)}{\partial p_{ij}}\tilde{\boldsymbol{E}}^{\mathcal{T}}\right]\\
&= \mathrm{tr}\left[(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}(\boldsymbol{e}_i\boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d - \tilde{\boldsymbol{Q}}\sum_{k=1}^{K}\tilde{d}_k^3\boldsymbol{a}_k^{\mathcal{T}}\boldsymbol{e}_j\boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{a}_k\boldsymbol{\Xi}_k)\tilde{\boldsymbol{E}}^{\mathcal{T}}\right]\\
&= \boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{E}}^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\boldsymbol{e}_i - \sum_{k=1}^{K}\boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{a}_k\tilde{d}_k^3\mathrm{tr}\left[(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\tilde{\boldsymbol{Q}}\boldsymbol{\Xi}_k\tilde{\boldsymbol{E}}^{\mathcal{T}}\right]\boldsymbol{a}_k^{\mathcal{T}}\boldsymbol{e}_j\\
&= \boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{E}}^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\boldsymbol{e}_i - \sum_{k=1}^{K}\boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{a}_k\tilde{d}_k^3\left[\boldsymbol{e}_k^{\mathcal{T}}\tilde{\boldsymbol{E}}^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\tilde{\boldsymbol{Q}}\boldsymbol{e}_k\right]\boldsymbol{a}_k^{\mathcal{T}}\boldsymbol{e}_j\\
&\triangleq \boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{E}}^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\boldsymbol{e}_i - \boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\sum_{k=1}^{K}\boldsymbol{a}_k\tilde{d}_k^3\tilde{\boldsymbol{H}}_1(k,k)\boldsymbol{a}_k^{\mathcal{T}}\boldsymbol{e}_j\\
&\triangleq \boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{E}}^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\boldsymbol{e}_i - \boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{G}_1\boldsymbol{e}_j,
\end{aligned}$$

where

$$\tilde{\boldsymbol{H}}_1 \triangleq \tilde{\boldsymbol{E}}^{\mathcal{T}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\tilde{\boldsymbol{Q}}, \quad \boldsymbol{G}_1 \triangleq \sum_{k=1}^{K} \tilde{d}_k^3 \tilde{\boldsymbol{H}}_1(k,k)\boldsymbol{a}_k\boldsymbol{a}_k^{\mathcal{T}}.$$

According to the characteristics of trace operation, one has

$$\mathrm{tr}\left[\frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}}{\partial p_{ij}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)\tilde{\boldsymbol{E}}^{\mathcal{T}}\right] = \mathrm{tr}\left[(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)}{\partial p_{ij}}\tilde{\boldsymbol{E}}\right].$$

Therefore,

$$\mathrm{tr}\left[\frac{\partial(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}}{\partial p_{ij}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)\tilde{\boldsymbol{E}}^{\mathcal{T}}\right] = \boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{E}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\boldsymbol{e}_i - \boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}\boldsymbol{G}_2\boldsymbol{e}_j,$$

where

$$\boldsymbol{G}_2 \triangleq \sum_{k=1}^{K} \tilde{d}_k^3 \tilde{\boldsymbol{H}}_2(k,k)\boldsymbol{a}_k\boldsymbol{a}_k^{\mathcal{T}}$$

with

$$\tilde{\boldsymbol{H}}_2 \triangleq \tilde{\boldsymbol{E}}(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\tilde{\boldsymbol{Q}}.$$

Subsequently,

$$\frac{\partial(\|\tilde{\boldsymbol{E}}\|_F^2)}{\partial p_{ij}} = 2\mathrm{tr}\left[\frac{\partial\tilde{\boldsymbol{E}}}{\partial p_{ij}}\tilde{\boldsymbol{E}}^{\mathcal{T}}\right] = 2\left[\boldsymbol{e}_j^{\mathcal{T}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Q}}_d(\tilde{\boldsymbol{E}}^{\mathcal{T}} + \tilde{\boldsymbol{E}})(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)^{\mathcal{T}}\boldsymbol{e}_i - \boldsymbol{e}_i^{\mathcal{T}}\boldsymbol{P}(\boldsymbol{G}_1 + \boldsymbol{G}_2)\boldsymbol{e}_j\right],$$

which leads to

$$\frac{\partial(\|\tilde{\boldsymbol{E}}\|_F^2)}{\partial \boldsymbol{P}} = 2(\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{Q}}_d)(\tilde{\boldsymbol{E}}^{\mathcal{T}} + \tilde{\boldsymbol{E}})\tilde{\boldsymbol{Q}}_d\tilde{\boldsymbol{A}}^{\mathcal{T}} - 2\boldsymbol{P}\boldsymbol{G},$$

where

$$\boldsymbol{G} \triangleq \boldsymbol{G}_1 + \boldsymbol{G}_2 = \sum_{k=1}^{K} \tilde{d}_k^3 \left[\tilde{\boldsymbol{H}}_1(k,k) + \tilde{\boldsymbol{H}}_2(k,k)\right]\boldsymbol{a}_k\boldsymbol{a}_k^{\mathcal{T}}.$$

Using the definition of $\tilde{\boldsymbol{R}}$ in (51), $\boldsymbol{G}$ can be calculated as $\boldsymbol{G} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{R}}\tilde{\boldsymbol{A}}^{\mathcal{T}}$. This completes the proof.

References

[1] P. G. Casazza, G. Kutyniok, and F. Philipp, "Introduction to finite frame theory." In: P. G. Casazza, G. Kutyniok (eds.), *Finite Frames: Theory and Applications*, Applied and Numerical Harmonic Analysis, pp. 1-53, Birkhäuser, 2013.

[2] O. Christensen, *An Introduction to Frames and Riesz Bases (Second Edition)*, Birkhäuser, 2016.

[3] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34-81, Mar. 2009.

[4] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[5] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[6] J. Karmakar, A. Pathak, D. Nandi, and M. K. Mandal, "Sparse representation based compressive video encryption using hyper-chaos and DNA coding," *Digit. Signal Process.*, vol. 117, Article 103143, Oct. 2021.

[7] S. Ding, G. Zhao, C. Liberatore, and R. Gutierrez-Osuna, "Learning structured sparse representations for voice conversion," *IEEE-ACM Trans. Audio Speech Lang.*, vol. 28, pp. 343-354, 2020.

[8] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, and L. Chang, "Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1581-1594, Mar. 2015.

[9] I. Tošić and P. Frossard, "Dictionary learning: What is the right representation for my signal?," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27-38, Mar. 2011.

[10] K. Engan, S. O. Aase, and J. Håkon-Husøy, "Method of optimal directions for frame design," in *1999 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP99)*, Phoenix, USA, pp. 2443-2446, Mar. 1999.

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

[12] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.

[13] J. A. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[14] Q. Jiang, S. Li, Z. Zhu, H. Bai, X. He, and R. C. de Lamare, "Design of compressed sensing system with probability-based prior information," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 594-609, Mar. 2020.

[15] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[16] M. S. Asif and J. Romberg, "Sparse recovery of streaming signals using ℓ_1-Homotopy," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4209-4223, Aug. 2014.

[17] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "On the use of unit-norm tight frames to improve the average MSE performance in compressive sensing applications," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 8-11, Jan. 2012.

[18] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016-2029, Apr. 2013.

[19] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when *p* is much larger than *n*," *Ann. Stat.*, vol. 35, no. 6, pp. 2313-2351, 2007.

[20] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384-3389, Jun. 2010.

[21] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Part I)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86-104, Jul. 2007.

[22] J. J. Benedetto and M. Fickus, "Finite normalized tight frames," *Adv. Comput. Math.*, vol. 18, pp. 357-385, Feb. 2003.

[23] P. G. Casazza and J. Kovačević, "Equal-norm tight frames with erasures," *Adv. Comput. Math.*, vol. 18, pp. 387-430, Feb. 2003.

[24] P. Zhang, L. Gan, S. Sun, and C. Ling, "Modulated unit-norm tight frames for compressed sensing," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3974-3985, Aug. 2015.

[25] R. R. Naidu and C. R. Murthy, "Construction of unimodular tight frames for compressed sensing using majorization-minimization," *Signal Process.*, vol. 172, Article 107516, Jul. 2020.

[26] D. G. Mixon, "Unit norm tight frames in finite-dimensional spaces." In: K. A. Okoudjou (eds.), *Finite Frame Theory: A Complete Introduction to Overcompleteness*, Proceedings of Symposia in Applied Mathematics, AMS Short Course, San Antonio, Texas, Jan. 2015.

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

[28] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 3, pp. 922-932, Mar. 2012.

[29] F. Lieb and H.-G. Stark, "Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches," *Signal Process.*, vol. 153, pp. 291-299, Dec. 2018.

[30] O. Mokrý and P. Rajmic, "Audio inpainting: Revisited and reweighted," *IEEE-ACM Trans. Audio Speech Lang.*, vol. 28, pp. 2906-2918, Oct. 2020.

[31] M. Benzi, "Preconditioning techniques for large linear systems: A survey," *J. Comput. Phys.*, vol. 182, no. 2, pp. 418-477, Nov. 2002.

[32] E. Tsiligianni, L. P. Kondi, and A. K. Katsaggelos, "Preconditioning for underdetermined linear systems with sparse solutions," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1239-1243, Sept. 2015.

[33] H. Bai, C. Hong, and X. Li, "Construction of unit-norm tight frame based preconditioner for sparse coding," in *2021 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2021)*, Toronto, Canada, pp. 5400-5404, Jun. 2021.

[34] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887-2898, Jun. 2013.

[35] H. Bai, S. Li, and X. He, "Sensing matrix optimization based on equiangular tight frames with consideration of sparse representation error," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2040-2053, Oct. 2016.

[36] R. A. Horn and C. R. Johnson, *Matrix Analysis (second edition)*, Cambridge University Press, 2013.

[37] C. D. Sigg, T. Dikk, and J. M. Buhmann, "Learning dictionaries with bounded self-coherence," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 861-864, Dec. 2012.

[38] H. Bai, S. Li, and Q. Jiang, "An efficient algorithm for learning dictionary under coherence constraint," *Math. Probl. Eng.*, Article ID 5737381, 11 pages, 2016.

[39] M. Gevers and G. Li, *Parametrizations in control, estimation and filtering problems: Accuracy aspects, communication and control engineering series*, Springer, 1993.

[40] B. G. Bodmann and P. G. Casazza, "The road to equal-norm Parseval frames," *J. Funct. Anal.*, vol. 258, no. 2, pp. 397-420, Jan. 2010.

[41] P. G. Casazza, M. Fickus, and D. G. Mixon, "Auto-tuning unit norm frames," *Appl. Comput. Harmon. Anal.*, vol. 32, no. 1, pp. 1-15, Jan. 2012.

[42] X. Li, H. Bai, and B. Hou, "A gradient-based approach to optimization of compressed sensing systems," *Signal Process.*, vol. 139, pp. 49-61, Oct. 2017.

[43] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2421-2424, Nov. 2006.