# Combined Regularization Optimization For Separating Transient Signals From Strong Noise In Rolling Element Bearing Diagnosis

Liang Yu, Jerome Antoni, Quentin Leclere

Laboratory Vibrations and Acoustics, INSA-Lyon
Batiment St. Exupery 25 bis av. Jean Capelle, 69621 Villeurbanne cedex - FRANCE
{liang.yu, jerome.antoni, quentin.leclere}@insa-lyon.fr

## Abstract

The problem of rolling element bearing diagnosis can be viewed as that of extracting transient signals from strong additive noise. In order to separate these two kinds of signals, the problem is formulated as the minimization of an objective function which consists of two terms, a data fitting term, and a regularization term. The data fitting term reflects the conservation of energy and the regularization term reflects the prior information (initial supposition) on the objective signals to be separated. Since the objective is to separate two kinds of signals, a combined regularization scheme (distinct priori information for each signal with a target for separation) is naturally investigated. In this paper, two kinds of strategies are considered : the first one is a combination of two sparse regularization terms, and the second one is a combination of a low rank regularization term and a sparse regularization term.

## 1 Introduction

Rolling element bearings are widely used in rotating machines and their failure is one of most frequent reason for machine breakdown [1]. Spectral analysis has for long been used as a basic tool for the task of detection and diagnosis ; nevertheless, being a tool dedicated to stationary signals, it is unable to capture changes in time. Time-frequency analysis has been widely used for rolling element bearing diagnosis [2][3][4] due to its good 'localization' capability (Short Time Fourier Transform (STFT) [5]), 'Multi-Resolution' (wavelet [4]), 'Adaptivity (wavelet pack [6])', which is useful for extracting machinery fault information from measurements. More recently, dictionary (dictionary provides a more flexible representation as compared with orthogonal transforms of signal [7]) based sparse representation was brought in rolling element bearing diagnosis [8][9] ; the principle is to select a small subset of functions (bases) from a general over-complete dictionary which provides a more flexible decomposition of the signal. In this paper, combined regularization structures (in contrast to single regularization in classical sparse representation model) are investigated for rolling element bearing diagnosis which resort to obtaining a better separation result. An overview of this paper is given as follows.

In order to separate two kinds of signals $\mathbf{x} = \mathbf{Ac_1}$ and $\mathbf{y} = \mathbf{Bc_2}$ from their mixture signal $\mathbf{z} = \mathbf{x} + \mathbf{y}$, where $\mathbf{A} \in R^{N\times N1}$, $\mathbf{B} \in R^{N\times N2}$ and $\mathbf{c_1} \in R^{N1}$, $\mathbf{c_2} \in R^{N2}$, $\mathbf{x} \in R^{N}$, $\mathbf{y} \in R^{N}$, $\mathbf{z} \in R^{N}$, a general constrained optimization problem (a similar and generalized structure can be found in mathematical optimization literature [10] [11] [12]) is investigated in this paper :

$$\begin{aligned} &\text{Minimize } \; Reg_1(\mathbf{c_1}) + Reg_2(\mathbf{c_2}) \\ &\text{Subject to } \; \mathbf{Ac_1} + \mathbf{Bc_2} = \mathbf{z} \end{aligned} \tag{1}$$

where $\mathbf{Ac_1} + \mathbf{Bc_2} = \mathbf{z}$ is the data fitting term, which indicates the conservation of energy, and $Reg_1(\mathbf{c_1}) + Reg_2(\mathbf{c_2})$ is the regularization term, which reflects prior information (initial supposition) assigned to each signal to be separated. The specific forms of the prior $Reg_1(\mathbf{c_1})$ and $Reg_2(\mathbf{c_2})$ are elaborated in the following parts.

More specifically, in order to separate a transient signal $\mathbf{x} \in R^N$ from strong additive stationary noise $\mathbf{y} \in R^N$ in measurement $\mathbf{z} \in R^N$, two kinds of combined regularization forms are investigated :

- Sparse + Sparse Combination : it is assumed that $\mathbf{x}$ can be represented in an over-complete dictionary $\Psi$ as $\mathbf{x}^{N\times1} = \Psi^{N\times M}\alpha^{M\times1}$, where $M \gg N$ and $\alpha^{M\times1}$ is a sparse vector, while $\tilde{\alpha}$ is not sparse when $\mathbf{y}^{N\times1}$ is decomposed as $\mathbf{y}^{N\times1} = \Psi^{N\times M}\tilde{\alpha}^{M\times1}$. Similarly, $\mathbf{y}^{N\times1}$ can be represented by one over-complete dictionary $\Phi$ as $\mathbf{y}^{N\times1} = \Phi^{N\times M}\beta^{M\times1}$, where $M \gg N$ and $\beta^{M\times1}$ is a sparse vector, while, $\tilde{\beta}$ is not sparse when $\mathbf{x}^{N\times1}$ is decomposed as $\mathbf{x}^{N\times1} = \Phi^{N\times M}\tilde{\beta}^{M\times1}$. The signal separation problem is then formulated as :

$$\begin{aligned} &\text{Minimize}_{\alpha,\beta} \quad \|\alpha\|_{L_0} + \lambda \|\beta\|_{L_0} \\ &\text{Subject to} \;\; \mathbf{z}^{N\times1} = \Psi^{N\times M}\alpha^{M\times1} + \Phi^{N\times M}\beta^{M\times1} \end{aligned} \tag{2}$$

  where $L_0$ is defined as total number of non-zero elements in a vector [13], $\lambda$ is a weight which serves the purpose of adding a trade-off between $\|\alpha\|_{L_0}$ and $\|\beta\|_{L_0}$ ; model (2) is named Morphological Component Analysis (MCA) in [14][15]. Unfortunately, the noise $\mathbf{y}^{N\times1}$ can not be represented parsimoniously under one dictionary (term $\lambda\|\beta\|_{L_0}$ disappears), thus, in this paper, one modified model is proposed for separating the transient signal from strong additive noise :

$$\begin{aligned} &\text{Minimize}_{\;\alpha} \quad \|\alpha\|_{L_0} \\ &\text{Subject to} \;\; \mathbf{z}^{N\times1} = \Psi^{N\times M}\alpha^{M\times1} + \mathbf{y}^{N\times1} \\ &\qquad\qquad \frac{\|\Psi\alpha\|_{L_2}}{\|\mathbf{y}\|_{L_2}} = c \end{aligned} \tag{3}$$

  where $L_2$ denotes the Euclidean norm, and the small constant $c$ denotes the ratio between the transient signal and noise energies. It is noted that $\lambda$ is a implicit ratio parameter in model (2) and $c$ is a explicit parameter in model (3) ; -more details are argued in section 3.2 .
- Low rank + Sparse Combination : Spectrogram is defined as the operation of calculating signals spectrogram matrix. For signal $\mathbf{z} \in R^N$, $Spectrogram\{\mathbf{z}^{N\times1}\} = Spectrogram\{\mathbf{x}^{N\times1}\} + Spectrogram\{\mathbf{y}^{N\times1}\}$ can be represented as $\mathbf{M}^{P\times Q} = \mathbf{L}^{P\times Q} + \mathbf{S}^{P\times Q}$. Where $\mathbf{L}^{P\times Q}$ corresponds to the stationary noise signal and is low rank, and $\mathbf{S}^{P\times Q}$ corresponds to the transient signal and is sparse. The signal separation problem is then formulated as :

$$\begin{aligned} &\text{Minimize} \quad \text{Rank}(\mathbf{L}) + \lambda\|\mathbf{S}\|_{L0} \\ &\text{Subject to} \;\; \mathbf{L} + \mathbf{S} = \mathbf{M}. \end{aligned} \tag{4}$$

It is evident that model (4) is similar to form (1), where $Ac_1 = \mathbf{L}$ , $Ac_2 = \mathbf{S}$ and $Reg_1(c_1) = Reg_1(\mathbf{L})$, $Reg_2(c_2) = Reg_2(\mathbf{S})$. The paper is organized as follows. The problem of separating a transient signal from strong additive noise is formally stated in section 2. The Sparse + Sparse combination and Low Rank + Sparse combination models are separately discussed in section 3 and section 4. Sections 5 and 6 presents simulation results and draws the conclusion respectively.

## 2 Problem Statement

The problem of separating a transient signal $\mathbf{x} \in R^N$ from strong additive noise $\mathbf{y} \in R^N$ in measurement $\mathbf{z} \in R^N$ is described as

$$\mathbf{z} = \mathbf{x} + \mathbf{y} \tag{5}$$

In this work, the transient signal $\mathbf{x}$ is modelled as a generalised shot noise process $\mathbf{x}(k) = \sum_k (X_k * h(k - \tau_k))$, $k = 1...N$, where $h(k)$ is the impulse response resulting from a single impact and $\{X_k\}$ and $\{\tau_k\}$ are sequences of random variables which account for possibly random amplitudes and random occurrences of the impacts [3]. Noise $\mathbf{y}$ is assumed to be of zero mean and wide-sense stationary. A simple illustrative example is shown in Fig. 1, where the transient signal (in green color) is merged into strong additive noise (in blue color) : the challenge is to separate the latter from the former.

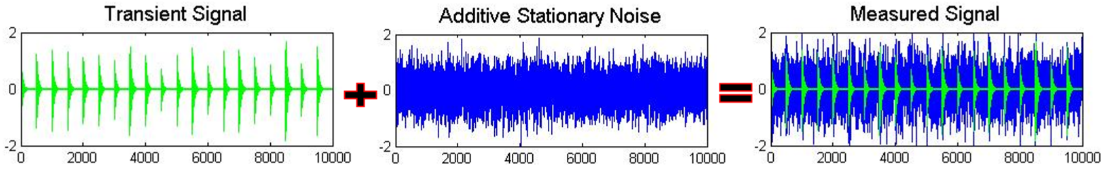


FIGURE 1 – Transient Signal **x** + Additive Stationary Noise **y** = Measured Signal **z**

## 3 Sparse + Sparse Combination Model

For the current problem (5), firstly, it is assumed that signal **x** can be represented in a over-complete dictionary $\Psi$ as $\mathbf{x} = \Psi\alpha$, and $\alpha$ is a sparse vector, which is obtained by the following equation :

$$\begin{aligned} &\text{Minimize}_\alpha \quad \|\alpha\|_{L_0} \\ &\text{Subject to} \quad \mathbf{x} = \Psi\alpha \end{aligned} \tag{6}$$

where $\|\alpha\|_{L_0}$ is a prior hypothesis for the coefficients distribution of the transient signal **x**, and the dictionary $\Psi$ can be built from existing time-frequency atoms or learned from a set of training samples (signals) [13]. In this paper, dictionary $\Psi$ is built from the discrete Gabor atoms [16] :

$$\Psi = \{\psi_{m,l}[n] = g(n-m)exp(\frac{i2\pi ln}{N})\}_{(m,l)\in R^2} \tag{7}$$

where $g(n)$ is the Hanning time window of period $N$ with unit norm $\|g\| = 1$, centered at $n = 0$. The atom $\psi_{m,l}$ is translated in time sample $m$ and frequency $l$, and each atom $\psi_{m,l}$ corresponds to a Heisenberg rectangle with a size $(\sigma_t, \sigma_\omega)$ which satisfies the uncertainty principle $\sigma_t\sigma_\omega \geq 1/2$.

Secondly, it is also known that the transient signal **x** is buried in strong noise **y** so that the measured signal reads $\mathbf{z} = \mathbf{x} + \mathbf{y}$ ; this condition is added into Eq. (6) to yield :

$$\begin{aligned} &\text{Minimize}_\alpha \quad \|\alpha\|_{L_0} \\ &\text{Subject to} \quad \mathbf{z} = \Psi\alpha + \mathbf{y} \end{aligned} \tag{8}$$

where $\mathbf{z} = \Psi\alpha + \mathbf{y}$ is the data fitting constrain. Equation (8) is not sufficient to accomplish the separation task, in order to approach the form of Eq. (2), thus, a prior hypothesis for the coefficient distribution of the noise **y** decomposition is searched. It is noted that noise **y** can not be represented parsimoniously under one determined Gabor dictionary $\Psi$, suggesting that the solution $\tilde{\alpha} = \Psi^\dagger \mathbf{y}$ († denotes pseudo inverse of a matrix) of equation $\mathbf{y} = \Psi\tilde{\alpha}$ is non-sparse (note that [17] claims that one realization of the noise can be represented parsimoniously under one realization of Gaussian dictionary, which is different with the Gabor dictionary $\Psi$ that was used here), which is congruous with the noise being non-compressible. Because noise **y** can not be decomposed with sparse coefficients, it seems that Eq. (8) is enough for separating the transient signal **x** from the strong additive noise **y**. Nevertheless, the trivial solution of Eq. (8) is a zero vector, therefore, if the measurement signal is all noise, the coefficient for representing the transient signal will be the most sparse (zero vector). Thus, one more constrained condition is needed to guarantee a non-trivial solution to exist.

Thirdly, it is observed that the noise energy $\sigma_\mathbf{y}{}^2$ of noise **y** is much larger than the signal energy $\sigma_\mathbf{x}{}^2$ of signal **x**, and it is assumed that the signal-to-noise ratio $\frac{\sigma_\mathbf{x}}{\sigma_\mathbf{y}} = \frac{\|\mathbf{x}\|_{L_2}}{\|\mathbf{y}\|_{L_2}} = \frac{\|\Psi\alpha\|_{L_2}}{\|\mathbf{y}\|_{L_2}} = c$ between the transient signal and noise energies is a small constant. It is assumed that the signal-to-noise ratio is a known parameter, which is considered as a constraint added to Eq. (8) to avoid a null sparse solution . Finally, all these constrains are collected together and problem (5) can be solved by the following equation :

$$\begin{aligned} &\text{Minimize}_\alpha \quad \|\alpha\|_{L_0} \\ &\text{Subject to} \quad \mathbf{z} = \Psi\alpha + \mathbf{y} \\ &\qquad \frac{\|\Psi\alpha\|_{L_2}}{\|\mathbf{y}\|_{L_2}} = c \end{aligned} \tag{9}$$

### 3.1 Matrix-Free Matching Pursuit Algorithm

There exists one difficulty to solve Eq. (9) which is explained here by an example. It is assumed that measurement $\mathbf{z} \in R^{10000}$, the dimension of the dictionary $\Psi$ is $10000 \times K$ and $K$ is much large than 10000, which makes it is impossible to tackle this huge dimension matrix with a desktop computer. Thus, a matrix-free algorithm based on Gabor transform is used to solve this problem. With this algorithm, the dictionary is constructed implicitly by the Gabor transform. As a preparation, the discrete Gabor transform of a signal $f$ of period $Nw$ is represented as [16]

$$
\begin{aligned}
\hat{f}[m,l] &= GT(f[n]) = \sum_{n=0}^{Nw-1} f[n] g_{Nw}[n-mR] exp(\frac{-i2\pi ln}{Nw}) \\
f[n] &= IGT(\hat{f}[m,l]) = \frac{1}{Nw} \sum_{m=0}^{Nw-1} g_{Nw}[n-mR] \sum_{l=0}^{Nw-1} \hat{f}[m,l] exp(\frac{i2\pi ln}{Nw})
\end{aligned}
\tag{10}
$$

where $Nw$ is the window length (Hanning window is used here), $R$ is the overlapping length between two continuous window blocks, and $R$ is set to be a constant parameter in this paper for simplicity. An matrix free algorithm is proposed : Firstly, the measurement signal $\mathbf{z}$ is assigned to $\hat{\mathbf{y}}$ ($\hat{\mathbf{y}}$ is set to be 0 initially). Then execute steps $2-6$ until $\frac{\|\hat{\mathbf{y}}\|_{L2}}{\|\mathbf{z}\|_{L2}} < (1-c)$. It is noteworthy that the trick is to calculate the Gabor transform $C^{Nw \times P} = GT(\hat{\mathbf{y}}, Nw)$ instead of calculating the matrix multiplication $\tilde{c}^{N\times 1} = \Psi^{\dagger}\hat{\mathbf{y}}$, which avoids storing a huge data matrix in memory. In step 3, $c^{*}_{m^*,l^*}$ is the element with the highest value in matrix $C$, $m^*$ and $l^*$ denote the position index in the matrix $C$, which is formally written as $c^{*}_{m^*,l^*} = max|C|$ ($|C|$ denotes absolute value for each entry in matrix $C$). The calculation $\hat{\mathbf{x}}_{\mathbf{i}} = d_i c^{*}_{m^*,l^*}$ ($d_i$ denote the $i$-th atom in dictionary $\Psi$) is implemented by inverse Gabor transform $\hat{\mathbf{x}}_i = IGT(c^{*}_{m^*,l^*}, Nw, m^*, l^*)$ of the corresponding time sample $m^*$ and frequency $l^*$ .

**Algorithm 1** Matrix Free Matching Pursuit Algorithm

**Input:**

Measured signal $\mathbf{z}$ ;
Signal-to-noise ratio $c$ ;
Window length $Nw$ ;

**Output:**

Estimated transient signal $\hat{\mathbf{x}}$ ;

1: **Set** $\hat{\mathbf{y}} = \mathbf{z}$ ;
2: **While** $\frac{\|\hat{\mathbf{y}}\|_{L2}}{\|\hat{\mathbf{z}}\|_{L2}} \geq (1-c)$ **do**
3: Compute $C^{Nw \times P} = GT(\hat{\mathbf{y}}, Nw)$ and find $c^{*}_{m^*,l^*} = max|C|$, where $1 < m < P$, $1 < l < Nw$ ;
4: $\hat{\mathbf{x}}_{\mathbf{i}} = IGT(c^{*}_{m^*,l^*}, Nw, m^*, l^*)$ ;
5: $\hat{\mathbf{y}} = \hat{\mathbf{y}} - \hat{\mathbf{x}}_{\mathbf{i}}$ ;
6: **End While**
7: Compute $\hat{\mathbf{x}} = \mathbf{z} - \hat{\mathbf{y}}$ ;

### 3.2 Connection with Morphological Component Analysis

In this section, the connection between Morphological Component Analysis(MCA) [14][15] and the proposed model (9) is explained. For the MCA model in Eq. (2), there are three essential considerations :

- Signal $\mathbf{x}$ can be represented by dictionary $\Psi$ and the coefficients $\alpha$ are sparse. According to the uncertainty principle[13][18], any two different representations of the same signal using an arbitrary dictionary cannot be jointly sparse. Therefore, the coefficients under the dictionary $\Psi$ for signal $\mathbf{y}$ is definitely dense. For the current problem, the transient signal $\mathbf{x}$ is represented by the Gabor dictionary $\Psi$ where the coefficients $\alpha$ are sparse, while noise $\mathbf{y}$ can not find a sparse representation under any dictionary. Thus, the coefficients under the dictionary $\Psi$ for noise $\mathbf{y}$ is also dense.
- Signal $\mathbf{y}$ can be represented by dictionary $\Phi$ where coefficients $\beta$ are sparse. Similarly, the coefficients under the dictionary $\Phi$ for signal $\mathbf{x}$ are definitely dense according to the uncertainty principle. For the current problem, this part ($\lambda \|\beta\|_{L_0}$) disappears because the noise $\mathbf{y}$ can not find a sparse representation under any dictionary.

- A constant parameter $\lambda$ is used to balance the ratio between the energy of signal $\mathbf{x}$ and the energy of signal $\mathbf{y}$ implicitly. The physical meaning is that the weight is assigned to the prior hypothesis and the these two combined prior hypothesis (regularization terms) terms are not equally important. When $\lambda$ is increased, $\|\beta\|_{L0}$ is given more weight, which means that the signal-to-noise ratio is increased. In fact, it is important to note that if $\lambda$ was assigned to $\|\alpha\|_{L0}$ (formally written as $\lambda\,\|\alpha\|_{L_0} + \|\beta\|_{L_0}$), the opposite conclusion would have been obtained : when $\lambda$ is increased, $\|\alpha\|_{L0}$ is given more weight, which means that the signal-to-noise ratio is decreased. For the current problem, the additional condition $\frac{\|\Psi\alpha\|_{L_2}}{\|\mathbf{y}\|_{L_2}} = c$ is used to balance the ratio between the energy of the transient signal $\mathbf{x}$ and the energy of noise $\mathbf{y}$ explicitly. When $c$ is increased, it suggests that the signal-to-noise ratio is increased.

# 4 Low Rank + Sparse Combination Model

For the rolling element bearing diagnosis problem (5), it is evident that the transient signal is sparse compared to the noise $\mathbf{y}$. Based on this assumption, Eq. (2) can be reformulated as follows [19][20] :

$$\begin{aligned} &\text{Minimize} \quad \|\beta\|_{L0} + \lambda\|\mathbf{x}\|_{L0} \\ &\text{Subject to} \;\; \mathbf{z} = \mathbf{x} + \Phi\beta. \end{aligned} \tag{11}$$

It is noted that the noise $\mathbf{y}$ can not be represented parsimoniously under one dictionary $\Phi$, thus, Eq. (11) can be further reformulated as follows. Define matrix $\mathbf{Y}$ ($[\mathbf{Y}]_{m,l}$ denotes the elements of matrix $\mathbf{Y}$) the Gabor transform of noise signal $\mathbf{y}$, that is $[\mathbf{Y}]_{m,l} = \hat{\mathbf{y}}[m,l] = GT(\mathbf{y}[n])$. The noise $\mathbf{y}$ is assumed to be stationary, thus $\mathbf{L} = E\{|\mathbf{Y}|^2_{ml}\}$ can be recognized as a low rank matrix, as a matter of fact $Rank(\mathbf{L}) = 1$ (see proof for more detail). Matrix $\mathbf{X}$ is defined as the Gabor transform of transient signal $\mathbf{x}$, which denotes as $[\mathbf{X}]_{m,l} = \hat{\mathbf{x}}[m,l] = GT(\mathbf{x}[n])$ and it is obvious that $\mathbf{S} = E\{|\mathbf{X}|^2_{ml}\}$ is a sparse matrix. Matrix $\mathbf{Z}$ is defined as the Gabor transform of measured signal $\mathbf{z}$, denoted as $[\mathbf{Z}]_{m,l} = \hat{\mathbf{z}}[m,l] = GT(\mathbf{z}[n])$ and the Spectrogram $\mathbf{M} = E\{|\mathbf{Z}|^2_{ml}\}$ of signal $\mathbf{z}$ can be represented as $\mathbf{M} = \mathbf{L} + \mathbf{S}$ . $\mathbf{L}$ corresponds to stationary noise signal which is low rank, and $\mathbf{S}$ corresponds to the transient signal which sparse. Thus, the signal separation problem can be solved by :

$$\begin{aligned} &\text{Minimize} \quad Rank(\mathbf{L}) + \lambda\|\mathbf{S}\|_{L0} \\ &\text{Subject to} \;\; \mathbf{L} + \mathbf{S} = \mathbf{M}. \end{aligned} \tag{12}$$

The minimization of the *Rank* and the $L_0$ norm function are both NP-hard problem, therefore Eq. (12) is relaxed into Eq. (13) entitled Robust Principal Component Analysis (RPCA) in the literature [21] :

$$\begin{aligned} &\text{Minimize} \quad \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_{L1} \\ &\text{Subject to} \;\; \mathbf{L} + \mathbf{S} = \mathbf{M} \end{aligned} \tag{13}$$

where $\|\mathbf{L}\|_* := \sum_{k=1}^{n} \sigma_k(\mathbf{L})$ is the nuclear norm of matrix $\mathbf{L}$ computed on the singular values $\sigma_1(\mathbf{L}) \geq \sigma_2(\mathbf{L}) \geq \ldots \sigma_n(\mathbf{L})$, and $\|\mathbf{S}\|_{L1} = \sum_{i=1,j=1}^{i=P,j=Q} |\mathbf{S}_{ij}|$. Nevertheless, matrix $\mathbf{S}$ can not be directly used to reconstruct the transient signal $\mathbf{x}$ because the phase information is lost. However, the latter can be recovered by taking over the phase information of matrix $\mathbf{Z}$, denoted by $[\mathbf{X}]_{m,l} = \sqrt{[\mathbf{S}]_{m,l}} e^{\angle[\mathbf{Z}]_{m,l}}$ and $[\mathbf{Y}]_{m,l} = \sqrt{[\mathbf{L}]_{m,l}} e^{\angle[\mathbf{Z}]_{m,l}}$. It is noteworthy to point out that the sparse matrix $\mathbf{X}$ is often contaminated by noise when $\lambda$ is chosen to be small (it suggests that some amount of noise is decomposed as sparse components). Therefore, a filter can be further used to refine the final results which is written as follows :

$$[\mathbf{F}]_{m,l} = \begin{cases} 1 & \text{if } |[\mathbf{X}]_{m,l}| > G|[\mathbf{Y}]_{m,l}| \\ 0 & \text{if } |[\mathbf{X}]_{m,l}| \leq G|[\mathbf{Y}]_{m,l}| \end{cases} \tag{14}$$

where $G$ is a parameter to control the amplitude gain ratio between $|[\mathbf{X}]_{m,l}|$ and $|[\mathbf{Y}]_{m,l}|$, similar filter was used in singing-voice separation scenario [22]. The purpose of this filter is to reduce noise in the sparse matrix $\mathbf{X}$. An illustrative example is seen in Fig. 2 : Fig. 2 (e) shows a matrix $\mathbf{L}$ that is rank-1, however, sparse matrix $\mathbf{S}$ is corrupted by noise in Fig. 2 (f). $X$ is also corrupted by noise since the difference between $\mathbf{S}$ and $\mathbf{X}$ is only that the phase information is added to $\mathbf{X}$ by $[\mathbf{X}]_{m,l} = \sqrt{[\mathbf{S}]_{m,l}} e^{\angle[\mathbf{Z}]_{m,l}}$ as compared to $\mathbf{S}$ (the noisy $\mathbf{X}$ figure is not shown in this paper). Figure. 3 shows the results after the filter operation $\mathbf{F} \odot \mathbf{X}$, where $\odot$ denotes Hadamard product. The parameter $\lambda$ in Eq. (13) is a trade-off between the low rank matrix and sparse matrix. When $\lambda$ is decreased , the decomposed matrix $\mathbf{L}$ has lower rank and $\mathbf{S}$ is less sparse, and vice versa. In Fig. 2 (a)(b), $\lambda$ is chosen to be 0.4, it is observed that no sparse components are decomposed ($\mathbf{S}$ is a zero matrix). In Fig. 2 (c)(d), $\lambda$ is decreased to be 0.1, it is found that more sparse components are decomposed for matrix $\mathbf{S}$. When $\lambda$ is further decreased to be 0.01, the rank-1 matrix $\mathbf{L}$ is fully decomposed, but the matrix $\mathbf{S}$ is no longer sparse due to the noise corruption resulting in the utilization of filter. Finally, the algorithm procedure for separating transient signal from noise is presented as follows :

**Algorithm 2** Algorithm Procedure For Low Rank + Sparse Combination Model

1: Calculate spectrogram matrix $\mathbf{M}$ of measured signal $\mathbf{z}$ by Eq. (10).
2: Decompose spectrogram matrix $\mathbf{M}$ into low rank real matrix $\mathbf{L}$ and sparse real matrix $\mathbf{S}$ by Eq. (13).
3: Recover sparse complex matrix $\mathbf{X}$ by $[\mathbf{X}]_{m,l} = \sqrt{[\mathbf{S}]_{m,l}} e^{\angle [\mathbf{Z}]_{m,l}}$.
4: Filter the noisy $\mathbf{X}$ by operation $\hat{\mathbf{X}} = \mathbf{F} \odot \mathbf{X}$, where $\hat{\mathbf{X}}$ is the filtered result.
5: Calculate Inverse Gabor Transform (IGT) of sparse complex matrix $\hat{\mathbf{X}}$ to reconstruct transient signal $\mathbf{x}$ by Eq. (10).

It is noteworthy to mention that without the expected value, matrix $\mathbf{L}$ in Eq. (12) is not a low rank matrix in practice. However, it is found that Eq. (13) is still close to the reality since the eigenvalues of matrix $\mathbf{L}$ decrease exponentially with the numerical index (more details about this are omitted here for the sake of brevity).

# 5 Simulation Experiment

- Simulation setup : a simulation setup to produce one transient signal $\mathbf{x}$ that is buried in the strong noise $\mathbf{y}$ is built by using the following equation :

$$s[n] = \sum_{m=0}^{Nm} \delta(n - mD_T) \quad \text{for} \quad n = 1, \ldots N \tag{15}$$

$$H[\mathbf{z}] = \frac{1}{1 - 2cos(2\pi \cdot 0.37) r\mathbf{z}^{-1} + r^2 \mathbf{z}^{-1}} \tag{16}$$

  Firstly, an impulse signal $s[n]$ (time duration is 1s, sampling frequency is $10^4$Hz) with length $N = 10^4$ is produced by Eq. (15) where $D_T = 5 \times 10^2$ and $Nm = 19$. Secondly, the impulse signal $s[n]$ is passed through a filter $H[\mathbf{z}]$ in Eq. (16) with $r = 0.98$ to synthesize the transient signal $\mathbf{x}[n]$. Finally, the simulated measured signal $\mathbf{z}[n]$ is synthesized by adding the transient signal $\mathbf{x}[n]$ to Gaussian noise $\mathbf{y}[n] \sim \mathcal{N}(0, 0.5)$.
- Simulation results for Sparse + Sparse model : in this simulation experiment, the parametric analysis focuses on the window length $Nw$ and signal-to-noise ratio parameter $c$. Firstly, the Hanning window length is fixed to $Nw = 256$, and the ratio $c$ is increased from 0.05 and 0.10 to 0.15. It is observed that more and more transient signal components are separated , see Fig. (4)(a)(c)(e). Secondly, the Hanning window length is set to $Nw = 1024$, and similar phenomenon is also observed, see Fig. (4)(b)(d)(f). It is also observed that the separation results are noisier with the increase of window length from $Nw = 256$ to $Nw = 1024$.
- Simulation results for Low Rank + Sparse model : Fig. 5 shows the separation results with different parameters by Low Rank + Sparse model. In Fig. 5 (a)(b)(c)(d), window length $Nw = 64$ and gain $G = 6$ are set to be constant, and it is observed that more and more transient signal is captured with the decrease of $\lambda$ from 0.02 to 0.001. It is noted that Fig. 5 (a) fails to capture the transient signal since too many sparse components in $\mathbf{X}$ are recovered ($\lambda$ is set too small) , which can be verified by Fig. 3 (a). Figure 5 (e)(f) show the results when the widow length is increased to $Nw = 256$.

# 6 Conclusion

In this paper, two kinds of combined regularization structures have been investigated to separate transient signals from strong additive noise in rolling element bearings diagnosis. First, the Sparse + Sparse combined

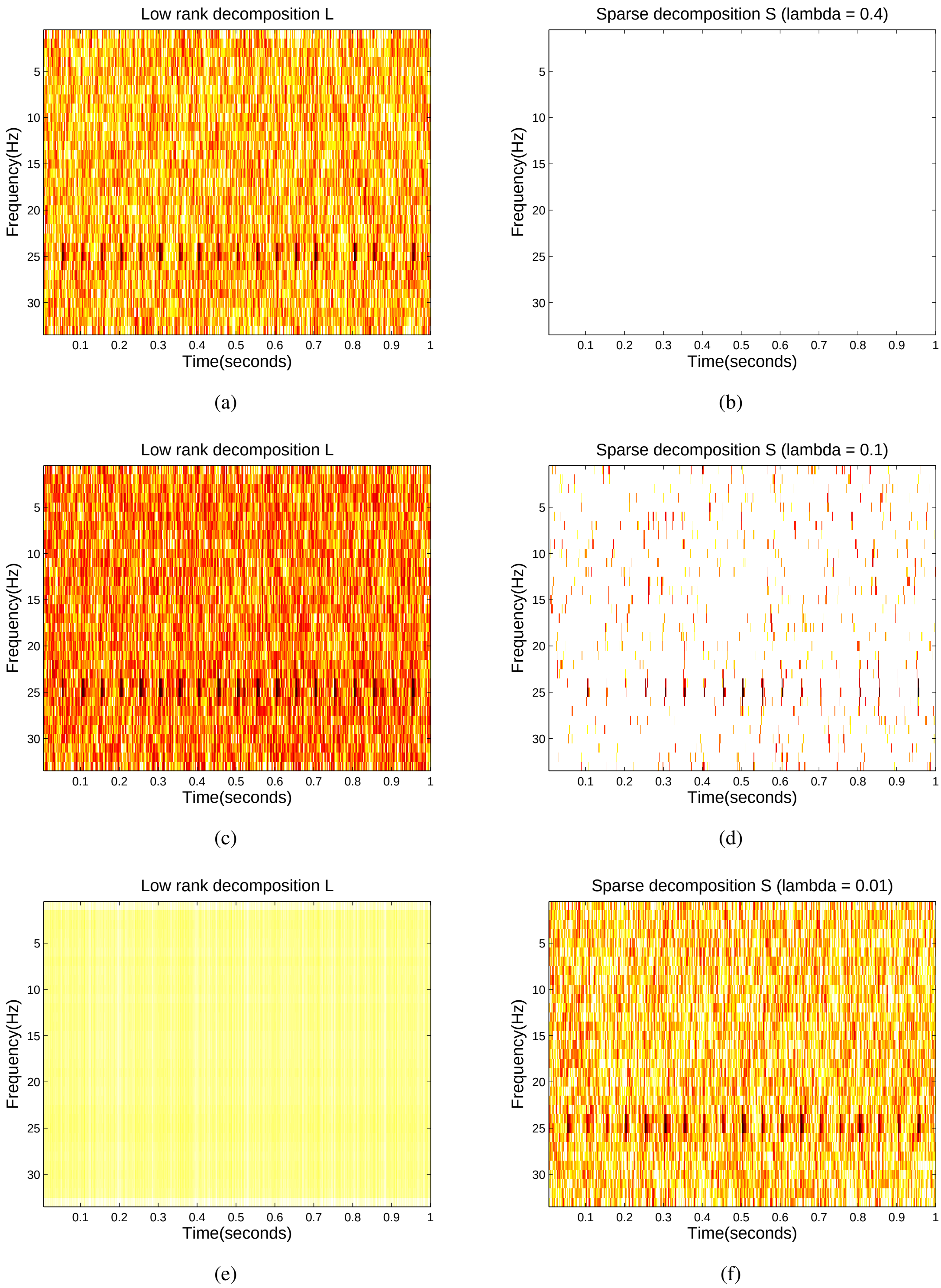


FIGURE 2 – Spectrogram $\mathbf{M}$ is decomposed into a low rank matrix $\mathbf{L}$ and a sparse matrix $\mathbf{S}$ with different parameters $\lambda$ : (a) Low rank decomposition $\mathbf{L}$ with $\lambda = 0.4$, (b) Sparse decomposition $\mathbf{S}$ with $\lambda = 0.4$, (c) Low rank decomposition $\mathbf{L}$ with $\lambda = 0.1$, (d) Sparse decomposition $\mathbf{S}$ with $\lambda = 0.1$, (e) Low rank decomposition $\mathbf{L}$ with $\lambda = 0.01$, (f) Sparse decomposition $\mathbf{S}$ with $\lambda = 0.01$,

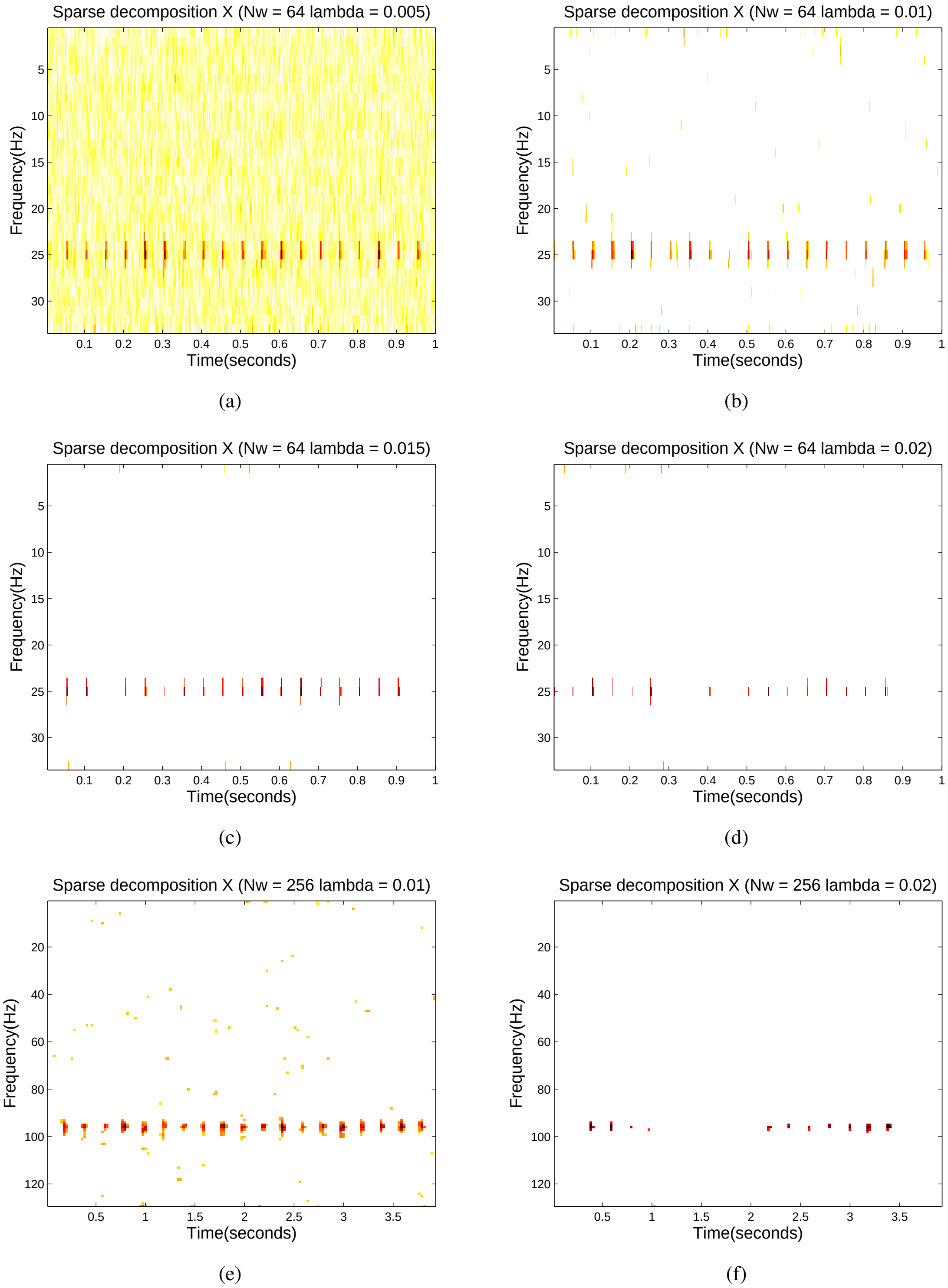


FIGURE 3 – Reconstructed sparse complex matrix **X** with different parameters : (a) Nw = 64, c = 0.005, (b) Nw = 64, c = 0.01, (c) Nw = 64, c = 0.015, (d) Nw = 64, c = 0.02, (e) Nw = 256, c = 0.01, (f) Nw = 256, c = 0.02,

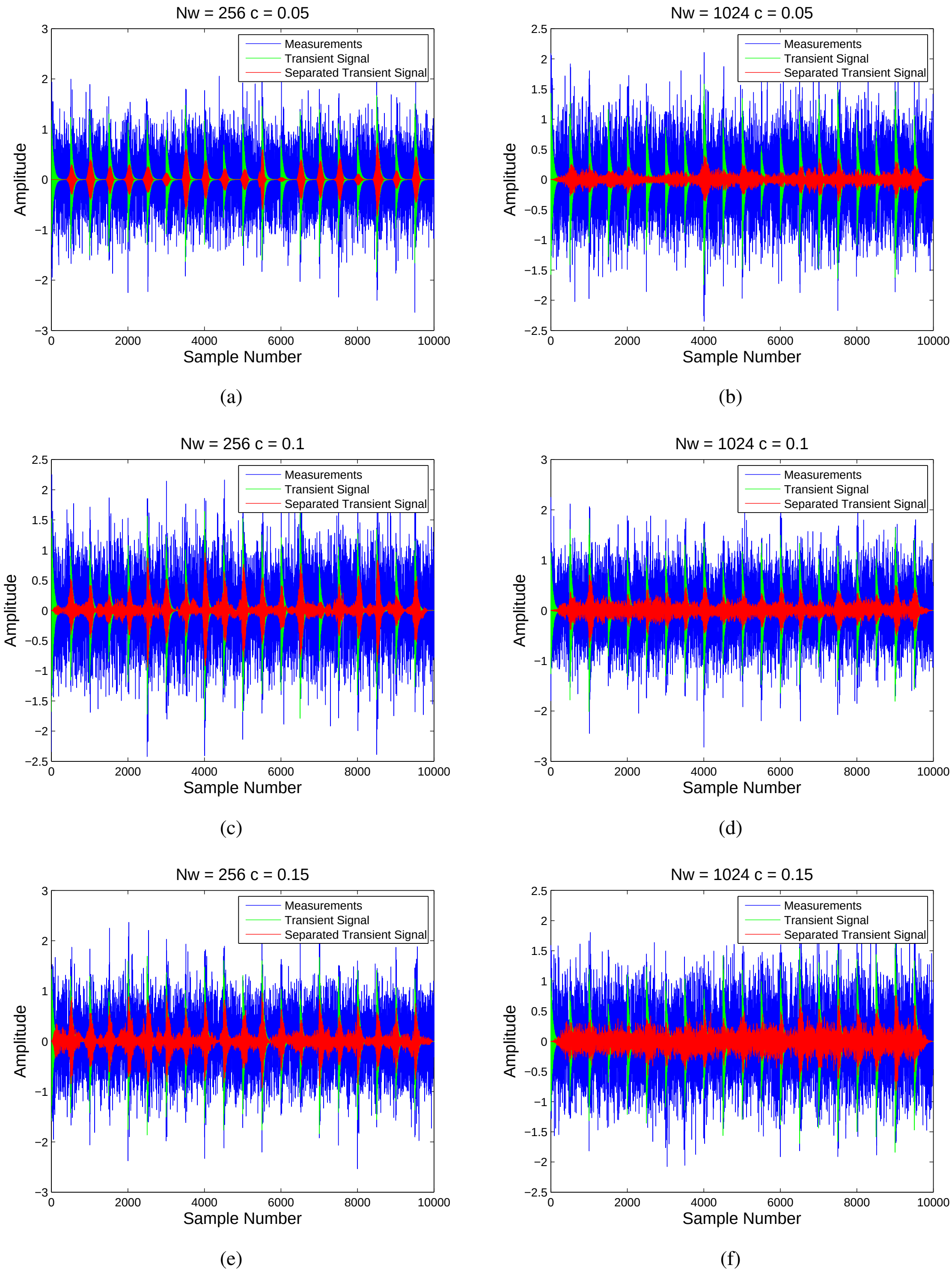


FIGURE 4 – Separation results of transient signal with different parameters by Sparse + Sparse model (a) Nw = 256, c = 0.05, (b) Nw = 1024, c = 0.05,(c) Nw = 256, c = 0.1,(d) Nw = 1024, c = 0.1,(e) Nw = 256, c = 0.15, (f) Nw = 1024, c = 0.15,

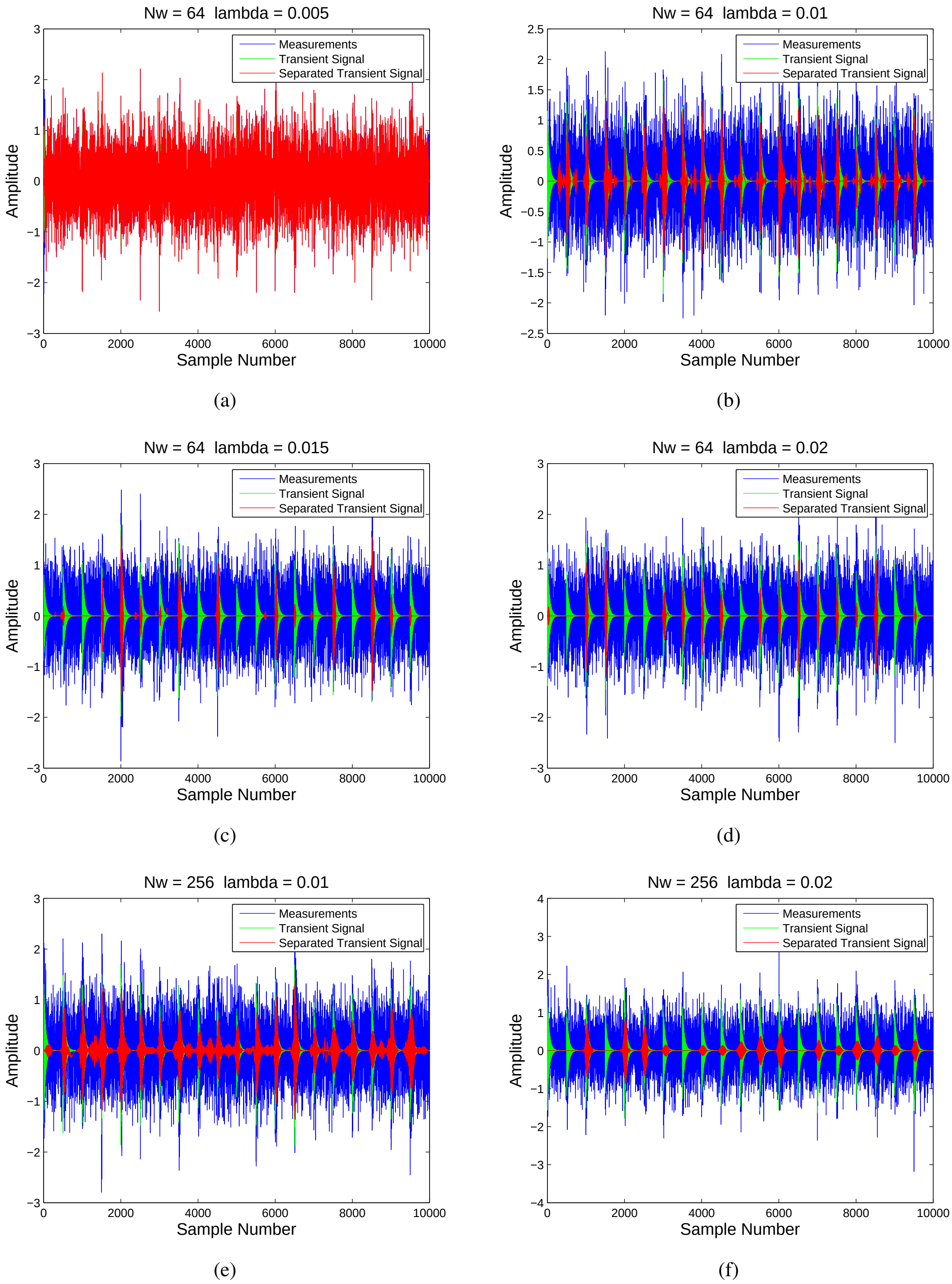


FIGURE 5 – Separation results of transient signal with different parameters by Low Rank + Sparse model : (a) Nw = 64, c = 0.005, (b) Nw = 64, c = 0.01, (c) Nw = 64, c = 0.015, (d) Nw = 64, c = 0.02, (e) Nw = 256, c = 0.01, (f) Nw = 256, c = 0.02,

regularization is investigated. It is found that this model can not be directly used in current problem, thus, a modified Sparse + Sparse combined regularization is proposed instead. Second, Low Rank + Sparse combined regularization is investigated and the connection with the Sparse + Sparse combined regularization is elucidated (two models are investigated in a uniform framework). The following topics can be further investigated :

- Is the Gabor dictionary good enough to represent the transient signal ? More specifically, how to choose a better dictionary ?
- How to choose the regularization parameter wisely since the separation performance highly depends on it ?

# A Proof of Low Rank Property For Matrix $\mathbf{L}$

Let us define the autocovariance $R_{2\mathbf{y}}[n-n^{'}] = E\{\mathbf{y}[n]\mathbf{y}^*[n^{'}]\}$ and $k = n-n^{'}$

$$
\begin{aligned}
L = [\mathbf{Y^2}]_{m,l} = E\{|\mathbf{Y}|^2_{ml}\} &= \sum_n \sum_{n'} \psi^*_{ml}[n^{'}]\psi_{ml}[n]E\{\mathbf{y}[n]\mathbf{y}^*[n^{'}]\} \\
&= \sum_n \sum_{n'} g_N[n-m]g_N[n^{'}-m]exp\{\frac{i2\pi l(n-n^{'})}{N}\}R_{2\mathbf{y}}[n-n^{'}] \\
&= \sum_n \sum_k g_N[n-m]g_N[n-k-m]exp\{\frac{i2\pi lk}{N}\}R_{2v}[k] \\
&= \sum_k (\underbrace{\sum_n g_N[n-m]g_N[n-k-m]}_{R_{NN}[k]})exp\{\frac{i2\pi lk}{N}\}R_{2\mathbf{y}}[k] \\
&= \sum_k R_{NN}[k]exp\{\frac{i2\pi lk}{N}\}R_{2\mathbf{y}}[k] \\
&= \sum_k R_{NN}[k]R_{2\mathbf{y}}[k]exp\{\frac{i2\pi lk}{N}\}
\end{aligned} \tag{17}
$$

The final results depends only on $l$. Thus, the matrix $\mathbf{L}$ is low rank.

# B Accelerated Proximal Gradient Algorithm

In this paper, Accelerated Proximal Gradient (APG) algorithm [23] is used to solved Eq. (13), which is rewritten as a penalized formulation as follows :

$$\text{Minimize} \quad \mu(\|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_{L1}) + \frac{1}{2}\|\mathbf{M}-\mathbf{L}-\mathbf{S}\|^2_F \tag{18}$$

where $\mu$ is a regularization parameter which is used as a trade-off between the regularization term $\|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_{L1}$ and the data fitting term $\frac{1}{2}\|\mathbf{M}-\mathbf{L}-\mathbf{S}\|^2_F$. It is also observed that the regularization term is a non-differentiable function and the data fitting term is a smooth function with Lipschitz gradient (Lipschitz constant $L_f = 2$). A soft thresholding operator $\mathscr{S}_{\frac{\lambda\mu}{L_f}}$ for any matrix $Q$ is defined as :

$$\mathscr{S}_{\frac{\lambda\mu}{L_f}}(Q) = \max(|q| - \frac{\lambda\mu}{L_f})\text{sgn}(q) \tag{19}$$

where $q$ denotes each entry of the matrix $Q$ and sgn() is the sign function. A soft thresholding $\mathscr{D}_{\frac{\mu}{L_f}}$ for any matrix $Q$ is defined as :

$$\mathscr{D}_{\frac{\mu}{L_f}}(Q) = U\mathscr{S}_{\frac{\mu}{L_f}}(\Sigma)V^T \tag{20}$$

where matrix $Q = U\Sigma V^T$ is decomposed by Singular Values Decomposition (SVD). Therefore, the accelerated proximal gradient algorithm can be iteratively written as algorithm 3. The iteration stops when SC $\leq \varepsilon$, where the stopping criterion $SC$ and $C_L, C_S$ are defined as :

$$\begin{aligned} \mathrm{SC} &= \frac{\|[C_L, C_S]\|_F}{L_f max\{1, \|[C_L, C_S]\|_F\}} \\ C_L &= L_f(G^k_{\mathbf{L}} - \mathbf{L}_{k+1}) + (\mathbf{L}_{k+1} + \mathbf{S}_{k+1} - G^k_{\mathbf{L}} - G^k_{\mathbf{S}}) \\ C_S &= L_f(G^k_{\mathbf{S}} - \mathbf{S}_{k+1}) + (\mathbf{L}_{k+1} + \mathbf{S}_{k+1} - G^k_{\mathbf{L}} - G^k_{\mathbf{S}}) \end{aligned} \tag{21}$$

where $[C_L, C_S]$ denotes the matrix constructor operator that vertically concatenates $C_L$ and $C_S$, and $\|\|_F$ denotes the Frobenius Norm. For the simulation in this paper, the parameters are chosen fixed as : $\mu = 10^{-3}$, $\eta = 0.9$ and $\varepsilon = 10^{-7}$.

**Algorithm 3** Accelerated Proximal Gradient Algorithm

**Input:**

Spectrogram matrix $\mathbf{M}$ and parameter $\lambda$ and $\mu$

**Output:**

Low rank real matrix $\mathbf{L}$ and sparse real matrix $\mathbf{S}$

1: Set $\mathbf{L}_0 = \mathbf{L}_1 = 0$, $\mathbf{S}_0 = \mathbf{S}_1 = 0$, $t_0 = t_1 = 1$, $\mu_0 = \|\mathbf{M}\|_F$
2: **While** not converged **do**
3: $G^k_{\mathbf{L}} = \mathbf{L}_k + \frac{t_{k-1}-1}{t_k}(\mathbf{L}_k - \mathbf{L}_{k-1})$
4: $G^k_{\mathbf{S}} = \mathbf{S}_k + \frac{t_{k-1}-1}{t_k}(\mathbf{S}_k - \mathbf{S}_{k-1})$
5: $\mathbf{L}_{k+1} = \mathscr{D}_{\frac{\mu_k}{L_f}}[G^k_{\mathbf{L}} - {L_f}^{-1}(\mathbf{M} - G^k_{\mathbf{L}} - G^k_{\mathbf{S}})]$
6: $\mathbf{S}_{k+1} = \mathscr{S}_{\frac{\lambda\mu_k}{L_f}}[G^k_{\mathbf{S}} - {L_f}^{-1}(\mathbf{M} - G^k_{\mathbf{L}} - G^k_{\mathbf{S}})]$
7: $t_{k+1} = \frac{1}{2}(1 + \sqrt{1 + 4t_k^2})$
8: $\mu_{k+1} = max(\eta\mu_k, \mu)$

## Références


[1] Robert B Randall and Jérôme Antoni. Rolling element bearing diagnostics–a tutorial. *Mechanical Systems and Signal Processing*, 25(2) :485–520, 2011.

[2] Zhipeng Feng, Ming Liang, and Fulei Chu. Recent advances in time–frequency analysis methods for machinery fault diagnosis : A review with application examples. *Mechanical Systems and Signal Processing*, 2013.

[3] Jérôme Antoni and RB Randall. The spectral kurtosis : application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2) :308–331, 2006.

[4] R Rubini and U Meneghetti. Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings. *Mechanical systems and signal processing*, 15(2) :287–302, 2001.

[5] Marco Cocconcelli, Radoslaw Zimroz, Riccardo Rubini, and Walter Bartelmus. Stft based approach for ball bearing fault detection in a varying speed motor. In Tahar Fakhfakh, Walter Bartelmus, Fakher Chaari, Radoslaw Zimroz, and Mohamed Haddar, editors, *Condition Monitoring of Machinery in Non-Stationary Operations*, pages 41–50. Springer Berlin Heidelberg, 2012.

[6] Weiguo Zhao, Lijuan Zhang, and Xujun Meng. Rolling bearing fault diagnosis using neural networks based on wavelet packet-characteristic entropy. In *Computer Science for Environmental Engineering and EcoInformatics*, pages 424–429. Springer, 2011.

[7] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6) :1045–1057, 2010.

[8] Haining Liu, Chengliang Liu, and Yixiang Huang. Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 25(2) :558–574, 2011.

[9] Hongyu Yang, Joseph Mathew, and Lin Ma. Fault diagnosis of rolling element bearings using basis pursuit. *Mechanical Systems and Signal Processing*, 19(2) :341 – 356, 2005.

[10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1) :1–122, 2011.

[11] Bing-sheng He Sheng-li Wang and Hai Yang. A modified variable-penalty alternating directions method for monotone variational inequalities. *Journal of Computational Mathematics*, 21(4) :495–504, 2003.

[12] Bingsheng He, Min Tao, Minghua Xu, and Xiaoming Yuan. An alternating direction-based contraction method for linearly constrained separable convex programming problems. *Optimization*, 62(4) :573–596, 2013.

[13] Michael Elad. *Sparse and redundant representations : from theory to applications in signal and image processing*. Springer, 2010.

[14] Jean-Luc Starck, Michael Elad, and David Donoho. Redundant multiscale transforms and their application for morphological component separation. *Advances in Imaging and Electron Physics*, 132 :287–348, 2004.

[15] Michael Elad, J-L Starck, Philippe Querre, and David L Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3) :340–358, 2005.

[16] Stephane Mallat. *A wavelet tour of signal processing : the sparse way*. Access Online via Elsevier, 2008.

[17] Ori Shental. Sparse representation of white gaussian noise with application to l0-norm decoding in noisy compressed sensing. *arXiv preprint arXiv :1104.2215*, 2011.

[18] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2) :489–509, 2006.

[19] John Wright and Yi Ma. Dense error correction via l1-minimization. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3033–3036. IEEE, 2009.

[20] Arvind Ganesh, John Wright, Xiaodong Li, Emmanuel J Candes, and Yi Ma. Dense error correction for low-rank matrices via principal component pursuit. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1513–1517. IEEE, 2010.

[21] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis ? *Journal of the ACM (JACM)*, 58(3) :11, 2011.

[22] Po-Sen Huang, Scott Deeann Chen, Paris Smaragdis, and Mark Hasegawa-Johnson. Singing-voice separation from monaural recordings using robust principal component analysis. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 57–60. IEEE, 2012.

[23] Zhouchen Lin, Arvind Ganesh, John Wright, Leqin Wu, MINMING Chen, and Yi Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 61, 2009.