Int J Adv Manuf Technol (2014) 70:185–199
DOI 10.1007/s00170-013-5258-5

ORIGINAL ARTICLE

Sparse representation and its applications in micro-milling condition monitoring: noise separation and tool condition monitoring

Kunpeng Zhu · Birgit Vogel-Heuser

Received: 25 April 2013 / Accepted: 12 August 2013 / Published online: 30 August 2013

Abstract This paper presents a new approach for cutting force denoising in micro-milling condition monitoring. In micro-milling, the comparatively small cutting force signal is contaminated by heavy noise, and as a result, it is necessary to denoise the force signal before further processing it. The traditional denoising methods, based on Gaussian noise assumption, are not effective in this situation because the noise is found to contain high non-Gaussian component. Based on the force and noise's sparse structures in the time–frequency domain, this approach employs a sparse decomposition approach and solves denoising as a convex optimization problem. It is shown that the proposed approach can separate the heavy non-Gaussian noise and recover useful information for condition monitoring.

Keywords Micro-milling · Condition monitoring · Sparse representation · Non-Gaussian noise

1 Introduction

CNC machining is fundamental to modern manufacturing. During machining, the contact between the cutting tool, work piece, and the chips imposes pressure on the tool and causes the shape of the tool to change, either gradually as tool wear or abruptly as tool fracture or breakage. Tool abnormal conditions can range from slightly exceeding the tolerances specified for the product to catastrophic failure of a tool which destroys an extremely expensive work piece.

Micro-milling is an advanced precision manufacturing technology which can produce high precision micro-scale components with hard materials. It is one of the most versatile micromachining operations and has been applied in high precision electronic instruments, aerospace, and medical and optical industries [1]. On the other hand, micro-milling is a discontinuous operation, with the teeth of the milling tool entering and exiting the work piece during each rotation. This interrupted cutting process brings shock and insatiability on the cutting tool. The cutting tool condition has direct influence on the final product quality. Due to its importance, tool condition monitoring has been extensively studied in manufacturing systems and many of them have been reviewed lately in [2, 3]. Generally, condition monitoring is to apply appropriate sensor layout and signal processing techniques to identify and predict the machine tool's state, so as to reduce loss brought about by low product quality or machine failures. The sensory signals can be used independently [4, 5] or combined as sensor fusion with various types of signals, such as vibration, force, and acoustics [4, 6–8] to provide compensated information for tool condition monitoring (TCM). With the implementation of different approaches to interpret these monitored signals, various goals of TCM can be reached. One of the prominent and most widely studied aspects is the detection of tool wear [4–6, 9–15], which is the main problem when the tool works under normal conditions. Other important tool failure modes are breakage and chipping [15–18], which lead to abrupt changes of tool geometry and result in unqualified products. The tool needs to be replaced immediately under these conditions. Studies on machining condition monitoring have also been carried from a system point of view, such as

K. Zhu (✉)
Institute of Advanced Manufacturing Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Huihong Building, Changwu middle road 801#, Changzhou 213164, China
e-mail: kunpengz@hotmail.com

K. Zhu · B. Vogel-Heuser
Department of Mechanical Engineering, Institute of Automation and Information Systems, Technical University of Munich, Boltzmannstr. 15, 85748 Garching, Munich, Germany

B. Vogel-Heuser
e-mail: vogel-heuser@ais.mw.tum.de

chatter [19, 20], vibration [21], ploughing [22], and tool deflection detection [23].

Condition monitoring is most important but difficult to implement in micromachining due to the component's tiny structure and high precision requirements. Signal noise always exists in machining, and denoising is prerequisite in condition monitoring in mechatronic systems [24–28]. The noise problem in milling is due to nature of the impulse response of the intermittent milling process and the system vibration brought by the (almost) periodical variations of the cutting forces. The problem is much magnified when the machining is scaled down to microlevel, and the cutting force is at very low amplitude. Many researchers have noticed this noisy problem. Tansel et al. [15] found the force was too noisy and then filtered it with a band pass filter to reduce the noisy effect. The authors also sort for acoustic emission (AE) signals for a combination. Similar studies were reported in [7] with a combination of force and AE signals. AE signals were also applied in [5, 17] for micro-milling monitoring. The AE signal is sensitive with very wide dynamic bandwidth, while the wide-band property also limits its application, as the data needs to be sampled at very high sampling rate (>1 MHz). An online monitoring with AE is almost impossible with current computing facilities, and at the same time, AE signals are very sensitive to sensor position and mounting surface. In [17], different schemes of sensor fusions were studied from vibration, force, and AE signals. The signal features were fused to a neuro-fuzzy network. It was observed that the neuro-fuzzy algorithm could effectively monitor the tool conditions in micro-milling. Vibration features were studied to monitor the tool conditions in [11, 16, 17]. In [11], the spindle vibration frequency features were elected as inputs to a neural network to classify the tool wear. It was reported that this framework could provide good classification rate. In [18], the authors applied motor current signals to detect tool breakage. Features from wavelet and statistical domains were extracted to correlate with different tool conditions. This provides an economical and convenient solution to TCM. However, these signals are not as sensitive as force and AE.

Cutting force has been found to be one of the most sensitive measurements for machining condition monitoring [2]. However, unlike macro machining, a difficulty of denoising in micromachining is that the noise is usually very high [low signal-to-noise ratio (SNR)] and found to be non-Gaussian. While most current denoising algorithms are based on Gaussian noise assumption [29], they are limited to apply to this case. On the other hand, it has been found that the noise harmonics distribute widely in the frequency domain. As a result, the traditional approaches, such as filtering and wavelet thresholding, are not effective in this situation. It should be noted that this denoising preprocessing may be skipped when the analysis is based on the averaged force features, as the noise is rather symmetric in general. The important instantaneous information was lost under this condition however.

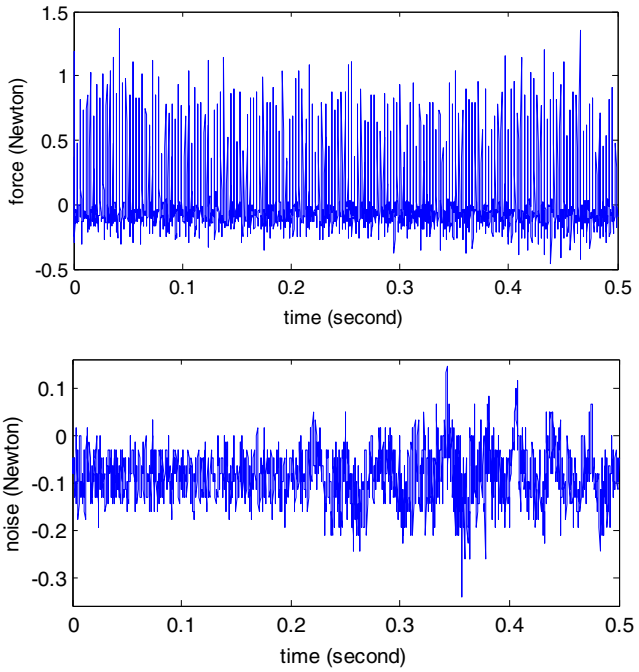

Fig. 1 Cutting force and noise in micro-milling. Both the force and noise are sampled at 6,000 Hz, and 0.5 s accounts for 3,000 sample points

Figure 1 shows a typical sample of the cutting force and noise in micro-milling. As can be seen from Fig. 1, the SNR is very low and the absolute value of the noise is nearly one third that of the force. At the same time, the frequency bands of noise and force are overlapped quite a lot (this will be shown in later sections). An effective noise removal procedure is an essential first step towards accurate estimation of the tool wear state in micromachining.

Based on the non-Gaussian noise assumption, in [28], the authors applied the independent component analysis (ICA) approach to deal with this problem; it has been found to be effective but the extracted independent components are prone to variations under different initializations. The other problem is that the approach in [28] needs to compute at least three force components to extract the independent components; it will not be applicable in case of less channels monitored. In this paper, a new method is developed for force denoising, which takes advantages of the special properties of non-Gaussian noise in the time–frequency domain.

Sparse representation (also named compressive sensing, sparse coding in information theory [30, 31]) has received a great deal of attentions in recent years from applied mathematics [30] to signal classification and image recognition [31–36]. Sparse representation originates from wavelet analysis while it has widened and deepened the scope of wavelet analysis. The idea of sparse representation is to represent a signal sparsely in terms of linear combination of atoms in an overcomplete (redundant) dictionary [29, 30]. The dictionary's

redundancy offers a wide choice of atoms to represent the signal so that a smallest linear combination of atoms could be selected to approximate the signal. The affront mentioned ICA approaches have been found to be closed related to the sparse representation [37].

In this study, based on both the signal and noise properties in the time–frequency domain, a sparse decomposition is developed for cutting force denoising and condition monitoring in micro-milling. The paper is organized as follows: Section 2 identifies the noise properties in micro-milling, and then the principle of sparse decomposition for denoising is developed in Section 3. Section 4 illustrates the denoising results in the time–frequency domain and discusses the results. The conclusions are then drawn in Section 5.

2 Noise and force properties and their properties in the time–frequency domain

2.1 The STFT

For any signal $y(t)$, the Fourier transform $\widehat{y}(f)$ is obtained by inner product of $y(t)$ with a sinusoidal wave $e^{j2\pi ft}$,

$$\widehat{y}(f) = \langle y(t), e^{j2\pi ft} \rangle = \int_{-\infty}^{\infty} y(t) e^{-j2\pi f} dt \tag{1}$$

The Fourier transform transforms the signal $y(t)$ from the time domain into the frequency domain and results in the amplitudes of the frequency f. To overcome the limitation of FT's lack of local information, a time sliding window is applied to the Fourier transform. The resultant transform is denoted windowed or short-time Fourier transform (STFT),

$$STFT_y(t,f) = \int_{t-T/2}^{t+T/2} y(\tau) g(\tau - t) \; e^{-2\pi jf\tau} d\tau \tag{2}$$

The corresponding energy density $|\mathrm{STFT}_y(t,f)|^2$ is called a spectrogram, which are widely used for time–frequency analysis. The STFT coefficients represent the signal in the sparse time–frequency blocks at different positions.

2.2 Noise properties

As can be seen from Fig. 1, the SNR is very low and the amplitude range of the noise is nearly one third of that of the force. The SNR is calculated as,

$$SNR = 10 \times \log_{10} \frac{s_i^2}{n_i^2} = 10.1 \;\; dB \tag{3}$$

According to formula (3), the SNR is generally around (9–15 dB) for micro-milling in this study. This value is rather low, and denoising is prerequisite for a reliable correlation of force tool conditions. The noise statistics are listed in Table 1. To find a suitable denoise approach, the noise properties are demonstrated in Fig. 2. The noise is collected during the recess of the tool hold during machining, which means the segment is in between two cutting process and is pure noise.

The density estimation shows that the noise is super Gaussian, with a longer tail than Gaussian distribution and is well fitted with Laplace distribution (see Fig. 2b). The normalized frequency representation (power spectrum density, PSD) and autocorrelation show that the noise harmonics distribute widely in the frequency domain, and the autocorrelation is high with value around 0.8. This means that the noise is neither Gaussian nor white. As a result, the traditional approaches, such as filtering and wavelet thresholding, are not effective under this case.

The interesting phenomenon in the above figures is that only several amplitudes in the PSD are prominent and the others are very small. When the PSD is integrated along the time, it is the time–frequency representation (it is named spectrogram and will be discussed in details in Section 3.3). Not like dense values in the time domain, the noise has sparse structure in the time–frequency domain (Fig. 3). On the other side, the force signal also shows such a sparsity in the time–frequency domain, which will be discussed in the next section.

Based on the sparse property of noise and force in the time–frequency domain, a sparse representation approach is applied for noise separation in the micro-milling condition monitoring. The idea of this approach is based on the findings that both force and noise could be sparsely represented in the time–frequency domain, and that the dictionaries (time–frequency bases) can be learned differently for force and noise signals. The details are discussed in the following sections.

Table 1 The statistics of reference noise

Statistics	Value
Mean	−0.1876
Variance	0.2103
Skew	0.006
Kurtosis	2.2621

2.3 Force properties

In Fig. 4, a force component in the time domain is illustrated and its corresponding representation in the frequency domain by applying PSD. As can be seen from the figure, the signal has nonzero coefficients for almost every time point in time domain, but in the frequency domain, only some frequencies (harmonics) have large coefficient. These properties motivate the sparse representation of both noise and force in the time–frequency domain.

3 Sparse force and noise representation in the time–frequency domain

3.1 Sparse representation theory

As the sparse representation theory is quite new to manufacturing society, a little more details are introduced in this section. When treated as a vector, a signal is said sparse if most of its entries are equal or close to zero. The idea of sparse representation is to code a signal with linear combination of atoms in an overcomplete (redundant) dictionary so that the representation is sparse [29]. The motivation for overcomplete is that larger dictionaries incorporating more patterns and increase sparsity. It improves the applications to compression, denoising, and pattern recognition as a result.

In signal theory, any square integrable function $y(t)$ is regarded as signal. With the square integrable constraint, $\int_{-\infty}^{\infty}|y(t)|^2dt<+\infty$, the signal $y(t)$ is an energy signal and belongs to the Hilbert space, $L_2(R)$ (the reader unfamiliar with Hilbert space may simply think Hilbert space is a generalization of the three-dimensional Euclidean space in which the concepts of distance and angle are extended as norm and inner product, respectively. For any signal in the Hilbert space, it has the basic operation properties that can be added, subtracted, and multiplied by constants. See [38] for more details.)

In the Hilbert space, the l_p norm for a given signal $y=[y_1, y_2,\ldots,y_n]^T$ is defined as

$$\|y\|_p = \left[\sum_{i=1}^{n}|y_i|^p\right]^{1/p} \quad (p\geq 1) \tag{4}$$

In particular, $\|y\|_1 = \sum_{i=1}^{n}|y_i|$, $\|y\|_2 = \sqrt{\sum_{i=1}^{n}|y_i|^2}$, $\|y\|_\infty = \max_{1\leq i\leq n}\{|y_i|\}$

The so-called l_0 norm, which is excluded from the above definition domain, is an extension of the l_p norm definition when $p\to 0$. Specifically, the sparsity of a discrete signal (vector) y is defined by

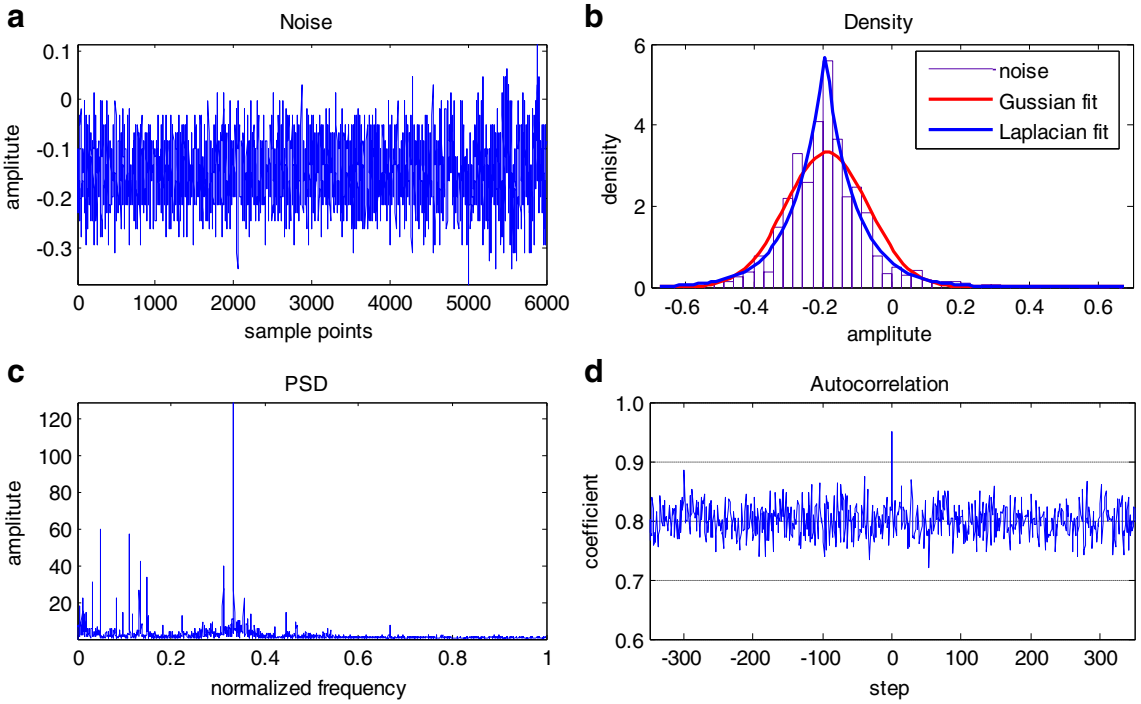

Fig. 2 Noise properties: **a** the noise signal, **b** noise distribution compared to Gaussian and Laplace distributions, **c** in the normalized frequency domain, **d** the autocorrelation coefficients. The noise is sampled with 6,000 Hz, and the segment in **a** is just 1 s. **b** shows that it is far from Gaussian

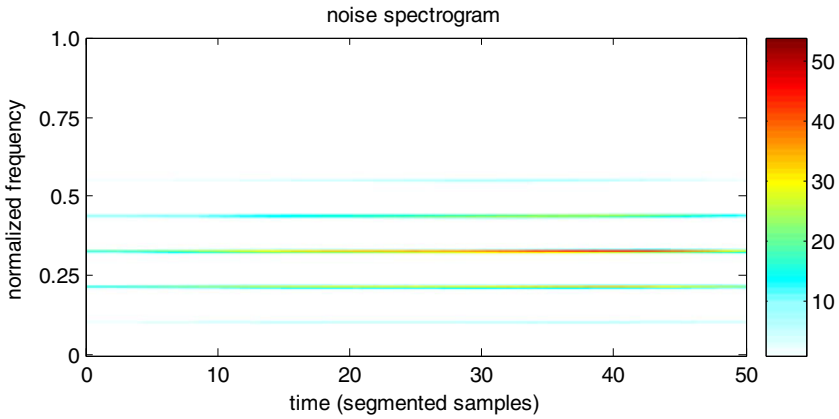

Fig. 3 The sparse property of noise spectrogram. The spectrogram is an integrated PSD along the time axis. In the figure, *darker points* indicate larger value coefficients and *white area* indicates coefficients are or very close to zero

$$\|y\|_0 = \sum_{i=1}^{n} |y_i|^0 = \text{number of nonzeros in y} \tag{5}$$

In the Hilbert space, one can always find a set of functions $\{\varphi_i\}$ in $L_2(R)$, such that the signal y can be expressed as the linear combination of N elementary waveforms, signal atoms $\{\varphi_i\}$, such that

$$y = \Phi x = \sum_{i=1}^{N} x_i \varphi_i \tag{6}$$

Where x_i is the representation coefficients of y in the dictionary $\Phi = (\varphi_1, \ldots, \varphi_N)$.

Clearly, x and y are equivalent representations of the same signal, with y in the time (or any other domain) and x in the Φ domain. The signal y is said to be K-sparse in Φ if it is a linear combination of only K basis vectors, that is, only K of the x_i coefficients in (6) is nonzero. In practice, x is considered sparse if it has just a few large coefficients and other coefficients are relatively small.

Figure 5 demonstrates such an idea. As it has been known, only two basis vector is needed to fully represent a vector in two dimension. The coefficients a_1 and a_2 are both large in the two-dimension representation (Fig. 5a). While it is represented in the three-dimensional space, Fig. 5b shows that in higher dimension, better basis (φ_1) can be found to represent the signal. The basis φ_1 is close to y, and the coefficient of this basis is large and the other twos are small. At the same time, there will be redundant information with three basis vectors representation. This is because the three basis vectors are linear dependent, as $\varphi_1 + \varphi_2 + \varphi_3 = 0$. This redundant

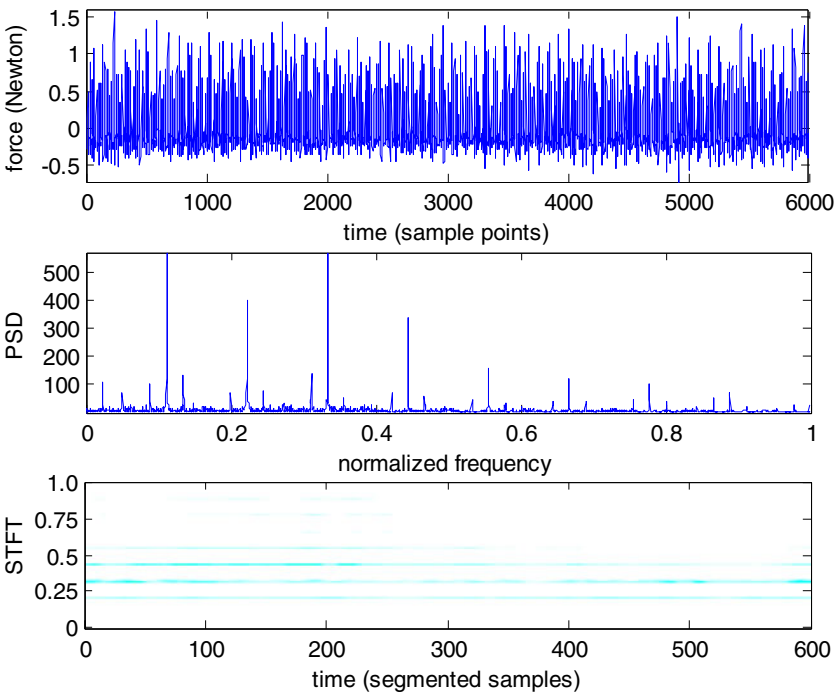

Fig. 4 The force and its sparse representation in PSD and STFT domain

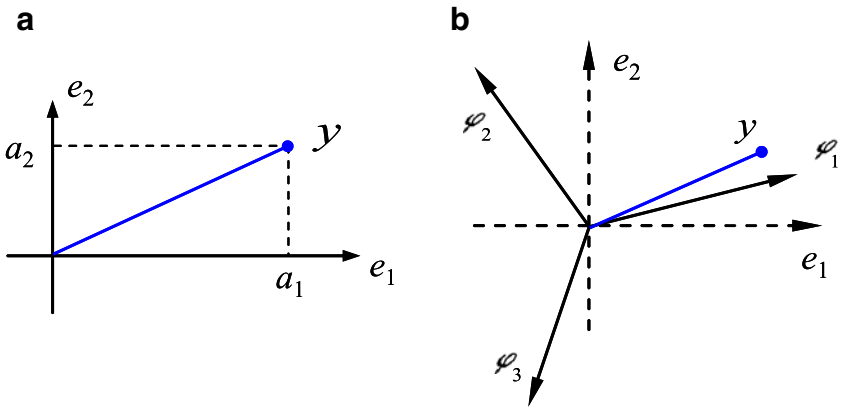

Fig. 5 The expansion of vector x in space $L_2(R)$. **a** Represented in complete base $\{e_1,e_2\}$ with coefficients a_1 and a_2. **b** Represented in overcomplete base $\{\varphi_1,\varphi_2,\varphi_3\}$

representation approach is close to the frame theory in the traditional wavelet analysis, and details could be found in [29].

With the l_0 norm definition in (5), the sparse coding can be formally expressed as a mathematical optimization problem: Given a signal $y \in R^n$ and a dictionary $\Phi \in R^{n\times k}$, the sparse representation problem can be stated as

$$\min_x \|x\|_0 \quad \text{s.t. } y = \Phi x, \quad \left(x \in R^k\right) \tag{7}$$

This form is also known as compressive sensing (sampling) in information theory [30]. This problem is NP-hard difficult to solve. On the other side, in general, the l_1 constraint also induces sparse solutions for the coefficient vectors x. By substituting the l_0 norm in Equation (7) with l_1 norm, it results in the Lasso-type problem [38],

$$\min_x \|y-\Phi x\|_2^2 + \lambda\|x\|_1 \tag{8}$$

Where λ is a parameter that balances the trade-off between reconstruction error and sparsity. The equation (8) is a convex problem that can be solved very efficiently using for example the LARS–Lasso algorithm [39].

3.2 The properties of sparse representation

The two key points in the development of sparse representation are sparsity and incoherence [40]. Sparsity indicates that the information carried by the signal is much smaller than its bandwidth. Incoherence means that the basis must be significantly different, and the dictionary consists of redundant (overcomplete) bases as a result. Typically, such overcomplete dictionaries are wavelet packets, Gabor dictionaries, and local cosine dictionaries.

Overcomplete dictionaries An overcomplete dictionary can be represented as a matrix $\Phi \in C^{m\times n}$ with $n > m$. The n column (atom) contains at least m linearly independent vectors, and it means the dictionary contains more than necessary vectors to represent signals, and so it is also called redundant dictionaries. With the overcomplete dictionary, a dense signal could be transformed into a sparse signal.

Figure 6 gives a demonstration of sparse representation with overcomplete dictionary representation (sparse coding). The original measurement $y_{20\times 1}$ has 20 nonzero coefficients,

Fig. 6 Sparse representation of the signal vector

which is a dense vector. When it is represented in terms of an overcomplete dictionary $\Phi_{20\times50}$, the coefficient vector $x_{50\times1}$ is sparse with $K=10$ in a higher dimension.

Incoherence between bases Incoherence means that the basis must be significantly different. For a dictionary $D=[\Phi,\Psi]$, the coherence between the two orthonormal matrices $\Phi_{n\times n}$ and $\Psi_{n\times n}$ is defined as [40]

$$\mu(\Phi,\Psi)=\sqrt{n}\max_{1\le k,j\le n}\left|\phi_k^T\psi_j\right| \tag{9}$$

Where ${\phi_k}^T$ and ψ_j are the k-th row and j-th column of Φ and Ψ. The coherence $\mu(\Phi,\Psi)$ measures similarity of basis in Φ and Ψ. Incoherence refers to the bases with small coherence, which means the bases are less similar. As a result, a dictionary with incoherent bases has large varieties and "richer vocabulary" in the dictionary. At the same time, the dictionary D has to meet some properties so as to represent the signal sparsely and compressively. Under the l_1 minimization, sufficient conditions have been raised that D satisfies the restricted isometry property (RIP) [41] or null space property (NSP) [42]. The RIP is robust and is a stronger condition than the NSP which is both necessary and sufficient [42].

As demonstrated in Section 2, the STFT coefficients represent the signal in the sparse time–frequency blocks at different positions. The STFT analysis is characterized by the window function, overlap, and duration. With proper choosing of window size and overlapping interval, these time–frequency blocks $|\mathrm{STFT}_y(t,f)|^2$ forms a redundant dictionary.

3.3 Sparse representation as a convex optimization problem

From the pattern classification point of view, the sparse representation in 3.1 is a reconstructive approach, as it is basically decomposed to contain sufficient information for reconstruction. While in this study, the aim is for noise separation and then the tool state estimation. The basis' discrimination power is concerned besides the signal's reconstruction error from the commonly applied sparse decomposition approach. The basis is the combination of the pre-learned basis of the pure noise Ψ and the updating signal basis $\Phi^{(i)}$, $[\Phi(i);\Psi]$.

This work introduces a metric which includes both reconstruction and discrimination terms in the dictionary learning process, benefitting from the best of both discriminative and reconstructive worlds. It is illustrated to optimize objective function J,

$$\begin{aligned}J&=\max(\text{Discrimination})+\min(\text{Reconstruction})+\max(\text{Sparsity})\\&=\max(J_1)+\min(J_2)+\max(J_3)\end{aligned} \tag{10}$$

Where J_1 maximizes the discrimination of the basis in the STFT, and J_2 controls the reconstruction error, and J_3 domain maximizes the sparsity of estimated force. This formulation of the final problem is similar to the ideas that have been studied for texture classification [34], face recognition [36], and handwritten digit recognition [34–36].

A measure of basis discrimination is adapted from Fisher's discriminant analysis [43],

$$J(w)=\frac{S_B}{S_W} \tag{11}$$

Where S_B is the between-basis scatter and S_W is the within-basis scatter.

For a c- basis problem, define the within-basis scatter,

$$S_w=\sum_{i=1}^{c}S_i \tag{12}$$

and the between-basis scatter,

$$S_B=\sum_{i=1}^{c}n_i(\mu_i-\mu_c)(\mu_i-\mu_c)^T \tag{13}$$

$$\text{Where } S_i=\frac{1}{n_i}\sum_{i=1}^{n}(\varphi_i-\mu_i)(\varphi_i-\mu_i)^T \tag{14}$$

is the covariance matrix measure the variability of sample within basis φ_i, $\mu_i=\frac{1}{n_i}\sum_{i=1}^{n}\varphi_i$ is the mean of samples from basis φ_i, n_i is the number of data in basis φ_i, and μ_c is the mean of basis φ_c.

To find the maximization of $J(w)$, this question can be formulated equivalently as an optimization problem by minimizing the inverse of $J(w)$,

$$J_1=\min_w\ \frac{1}{2}\frac{S_W}{S_B} \tag{15}$$

Where the role of within and between scatter is switched.

The measure J_2 is the reconstruction error l_2 norm that illustrates the robustness of the algorithm,

$$J_2=\min\left\|y_i-{\varphi_i}^Tx_i\right\|_2^2 \tag{16}$$

The measure J_3 forces the sparsity of the representation the STFT,

$$J_3=\min|x_i|_1 \tag{17}$$

So finally, the sparse representation problem is formalized as a convex optimization problem,

Table 2 The modified K-SVD algorithm

Initialization : set the random normalized dictionary matrix
$\Phi^{(0)}$=abs(randm(m,k)), $X^{(0)}$=abs(randm(k,n)),$k<m,n$
Set J=1, $\Phi^{(0)}=[\Phi^{(0)};\Psi]$
Repeat until convergence
Sparse coding: use OMP algorithm to compute the sparse representation $\{\varphi_i\}$ for each signal $\{y_i\}$
for i=1,2,…,N
min $\frac{1}{2}\frac{S_W}{S_B}+\frac{1}{2}\lambda_1\|y_i-\varphi^T x_i\|_2^2+\lambda_2|x_i|_1$
Dictionary update: For $k=1, 2, \ldots,K$
- Define the group of examples that use φ_k, $w_k=\{i|1\le i\le N, x_i(k)\neq 0\}$
- Compute residual matrix
$$E_k = y-\sum_{j\neq k}\varphi_j^T x_j$$
- Restrict E_k by choosing only the columns corresponding to those elements that initially used φ_k in their representation, and obtain E_k^R.
- Apply SVD decomposition $E_k^R=U\Delta V^T$.

Update: $\varphi_k=u_1, x_R^k=\Delta(1,1)\cdot v_1$
Set $J=J+1$

$$\min\frac{1}{2}\frac{S_W}{S_B}+\frac{1}{2}\lambda_1\left\|y_i-\varphi_i{}^T x_i\right\|_2^2+\lambda_2|x_i|_1 \quad (18)$$

In (18), λ_1 is the l_2 norm regularization parameter, and λ_2 is the l_1 norm regularization parameter.

How to learn the dictionary $\{\varphi_i\}$ is most important in this approach. The best dictionary is the one that leads to the sparsest representation. Hence, it could be desirable to have a huge dictionary (i.e., $T>>N$) but it leads to a prohibitive computation time cost for calculating the x coefficients. Therefore, there is a trade-off between the complexity of our analysis (i.e., the size of the dictionary) and computation time. Pursuit algorithms can nearly reach optimal M-term approximations in incoherent dictionaries that include vectors that are sufficiently different [29].

In this work, an l_1 variation of K-SVD [32] is adapted to learn the dictionary by solving the optimization problem in formula (18) in which the orthogonal matching pursuit (OMP) [29, 32] is implemented. The optimization is carried out iteratively in three steps:

a Learn the dictionary Ψ from pure noise.
b Sparsely code the y given the current dictionary estimate.
c Update the dictionary atoms given the sparse representations.

The dictionary update is performed one atom at a time, optimizing the target function for each atom individually while keeping the remaining atoms fixed. The time domain signal then can be recovered by inversion of STFT transformation. The detail implementation of the algorithm is described in Table 2.

4 Case Studies

4.1 Experiment Setup

A total of 27 test experiments are conducted with different spindle speed, depth of cut, radial depth of cut, and feed rate for Φ500 μm and Φ800 μm tools [26]. The materials used are either copper or steel. The tool wear was measured using the Olympus Toolmakers microscope at 213 times enlargement. Figure 7a shows a fresh tool with no flank wear, while Fig. 7b shows an average flank wear of (64 μm+58 μm)/2=61 μm.

The machine used in the micro-milling experiment is an end milling machine driven by a 22 kw spindle drive motor, and the spindle speed variation is between 5,000–30,000 rpm. The cutting force was measured with a three-channel dynamometer mounted under the work piece. The cutting force output was recorded as voltage variations on a digital tape recorder and down-sampled at 6,000 Hz. The working conditions are attached in the Appendix, which was previously

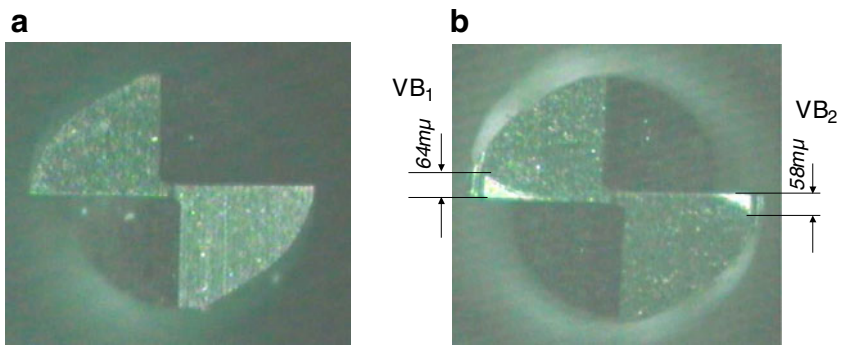

Fig. 7 Flank wear a micro-mill tool with diameter Φ800 μm. **a** fresh tool **b** worn tool

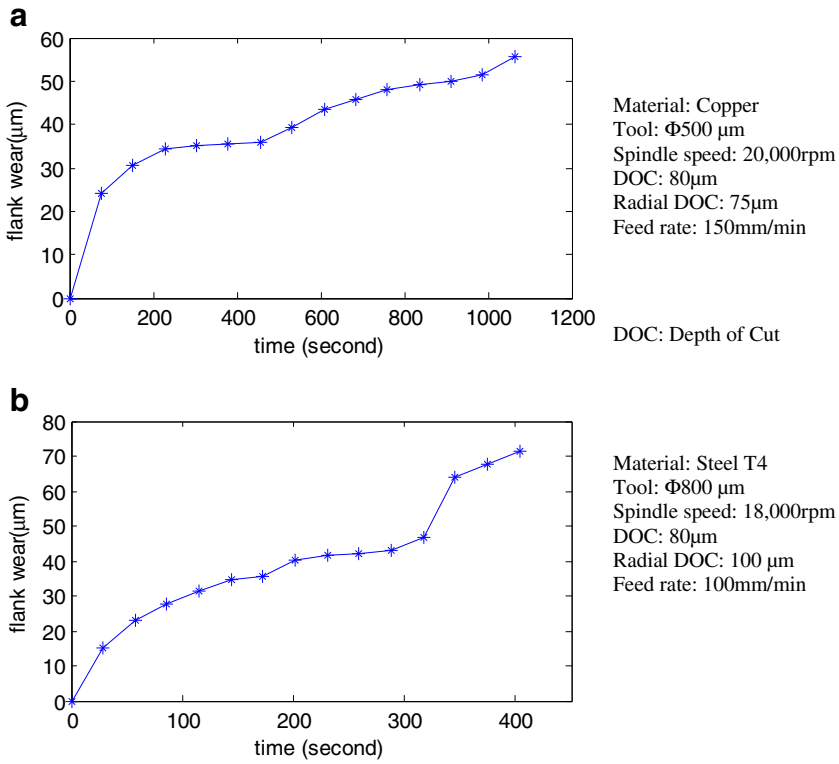

Fig. 8 Tool life of the experiments. **a** Tool life curve of Test 6. **b** Tool life curve of Test 22

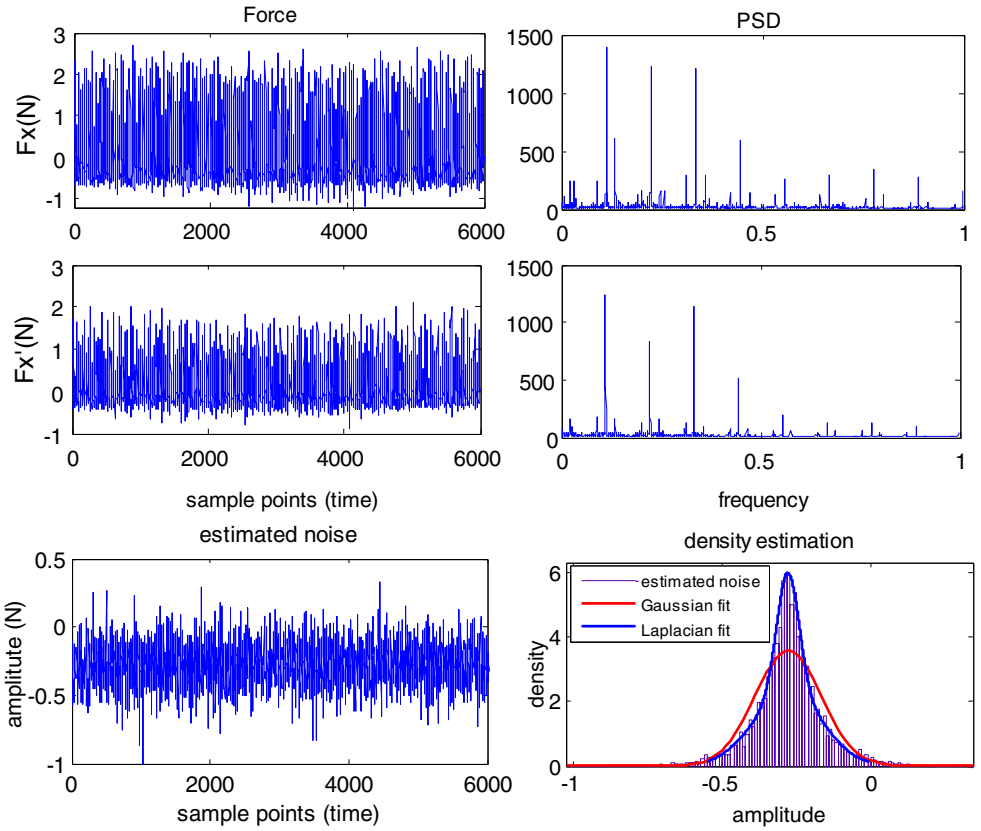

Fig. 9 *Top* the sensor output and their corresponding power spectrum; *Middle* after denoising; *Bottom* the estimated noise density

tabled in [26]. Figure 8 shows the tool life curves of tests of pure copper and steel T4 with coated carbide tool.

4.2 Results and discussions

In this study, a dictionary is learned for every working condition; as a result, there are a total of 27 dictionaries that represent 27 different experiments. The experimental setup is the same as that reported in [26] and listed in the Appendix. Due to the discontinuous cutting nature of micro-milling, the noises are recorded in between every machining pass and during the recess of tool holder. This reference noise and the cutting force from the first cutting pass are applied to learn the dictionary for its corresponding working conditions. While all the results are consistent, the following shows the results from test 22 (Fig. 8b).

In this study, the STFT window is chosen as Gaussian. The window size is selected with 512 points and with 64 points overlapping. The selection of optimal values for λ_1 and λ_2 is a delicate and difficult task. Their values are chosen with empirical studies. The regularization parameter λ_2 is directly related to the sparsity of noise in STFT: the higher the λ_2, the higher the sparsity, λ_2=0.89. The parameter λ_1 is chosen to control the reconstruction: the higher the value of λ_1, the better reconstruction, λ_1=0.41.

Figure 9 shows the force signals and their corresponding power spectrum at severe flank wear state. As can be seen from top figure, the low frequency components have relatively larger peaks, and as a result, the desired frequency components (the harmonics of rotation frequency *fr*) are largely affected. This phenomenon is even worse when the tool wear value is lighter. The denoised force and its corresponding power spectrum are shown in the middle figure. As can be seen from this figure, the minor lobes beyond the characteristic frequencies are largely compressed or eliminated. At the same time, the ratio between the first two characteristic frequencies *fr* and *2fr* is raised, which is an important indication of tool wear conditions.

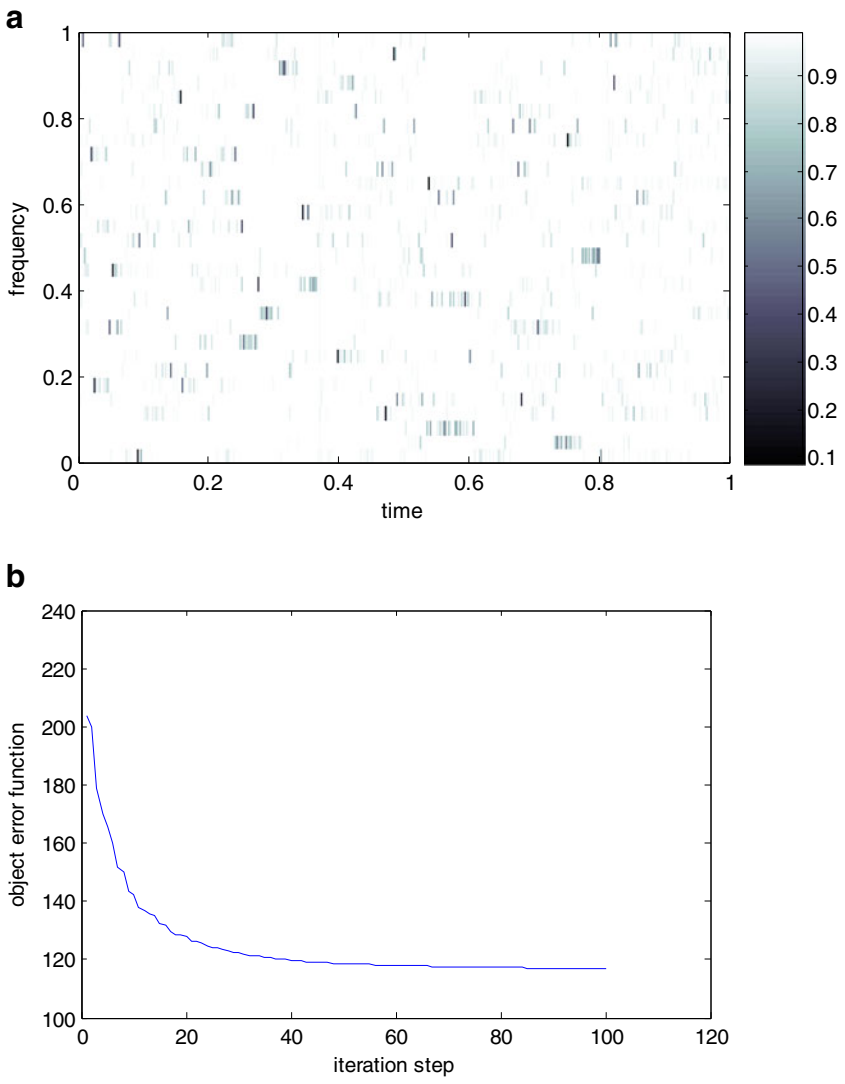

Fig. 10 Sparse representation and dictionary learning. **a** The learned dictionary; note that in this figure, for better demonstration, *darker points* indicate lower values while *lighter point* indicate higher values. **b** The convergence rate of the proposed dictionary learning

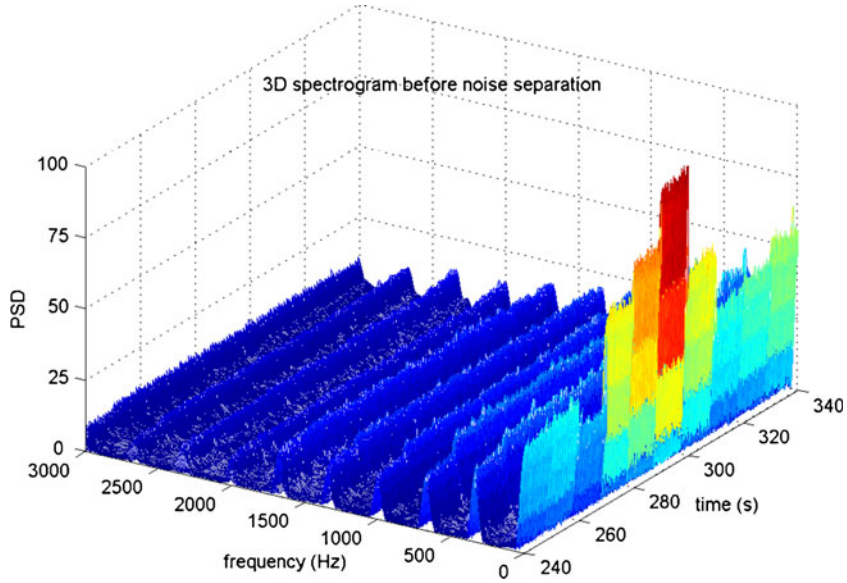

Fig. 11 Force spectrogram before denoising *Fx*

To verify the results, the separated noise is shown in the bottom figure. The statistics of the noise are mean=−0.2813, variance=0.1862, skew=0.0003, and kurtosis=2.8930. This indicates the estimated noise is super Gaussian and having heavy tails. It coincides with the reference noise distribution shown in Fig. 2, with heavy tail distribution fitted well by a Laplace distribution, which has skew=0 and kurtosis=3.0. Meanwhile, under the same mean and variance fit, the noise is far from Gaussian distribution. The results show that the proposed approach could separate the heavy super Gaussian noise well. This could be a good supplement to most current denoising algorithms that based on the Gaussian noise assumption. The underlying assumption of this approach, the noise having sparse representation in the frequency domain, is a common phenomenon and adaptable to most machining force signals.

On the other side, it has been found that the learned dictionaries are quite dense and it has more than 95 % nonzero elements (Fig. 10a). The learned dense dictionary reflects the idea of sparse coding with highly redundant dictionary, and its corresponding dense basis makes it possible for the discrimination between different signals with various elements. Figure 10b shows that the algorithm is fast and converges quickly.

To demonstrate the denoising effect, Fig. 11 and 12 show the force spectrogram before and after denoising. Due to the limit of display memory, only 100 s force signals are displayed. This segmental force is focused in the last few machining passes of test 22 (Fig. 8b) before tool is severely worn. As can be seen from Fig. 11, the force harmonics are completely immersed within the noise and the difference

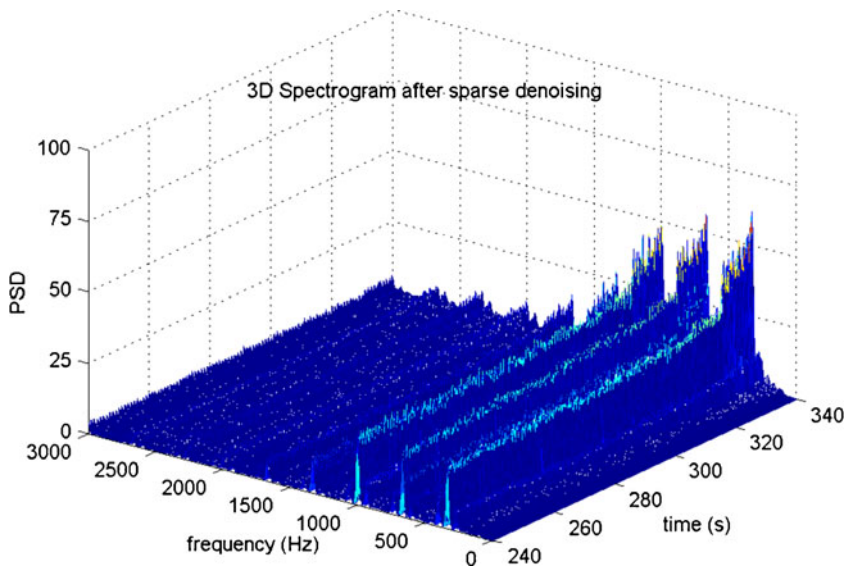

Fig. 12 Tool wear with sparse STFT denoised force

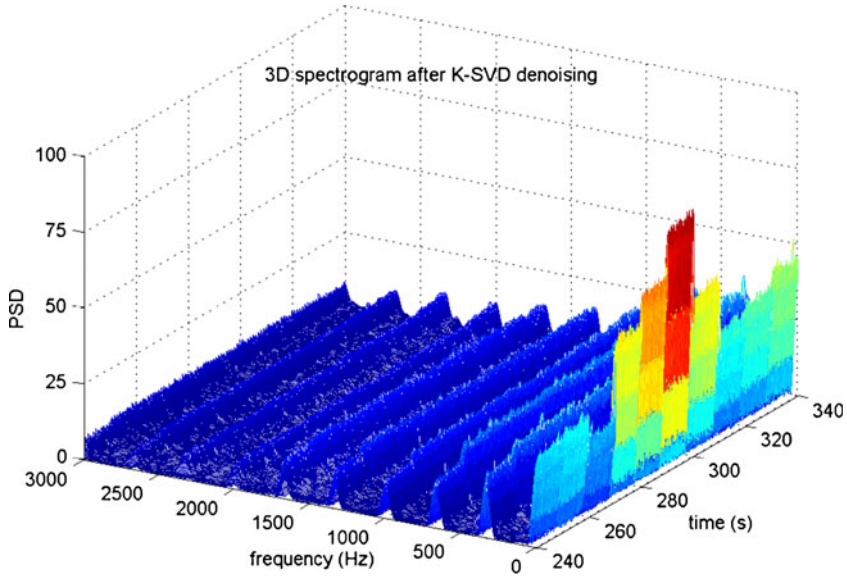

Fig. 13 Tool wear with K-SVD denoised force

cannot be discriminated from the figure when the tool flank wear dramatically increases. The sparse STFT denoised force spectrogram shows much better properties in Fig. 12. It is observed that after 80 s, 1*Fr*, 2*Fr*, and 3*Fr* (*Fr*=300 Hz) of the reconstructed force increased dramatically, corresponding to the sharp increase of tool flank wear in Fig. 8b) around 320 s. Comparing Fig. 12 to 11, it is clear that Fig. 12 shows better diagnostic information since the harmonics stand out with the first three harmonics increasing dramatically according to the fast increase in flank wear.

The K-SVD approach, which was originally applied to remove additive Gaussian noise in image [44], is also adapted to this denoising problem. The parameters are optimized, and the denoised force STFT is shown in Fig. 13. As can be seen from the figure, the significant low frequency noises are not removed. Basically, the K-SVD approach removes Gaussian noise and all the harmonics are compressed almost evenly. The denoising performance of K-SVD is essentially similar to that of wavelet thresholding approach, which is related to the current approach as discussed in the next section.

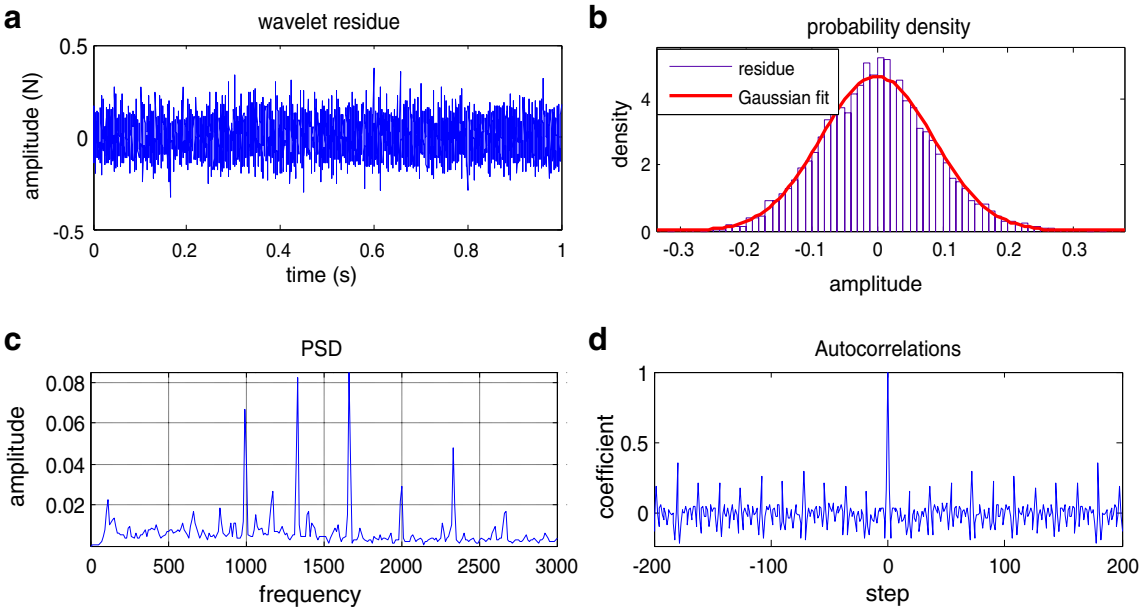

Fig. 14 Illustration of the of wavelet residue: **a** residue (noise), **b** noise distribution compared with Gaussian distribution, **c** the corresponding power spectrum density, **d** the autocorrelation coefficients

On the other hand, besides denoising, the proposed algorithm can also be applied to tool condition estimation. The aim of this research is TCM in micromachining, and denoising is only the first step. In fact, the proposed sparse decomposition algorithm can be further slightly modified to learn various dictionaries from different tool conditions and these dictionaries are constrained to be discriminant in the learning process. The learnt basis can be applied to the tool state estimation when the force features are matched to the respective dictionaries. The detail study of this approach is to be presented in another work.

4.3 The proposed approach and its relation to other methods

4.3.1 Sparse decomposition and its relation to wavelet thresholding for noisy data

An important feature of compressed sensing is that it is robust to noise. For a signal contaminated with noise, the signal model is given by

$$y = Ax + w \tag{19}$$

where w represents noise

In the theory of sparse decomposition, signal x in (19) may be estimated from noisy measurement y by solving the convex minimization problem, as follows:

Minimize $\|x\|_1$,

subject to $\|Ax-y\|_2 \leq \varepsilon$,

Where ε is bound of the amount of noise in the data.

The above convex minimization problem is in fact the same as formula (8), the Lasso problem. Under a special condition, when the noise is assumed both Gaussian and white, a proper choice of ε leads to the celebrated wavelet thresholding approach [45]. It is a special form of this study in case the first term of formula (18) diminishes. In this study, the noise is neither Gaussian nor white and it is not effective with wavelet thresholding.

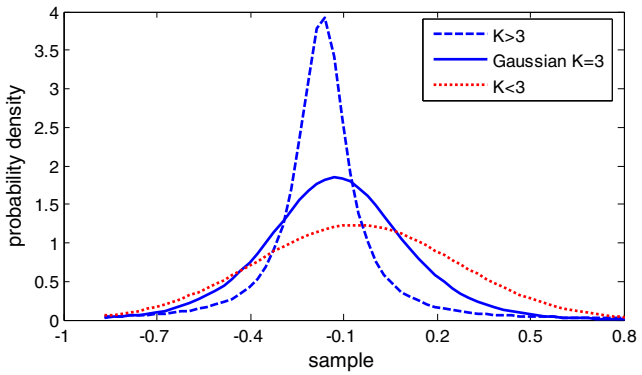

Fig. 15 Different distributions together with their kurtosis measure

Figure 14 shows the residue of the denoised force with wavelet thresholding, which is the noise separated. From Fig. 14c, the PSD harmonics distribute nearly evenly in all frequency bands; this explains why it cannot be separated even in the low frequency noise environmental. In Fig. 14d, the autocorrelation is close to zero; this means that the estimated noise is purely random, not as shown in Fig. 14d. Figure 14b shows that it is typically a Gaussian noise. The Donoho's universal thresholding is not as effective in non-Gaussian noise separation. The noise is identified to be Laplace distribution in this study, and under this condition, the wavelet decomposition coefficients for noise are not evenly distributed among all scales as Gaussian noise. As a result, some of the noise coefficients are not small in certain scales and the threshold is too low under this condition. This is similar to the findings in [46].

4.3.2 Sparse representation and ICA

In the ICA approach, the extracted signals are measured by their Gaussianity (kurtosis). Super Gaussian is defined as with normalized kurtosis K>3 (see [47] for the definition of kurtosis). The intuitive idea is that the super-Gaussian density has heavy tails and a peak at mean. The data distribute densely around the mean while most other areas have small values or close to zero. This property is in fact closely related to the definition of sparseness. Figure 15 shows an example of probability density functions with different values of kurtosis. Kurtosis can be used to check if a distribution is Gaussian. For a symmetrical distribution, super Gaussian has kurtosis higher than 3, while sub-Gaussian lower than 3. In [37], Daubechies et al. have found that the solution to the ICA approach is in fact the sparse solution.

5 Conclusion and future works

In micro-milling, there is machining noise that is relatively heavy with non-Gaussian distribution and has wide harmonics. This study has developed a denoising approach that imposes the constraints on the signal and force's representation in the time–frequency domain. The results show that the developed sparse coding approach separates the non-Gaussian noise, and the noise is found to be consistent with the reference noise. The approach has potential in any non-Gaussian noise separation provided it has sparse representation in the time–frequency domain. Further studies will be carried out on the approach to other precision machining types such as micro-drilling and micro-turning and on how to simultaneously denoise signal and learn different dictionaries for tool condition estimation.

Acknowledgments The first author would like to thank the Alexander Humboldt Foundation of Germany for its support of the research. The authors would also like to thank National University of Singapore for the permission to use the data.

Appendix

Table 3 The micromachining conditions in this study [26]

Experiment setup

Test	Material	Tool diameter (μm)	Spindle speed (rpm)	Depth of cut (μm)	Radial depth of cut (μm)	Feed rate (mm/min)
1	Copper	500	20,000	30	75	80
2	Copper	500	20,000	30	75	120
3	Copper	500	20,000	40	75	100
4	Copper	500	20,000	60	75	120
5	Copper	500	20,000	60	75	150
6	Copper	500	20,000	80	75	150
7	Copper	500	20,000	80	75	180
8	Copper	500	18,000	60	75	120
9	Copper	500	18,000	80	150	150
10	Copper	500	18,000	100	150	150
11	Copper	500	18,000	100	225	150
12	Copper	500	18,000	100	225	180
13	Copper	500	18,000	120	225	150
14	Copper	500	18,000	100	300	180
15	Copper	500	18,000	120	300	180
16	Copper	500	18,000	150	225	150
17	Copper	500	18,000	120	300	150
18	Steel	800	18,000	80	240	120
19	Steel	800	18,000	80	80	100
20	Steel	800	18,000	80	120	100
21	Steel	800	18,000	80	80	120
22	Steel	800	18,000	80	100	100
23	Steel	800	18,000	80	80	80
24	Steel	800	18,000	60	80	80
25	Steel	800	18,000	60	80	60
26	Steel	800	18,000	60	60	60
27	Steel	800	18,000	60	80	100

References

1. Dornfeld D, Min S, Takeuchi Y (2006) Recent. Adv Mech Micromachining, CIRP Ann–Manuf Technol 55(2):745–768
2. Teti R, Jemielniak KO, Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. CIRP Ann–Manuf Technol 59(2):717–739
3. Abellan-Nebot JV, Subirón FR (2010) A review of machining monitoring systems based on artificial intelligence process models. Int J Adv Manuf Technol 47(1–4):237–257
4. Deiab I, Assaleh K, Hammad F (2009) On modeling of tool wear using sensor fusion and polynomial classifiers. Mech Syst Signal Process 23(5):1719–1729
5. Ai CS, Sun YJ, He GW, Ze XB, Li W, Mao K (2012) The milling tool wear monitoring using the acoustic spectrum. Int J Adv Manuf Technol 61(5–8):457–463
6. Cuneyt Aliustaoglu H, Ertunc M, Ocak H (2009) Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system. Mech Syst Signal Process 23(2):539–546
7. Jemielniak K, Arrazolain PJ (2008) Application of AE and cutting force signals in tool condition monitoring in micro-milling. CIRP J Manuf Sci Technol 1(2):97–102
8. Cho S, Binsaeid S, Asfour S (2010) Design of multisensor fusion based tool condition monitoring system in end milling. Int J Adv Manuf Technol 46(5–8):681–694
9. Bhattacharyya P, Sengupta D, Mukhopadhyay S (2007) Cutting force-based real-time estimation of tool wear in face milling using a combination of signal processing techniques. Mech Syst Signal Process 21(6):2665–2683
10. Zhu KP, Hong GS, Wong YS (2011) Multiscale singularity analysis of cutting forces for micromilling tool-wear monitoring. IEEE Trans Ind Electron 58(6):2512–2521
11. Hsieh Wan-Hao L, Ming-Chyuan CS-J (2012) Application of backpropagation neural network for spindle vibration-based tool wear monitoring in micro-milling. Int J Adv Manuf Technol 61(1–4):53–61
12. Huang SN, Tan KK, Wong YS, Silva CW, Goh HL, Tan WW (2007) Tool wear detection and fault diagnosis based on cutting force monitoring. Int J Mach Tool Manuf 47(3–4):444–451
13. Palanisamy P, Rajendran I, Shanmugasundaram S (2008) Prediction of tool wear using regression and ANN models in end-milling operation. Int J Adv Manuf Technol 37(1–2):29–41
14. Tobon-Mejia DA, Medjaher K, Zerhouni N (2012) CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. Mech Syst Signal Process 28(4):167–182
15. Tansel IN, Rodriguez O, Trujillo M, Paz E, Li W (1998) Micro-end-milling—I. wear and breakage. Int J Mach Tools Manuf 38(12):1419–1436
16. Zhang JZ, Chen JC (2008) Tool condition monitoring in an end-milling operation based on the vibration signal collected through a microcontroller-based data acquisition system. Int J Adv Manuf Technol 39(1–2):118–128
17. Malekian M, Park SS, Jun Martin BG (2009) Tool wear monitoring of micro-milling operations. J Mater Process Technol 209(10):4903–4914
18. Sevilla-Camacho P, Herrera-Ruiz G, Robles-Ocampo J, Jáuregui-Correa J (2011) Tool breakage detection in CNC high-speed milling based in feed-motor current signals. Int J Adv Manuf Technol 53(9):1141–1148
19. Afazov SM, Ratchev SM, Segal J, Popov AA (2012) Chatter modeling in micro-milling by considering process nonlinearities. Int J Mach Tool Manuf 56(5):28–38
20. Shi Y, Mahr F, von Wagner U, Uhlmann E (2012) Chatter frequencies of micromilling processes: influencing factors and online detection via piezoactuators. Int J Mach Tool Manuf 56(5):10–16
21. Bisu CF, Zapciu M, Cahuc O, Gérard A, Anica M (2012) Envelope dynamic analysis: a new approach for milling process monitoring. Int J Adv Manuf Technol 62(5–8):471–486
22. Yun HT, Heo S, Lee MK, Min B-K, Lee SJ (2011) Ploughing detection in micromilling processes using the cutting force signal. Int J Mach Tool Manuf 51(5):377–382
23. Ryu SH (2012) An analytical expression for end milling forces and tool deflection using Fourier series. Int J Adv Manuf Technol 59(1–4):37–46

24. Jung S-C, Kim M-S, Yang Y (2012) Baseband noise reduction method using captured TX signal for UHF RFID reader applications. IEEE Trans Ind Electron 59(1):592–598
25. Rosado-Munoz A, Bataller-Mompean M, Soria-Olivas E, Scarante C, Guerrero-Martinez JF (2011) FPGA Implementation of an adaptive filter robust to impulsive noise: two approaches. IEEE Trans Ind Electron 58(3):860–870
26. Zhu KP, Wong YS, Hong GS (2009) Multi-category micro-milling tool wear monitoring with continuous hidden Markov models. Mech Systems Signal Process 23(2):547–560
27. Serviere C, Fabry P (2004) Blind source separation of noisy harmonic signals for rotating machine diagnosis. J Sound Vib 272(1–2):317–339
28. Zhu KP, Hong GS, Wong YS, Wang WH (2008) Cutting force denoising in micro-milling tool condition monitoring. Int J Prod Res 46(16):4391–4408
29. Mallat S. G. (2008) A wavelet tour of signal processing: the sparse way, 3rd edition, Academic Press
30. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306
31. Plumbley MD, Blumensath T, Daudet L, Gribonval R, Davies ME (2010) Sparse representations in audio and music: from coding to source separation. Proc IEEE 98(6):995–1005
32. Aharon M, Elad M, Bruckstein AM (2006) The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representations. IEEE Trans Signal Process 54(11):4311–4322
33. Zibulevsky M, Pearlmutter BA (2001) Blind source separation by sparse decomposition in a signal dictionary. Neural Comput 13(4): 863–882
34. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2008) Supervised dictionary learning. the Neural Inf Process Systems (NIPS) 21:1033–1040
35. Huang K, Aviyente S (2007) Sparse representation for signal classification. The Neural Inf Process Systems (NIPS) 19:609–617
36. Yang M, Zhang L, Feng X, Zhang D (2011) Fisher discrimination dictionary learning for sparse representation. IEEE Int Conf Comput Vis (ICCV) 2011:543–550
37. Daubechies I, Roussos E, Takerkart S, Benharrosh M, Golden C, D'Ardenne K, Richter W, Cohen JD, Haxby J (2009) ICA Algorithms for fMRI do NOT select for independence. Proc Nat Acad Sci (PNAS) 106(26):10415–10422
38. Debnath L., Mikusinski P. (2005) Introduction to Hilbert Spaces with Applications, 3rd Edition, Academic Press
39. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Royal Statist Soc B 58(1):267–288
40. Efron B, Johnstone I, Hastie T, Tibshirani R (2002) Least angle regression. Ann Stat 32(2):407–499
41. Candès E, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52(2):489–509
42. Cohen A, Dahmen W, DeVore R (2009) Compressed sensing and best k-term approximation. J Am Math Soc 22(1):211–231
43. Theodoridis S., Koutroumbas K. (2003) Pattern Recognition, Academic Press
44. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(12):3736–3745
45. Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 90(432):1200–1224
46. Averkamp R, Houdr C (2003) Wavelet thresholding for non-necessarily Gaussian noise: idealism. Ann Stat 3(1):110–151
47. Hyvarinen A., Karhunen J., Oja E. (2001) Independent Component Analysis, Wiley