Denoising Raman spectra by Wiener estimation with a numerical calibration dataset

YANRU BAI AND QUAN LIU*

School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore
*quanliu@ntu.edu.sg

Abstract: Most denoising methods that are currently used in the processing of Raman spectra require significant user interaction in order to optimize their performance across a range of signal-to-noise ratios. In this study, we proposed a method based on the principle of spectral integration followed by Wiener estimation using a numerical calibration dataset, which eliminates the need of experimental measurements for calibration as in the previous Wiener estimation based denoising method. The new method was tested on three types of samples, including a phantom sample, human fingernail and leukemia cells. Compared to two common denoising methods, i.e. moving-average filtering and Savitzky-Golay filtering, the performance of the proposed method is significantly less sensitive to the choices of parameters. Moreover, this method provides comparable or even better denoising performance in the cases with low signal-to-noise ratios.

© 2019 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Raman spectroscopy has been well accepted as a non-destructive analytical tool for chemical and biological analysis in the last few decades. It is a powerful technique for both qualitative and quantitative characterization of materials including biological samples. However, the Raman signal is inherently weak, making interesting Raman peaks susceptible to noise, especially in biological samples. There are two major sources of noise: the camera and signal itself (shot noise) [1]. As one of the most important preprocessing steps, noise reduction is usually performed before any subsequent Raman spectral analysis. The reduction of noise and recovery of the clean Raman signal from Raman measurements with low signal-to-noise ratio (SNR) is a fundamental task to obtain the valuable characteristics of the sample under study.

Smoothing algorithms are commonly used to reduce noise; however, spectral information might be lost in the smoothing process because smoothing removes noise by essentially reducing the spectral resolution of original measurements [2]. Among various smoothing methods, moving-average (MA) filtering is one of the simplest smoothing algorithms where the value of each spectral point is replaced by the average value of all points in a predefined spectral window centred on it. However, it works well only for those spectra with very broad features [2], and the performance is highly dependent on the proper selection of window length. Another commonly used smoothing method for reducing noise in Raman spectra is the Savitzky-Golay (SG) algorithm [1,3]. SG smoothing is conducted by fitting each segment of the original Raman spectrum within a specified window to a polynomial function [4]. The main performance index of the SG filter is determined by the polynomial order and window length [5]. Although this method is useful in noise reduction, it may also suffer from the degradation of the underlying Raman spectral features [1,5–7]. Particularly, a short window length results in a poor smoothing performance, while a long window length may degrade the spectral resolution due to distortion of the weak spectral features. Therefore, a prudent selection of the polynomial order and window length is of great importance to ensure a tradeoff between denoising performance and spectral resolution, which highly depends on a user’s experience. Another category of smoothing methods such as Fourier transform (FT) filtering [8] and wavelet transformation [9–12] are conducted in the frequency domain. These methods decompose a noisy Raman spectrum into high- and low-frequency
components in the frequency domain, making use of the fact that the high-frequency components are regarded to be contributed mainly by noise, and they are truncated with a low-pass filter for the denoising purpose. However, over-filtering or under-filtering are the main drawbacks to FT and wavelet filtering when it is not easy to separate the frequency components of Raman spectrum and noise [13]. In addition, FT or wavelet filtering usually introduces some subjectivity into the analysis because their performance heavily depends on the choices of several complex parameters [8,13], for instance, the composition of wavelet filters and threshold values in wavelet filtering. In sum, most existing denoising methods such as those mentioned above demand for significant experiences from users to ensure good and consistent performance for Raman spectra with different signal-to-noise ratios (SNR). Therefore, it is desirable to develop a denoising method for which the performance is less sensitive to parameter choices and still decent even for those parameter values in the margins of the usual ranges.

In our previous study [14], the clean Raman spectra have been successfully reconstructed from the original Raman measurements of phantoms and bacteria samples with extremely low SNR using the principle of spectral integration followed by Wiener estimation (WE). Based on the minimum mean square error (MMSE) criterion, Wiener estimation is capable of estimating the corresponding high-dimensional data from low-dimensional measurements. However, one significant disadvantage of this method is that a dataset taken from calibration samples similar to the test samples in Raman characteristics is required in Wiener estimation, which greatly limits the application of this method. To get rid of this limitation, later we came up with a method to recover a Raman spectrum using a universal calibration dataset that consisted of Raman spectra measured from each basic biochemical component in the test samples instead of the actual samples [15]. For example, a calibration dataset composed of Raman spectra of proteins, lipids and carbohydrates has been created to recover the Raman spectra of K562 cells and Caki-2 cells. The use of the universal calibration dataset has showed reconstruction performance comparable to that using the conventional calibration dataset measured from the actual biological samples. The Raman spectra of cells and biological fluid are successfully reconstructed using the universal calibration dataset based WE without sacrificing spectral resolution. Because of common biochemical components in most biological samples, the proposed universal calibration dataset, which requires only a limited number of Raman measurements, is applicable to any samples with similar Raman property and biochemical composition.

To further simplify the procedure and expand the universality of the WE based denoising method, in this study, we attempt to create a numerical calibration dataset by considering the common characteristics of Raman spectra such that the method does not require any experimental measurements, and is generally applicable to Raman spectra measured from any samples. For the purpose of denoising, the numerical calibration dataset includes both clean Raman spectra and noisy Raman spectra, in which the noisy spectra are created by adding random noise to the clean spectra. Principle component analysis (PCA) is carried out on the noisy spectra, which partially plays a role of denoising while reducing dimensionality. The Wiener matrix is created from the numerical calibration dataset, which associates the clean Raman spectra with the PC scores obtained by PCA from the corresponding noisy Raman spectra.

Experimental Raman measurements of three types of samples, i.e. a phantom sample (monosodium phosphate) and biological samples (human fingernail and leukemia cells) with various noise levels are used to evaluate the method. Raman spectra with low SNR are obtained by exciting samples sequentially with a series of laser power values during measurements, and a reference spectrum with high SNR is also obtained for comparison using a large laser power and long exposure time with multiple accumulations. Performance comparison in terms of relative root mean square error (rRMSE) between the denoised spectrum and reference spectrum is conducted with the other two commonly used denoising methods, i.e. Savitzky-Golay filtering (SG) method and moving-average filtering (MA). The results indicate that the proposed method
can recover Raman spectra from noisy measurements with different noise levels, and the method is demonstrated to be significantly less sensitive to the selection of involving parameters than the common SG method. The proposed method provides denoising performance comparable to the SG method and surpasses the MA method especially in low SNR cases.

2. Methods and materials

In this study, we define the SNR as in Eq. (1) to evaluate the level of noise in a spectrum:

\[
SNR = \frac{E}{\sigma}
\]  

\[
E = \sqrt{\frac{1}{N} \sum_{i=1}^{N} S_{\text{Ref}}(\lambda_i)^2}
\]

\[
\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [S_{\text{Noisy}}(\lambda_i) - S_{\text{Ref}}(\lambda_i)]^2}
\]

where \(E\) refers to the root-mean-square intensity of the reference spectrum, which is supposed to be free from noise, calculated according to Eq. (2), and all \(S_{\text{Ref}}(\lambda_i)\) refers to the spectral intensity at wavenumber \(\lambda_i\) (\(i = 1, \ldots, N\)) in the reference spectrum. \(\sigma\) is the root-mean-square-error of the noisy Raman spectrum relative to the reference Raman spectrum calculated according to Eq. (3), where \(S_{\text{Noisy}}(\lambda_i)\) is the intensity values of the noisy Raman spectrum at the wavenumber \(\lambda_i\). The fluorescence background is not removed from the reference and noisy Raman spectra prior to the SNR calculation; however, their intensity values are scaled such that the maximum and minimum values in each spectrum fall in the same range.

2.1. Spectral reconstruction by Wiener estimation based on a numerical calibration dataset (NCD-WE)

Denoising of Raman measurements using Wiener estimation involves two stages. The first stage is to create a Wiener matrix from the calibration dataset, i.e. the calibration stage, and the second stage is to reconstruct a denoised spectrum from noisy Raman measurements in the test dataset using the obtained Wiener matrix, i.e. the reconstruction stage.

Stage 1: Calibration

In this work, the calibration dataset is designed to include two sets of spectra, i.e. clean spectra and noisy spectra. The set of clean spectra is made up of a series of basic components (BCs) spectra and each component spectrum contains a single peak numerically generated using a Gaussian function as shown in Eq. (4). The wavenumber range and increment in the basic components spectra should be identical to those of the measured spectra in the test dataset.

\[
I = I_{\text{max}} \times e^{-\frac{(\lambda - \lambda_0)^2}{2\sigma^2}}
\]

where \(I_{\text{max}}\) is the maximum peak intensity with two possible values, i.e. 1 and 2. \(\lambda\) refers to wavenumber \((\text{cm}^{-1})\), the range of which depends on that of the spectra to be reconstructed in the test dataset. \(\lambda_0\) refers to the central peak position. \(\sigma\) is related to bandwidth varied within the interval 10-30 cm\(^{-1}\) with an increment of 5 cm\(^{-1}\).

In addition, the set of noisy spectra in the calibration dataset is obtained by adding one of two kinds of random noises to each clean spectrum: 1) Gaussian noise (GN) spectrum generated using the function ‘awgn’ in MATLAB, yielding an SNR range of 1.51 to 1.77 for the noisy spectra in the calibration dataset according to the definition in Eq. (1); 2) Poisson noise (PN) spectrum generated using the function ‘poissrnd’ function in MATLAB, where the original
clean spectrum after scaling with respect to intensity is used as the rate parameter to specify the Poisson distribution at each data point. The scaling factor is set to $10^2$ so that the SNR range of noisy spectra in the calibration dataset is varied from 1.26 to 3.09 according to the definition in Eq. (1). Although Gaussian noise is commonly used as random noise model, the Poisson noise model has been confirmed to be more suitable in Raman spectroscopy [16]. Therefore, we considered both cases to be comprehensive. In addition, we have verified that no significant difference exists in the denoising performance of our method as long as the SNR of the noisy spectra in the calibration dataset is lower than those of the spectra in the test dataset.

Then, principal-component analysis (PCA) is performed for the noisy spectra in the calibration dataset according to Eq. (5). The process compresses the spectra to reduce data dimension.

$$D_c = F S_{c,L}$$  \hspace{1cm} (5)

where $S_{c,L}$ is an $(m \times n)$ matrix of the noisy Raman spectra with low SNR, $m$ represents the number of spectra, and $n$ represents the number of discrete wavenumbers in the Raman spectrum. Subscript $c$ refers to the calibration dataset and $L$ refers to low SNR. The $(n \times l)$ matrix $F$ is the transmission spectra of principal components (PCs) based filters, and $l$ is the number of PCs. A truncated $(m \times l)$ matrix $D_c$, denoted as the PC score matrix, is thus obtained by mapping the original spectra matrix of $n$ variables to a new space of $l$ variables. Matrix $D_c$ will be used for Wiener estimation in the reconstruction later.

Subsequently, Wiener matrix $W$ is created from the calibration dataset according to Eq. (6) [17].

$$W = E[S_{c,H}D_c^T] \{E[D_cD_c^T]\}^{-1}$$  \hspace{1cm} (6)

where $S_{c,H}$ presents the set of clean spectra with high SNR in the calibration dataset, and $D_c$ is the PC score matrix obtained earlier according to Eq. (5). $E[\cdot]$ function refers to ensemble average, and the superscripts $T$ and $-1$ represent matrix transpose and matrix inverse, respectively. The Wiener matrix $W$ associates the clean Raman spectra with the PC scores obtained from the corresponding noisy Raman spectra by PCA, which partially contributes to the denoising power of our method.

As described above, the established numerical calibration dataset is made up of a series of clean and noisy spectra each with a single Gaussian peak at the same peak location. One of the general accepted assumptions in the quantitative analysis of Raman spectroscopy is that the spectrum of the sample is the linear summation of spectra of all basic components present in the sample [18]. Simplified the assumption to a linear model, a spectrum with multiple peaks can be modelled as the linear summation of several spectra with single peak after proper spectral shifts. It can be predicted that Wiener matrices obtained from calibration datasets in which the spectra possess different central peak locations are identical in column vectors except the differences in central peak location. Therefore, the Wiener matrix obtained from a calibration dataset with multiple peaks in spectra is equivalent to the linear summation of all Wiener matrices obtained from a series of calibration datasets each with a single peak in spectra after the peaks in the column vectors of the Wiener matrices are spectrally shifted along the spectral dimension properly. Figure 1 demonstrates two Wiener matrices obtained from calibration datasets with $\lambda_1$ and $\lambda_2$ as the central peak locations, respectively. The upper panel shows the clean and noisy spectra in the calibration dataset, and the obtained Wiener matrix according to Eq. (6) is presented column-wise in the lower panel, where each curve corresponds to one column vector of the Wiener matrix.

Stage 2: Reconstruction

In the reconstruction stage, the corresponding denoised Raman spectra are reconstructed from Raman spectra measured with low SNR in the test dataset using the Wiener matrix obtained in the calibration stage. This stage involves two steps. First, dimension reduction using PCA
is performed to the low-SNR Raman measurement to yield the PC score matrix, according to Eq. (7).

\[ D_t = FS_{t, L} \]

where \( S_{t, L} \) is a \((m \times n)\) matrix of the low-SNR Raman spectra, \( m \) represents the number of spectra, and \( n \) represents the number of discrete wavenumbers in the Raman spectrum. Subscript \( t \) refers to the test dataset and \( L \) refers to low SNR. \( F \) is the transmission spectra of PCs based filters, and it should be noted that the number of PCs is the same as that of the calibration dataset. PC score matrix \( D_t \) will be used for Wiener estimation in the following step.

Then, the reconstruction is performed step by step using the Wiener matrix obtained from the calibration dataset after spectral shifting with a shifting increment of 1 cm\(^{-1}\). Take the central location \( \lambda_j \) as an example, the reconstructed spectrum \( \hat{S}_{\lambda_j} \) is obtained according to Eq. (8), then only values falling in a spectral window with a length \( L \) and a central location of \( \lambda_j \) are retained, all values outside the window are set to 0. Issues regarding the determination of a proper window length \( L \) will be discussed in the subsequent section below.

\[ \hat{S}_{\lambda_j} = W_{\lambda_j}D_t \]

where \( D_t \) is the PC scores of the test dataset calculated according to Eq. (7). \( W_{\lambda_j} \) refers to Wiener matrix obtained from the calibration dataset with a central location at \( \lambda_j \). The final reconstructed spectrum is achieved by averaging all \( \hat{S}_{\lambda_j} \) as shown in Eq. (9)

\[ \hat{S} = \frac{1}{N} \sum_{j=1}^{N} \hat{S}_{\lambda_j} = \frac{1}{N} \sum_{j=1}^{N} W_{\lambda_j}D_t \]

where \( \hat{S} \) is the reconstructed spectrum and \( N \) refers to the total number of shifted Wiener matrix. \( N \) should be equal to the range of wavenumber in Raman spectra divided by the shifting increment (1 cm\(^{-1}\) in this work).

2.2. Sample preparation and Raman measurements

In this study, Raman measurements from three samples, a phantom, human fingernail and leukemia cells, were utilized to validate the effectiveness of the numerical calibration dataset.
based wiener estimation method (NCD-WE) for noise reduction. The results of the proposed method were compared with two traditional denoising techniques, i.e. Savitzky-Golay filtering (SG) and moving-average (MA) filtering. Parameter determination of the three methods will be discussed in the first part of “Results” section below.

The phantom was made by dissolving monosodium phosphate (20233-1 KG, Affymetrix, USA) in distilled water to create a solution with a concentration of 1M. The Raman spectra of phantom were measured over a wavenumber range from 600 to 1800 cm$^{-1}$ with an increment of 2 cm$^{-1}$, using a micro-Raman system (InVia, Renishaw, U.K.) coupled to a microscope (Alpha 300, WITec, Germany) with a 20× objective lens. The excitation wavelength was 785 nm (HPNIR785, Renishaw, UK). The maximum laser power reaching the samples was approximately 218 mW. Spectra with four levels of SNR were collected. This was achieved using 10%, 1%, 0.1%, and 0.05% of the maximum laser power, respectively, and the exposure time was kept at 10s. In addition, a reference Raman spectrum (with high SNR) was collected from the sample using 50% of the maximum laser power with an exposure time of 20 s and 3 accumulations.

Raman measurements were taken from a human fingernail sample ex vivo using the same system as above with laser light focusing on the surface of the fingernail sample. Spectra with four levels of SNR were collected. This was achieved by using 1%, 0.5%, 0.01%, and 0.05% of the maximum laser power, respectively, and the exposure time was 10s. A reference Raman spectrum (with high SNR) was also collected from the sample using 50% of the maximum laser power with an exposure time of 20 s and 3 accumulations.

Leukemia cells (CCL-243, USA) were cultured in Iscove’s Modified Dulbecco’s Medium (IMDM) added with 10% fetal bovine serum and incubated under a 37°C with 5% CO₂ circumstance. It should be noted that the cells were washed twice and immersed in PBS before Raman measurements to reduce fluorescence background originated from culture media. Raman measurement of cells was conducted using the same system as above. Spectra with four levels of SNR were collected using 10%, 5%, 1%, and 0.5% of the maximum laser power, respectively, and the exposure time was 10s. A reference Raman spectrum (with high SNR) was also collected from the sample using 50% of the maximum laser power with an exposure time of 60 s and 3 accumulations. The state of the cells was fine according to our visual observation and spectral evaluation. In fact, the integrity of cells has been confirmed in the literature when exposed to 115 mW laser power at 785 nm for more than one hour [19].

The parameters of all experimental measurements and the resulting SNRs of measured spectra are summarized in Table 1. The noisy spectra in the test dataset with four noise levels are denoted as P1 – P4 for the phantom sample, F1 – F4 for fingernail, and L1 – L4 for leukemia cells, respectively, hereinafter.

2.3. Performance assessment

The performance of noise reduction is evaluated according to the relative root mean square error (rRMSE) as defined in Eq. (10).

$$rRMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{S}(\lambda_i) - S_{ref}(\lambda_i))^2}{N \times \max[S_{ref}(\lambda_i)]^2}}$$  \hspace{2cm} (10)

where $\hat{S}$ is the reconstructed spectrum from the Raman measurement with low SNR in the test dataset. $S_{ref}$ refers to the reference Raman spectrum with high SNR. $\lambda_i$ refers to the i-th wavenumber, and the function $\max[\cdot]$ provides the maximum value in the entire spectrum.

It should be noted that the intensity values of both the reconstructed spectra and the reference spectrum went through z-score normalization to have a mean of 0 and a standard deviation of 1 before calculating rRMSE. The purpose of normalization is to minimize the difference in spectral
Table 1. Parameters of experimental measurements and SNR of test Raman spectra

| Sample    | Test | SNR  | Laser power (percent of maximum) | Exposure time (s) | Accumulation |
|-----------|------|------|----------------------------------|-------------------|--------------|
| Phantom   |      |      |                                  |                   |              |
| Reference | N.A. | 50%  |                                  | 20                | 3            |
| P1        | 25.77| 10%  |                                  | 10                | 1            |
| P2        | 11.77| 1%   |                                  | 10                | 1            |
| P3        | 3.57 | 0.1% |                                  | 10                | 1            |
| P4        | 2.51 | 0.05%|                                  | 10                | 1            |
| Fingernail|      |      |                                  |                   |              |
| Reference | N.A. | 50%  |                                  | 20                | 3            |
| F1        | 17.31| 1%   |                                  | 10                | 1            |
| F2        | 16.16| 0.5% |                                  | 10                | 1            |
| F3        | 6.30 | 0.01%|                                  | 10                | 1            |
| F4        | 4.60 | 0.05%|                                  | 10                | 1            |
| Leukemia Cells | |      |                                  |                   |              |
| Reference | N.A. | 50%  |                                  | 60                | 3            |
| L1        | 12.49| 10%  |                                  | 10                | 1            |
| L2        | 11.14| 5%   |                                  | 10                | 1            |
| L3        | 7.98 | 1%   |                                  | 10                | 1            |
| L4        | 4.51 | 0.5% |                                  | 10                | 1            |

Intensities due to varied measurement parameters. In addition, fluorescent background was also removed using a wavelet algorithm [20] to facilitate comparison and eliminate the influence of the background.

3. Results

3.1. Optimization of parameters

In addition to NCD-WE, another two widely utilized denoising methods in Raman spectroscopy, Savitzky-Golay filtering (SG) and moving-average filtering (MA), were evaluated on the Raman measurement in low SNR for comparison. Because the performance of a method depends on both the parameters involved and the SNR of the test spectra, we analysed the performance of each method according to the rRMSE of the denoised spectrum relative to the reference spectrum, using a range of values for every parameter for each method. In particular, the parameters involved in NCD-WE include the number of PCs and window length. The SG method also involves two parameters, i.e. polynomial order and window length. It is important to note that SG smoothing algorithm performs properly only if the window length is odd and the polynomial order is designated a value smaller than the window length [6]. In an earlier study [6], a polynomial order varied from 1 to 11 and a window length varied from 3 to 25 were adopted to test the performance of SG smoothing filter for denoising the Raman spectra of non-structural protein 1, a biomarker for flavivirus origin diseases. Results demonstrated that an optimal polynomial order of 3 and a window length of 13 provided the best smoothing effect, meanwhile, preserved most characteristics in the original Raman spectra. Inspired by the observations in this study, a polynomial order ranging from 3 to 9, and a window length ranging from 5 to 95 were considered to ensure that at least one combination of these parameters would yield the optimal performance in denoising spectra with various SNR. For the MA method, window length is the only adjustable parameter [21]. The range and increment of each parameter in every method are summarized in Table 2, which cover most commonly used values.

Figure 2 and 3 display the rRMSE distribution of denoised Raman spectra measured from the phantom sample using the NCD-WE method for the ranges of number of PCs and window length and that using the SG method for the ranges of order and window length, respectively, as
Table 2. Parameters tested for NCD-WE, SG and MA methods

|                | NCD-WE | SG          | MA          |
|----------------|---------|-------------|-------------|
| Number of PCs  | 3–9     | 3–9         | 1–20        |
| Window Length  | 1–20    | 5–95        | 1–20        |
| Order          |         | 1           |             |
| Window Lengtha |         | 6           | 1           |

*a*Note that the window length for the SG method has to be an odd number.

listed in Table 2. The graphs from left to right correspond to P1, P2, P3 and P4, respectively, in both figures. Here, we only demonstrate the rRMSE distribution of NCD-WE method with GN added in the calibration dataset. As for PN, the best and average performance in terms of rRMSE are summarized in Table 3. The window length appears to be an important parameter. The best denoising performance was achieved when the window length values of 5, 7, 11 and 19 were used, respectively, for P1, P2, P3, and P4. However, the method is relatively insensitive to the number of PCs for denoising Raman spectra measured from the phantom sample, especially in the range of 5 to 9. This can be clearly seen from the observation that the maximum rRMSE value is not 86% greater than the minimum rRMSE value at each SNR. In contrast, the performance of SG relies highly on window length especially for lower SNR, e.g. P3 and P4. Only a lower order combined with a longer window length can guarantee a reliable result for P3 and P4, while a short window length (less than 30) works very poorly. There is an obvious trend that a longer window length yields a better result for lower SNR. The maximum rRMSE value can be anywhere from 180% to 406% greater than the minimum rRMSE value across all SNR. Figure 4 displays the rRMSE of denoised Raman spectra measured from the phantom sample using the MA method for a range of window lengths. The lower the SNR, the longer the window length required to yield better denoising performance.

![Fig. 2. Relative RMSE (rRMSE) of denoised Raman spectra measured from the phantom sample using the NCD-WE method (calibration dataset with GN). Graphs from left to right correspond to P1, P2, P3 and P4.](image)

The rRMSE distribution of denoised Raman spectra measured from the human fingernail and leukemia cells using the three methods are not shown here due to limited space. In general, the performance of NCD-WE is relatively consistent in the entire range of selected parameters according to the observation of smaller range of rRMSE values compared to that for the SG method. The two parameters of SG method are dependent with each other resulting a noticeable pattern. For MA method, it is observed that a short window is more suitable for dealing with spectra with high SNR, while a long window yields better performance in low SNR cases.

### 3.2. Denoising performance

The best and average denoising performance based on rRMSE of the NCD-WE, SG and MA methods are summarized in Table 3.
Fig. 3. Relative RMSE (rRMSE) of denoised Raman spectra measured from the phantom sample using the SG method. Graphs from left to right correspond to P1, P2, P3 and P4. Notice the presence of NAN (not a number) values at the right corner is due to the fact that the SG algorithm cannot be properly conducted when the window length is smaller than the polynomial order.

Fig. 4. Relative RMSE (rRMSE) of denoised Raman spectra measured from the phantom sample using the MA method. Graphs from left to right correspond to P1, P2, P3 and P4.

Table 3. Best and average denoising performance (rRMSE) for the ranges of parameters for the NCD-WE (GN and PN), SG and MA methods

| rRMSE($10^{-3}$) | Best | Average |
|------------------|------|---------|
|                  | NCD-WE (GN) | NCD-WE (PN) | SG | MA | NCD-WE (GN) | NCD-WE (PN) | SG | MA |
| Phantom          |      |         |    |    |      |         |    |    |
| P1               | 0.72 | 0.64    | 0.39 | 0.57 | 1.00 | 0.81    | 0.79 | 1.50 |
| P2               | 0.81 | 0.72    | 0.54 | 0.67 | 1.00 | 0.93    | 0.89 | 1.50 |
| P3               | 1.50 | 1.60    | 1.40 | 2.30 | 1.70 | 1.80    | 2.20 | 2.70 |
| P4               | 2.30 | 2.60    | 2.00 | 3.30 | 2.60 | 3.10    | 3.70 | 3.90 |
| Fingernail       |      |         |    |    |      |         |    |    |
| F1               | 1.10 | 1.10    | 1.00 | 1.30 | 1.40 | 1.70    | 1.70 | 2.20 |
| F2               | 1.10 | 1.10    | 0.86 | 1.00 | 1.40 | 1.90    | 1.40 | 1.90 |
| F3               | 1.20 | 1.50    | 1.30 | 1.80 | 1.60 | 2.20    | 2.20 | 2.60 |
| F4               | 1.50 | 1.40    | 1.60 | 2.50 | 1.90 | 2.10    | 2.50 | 3.00 |
| Leukemia Cell    |      |         |    |    |      |         |    |    |
| L1               | 2.02 | 2.09    | 1.79 | 2.41 | 3.00 | 2.90    | 2.70 | 3.50 |
| L2               | 2.44 | 1.89    | 1.79 | 2.26 | 3.10 | 2.60    | 2.90 | 3.30 |
| L3               | 2.58 | 2.60    | 2.61 | 3.93 | 3.20 | 3.80    | 4.00 | 4.80 |
| L4               | 2.33 | 2.46    | 2.87 | 3.89 | 3.00 | 3.30    | 4.30 | 5.00 |

In general, the MA method yields the worst performance in nearly all cases except for P1 and P2 for the phantom sample. The SG method with optimal parameters works better than the NCD-WD method for denoising spectra with higher SNR. However, the NCD-WE method is comparable to or even better than the SG method in denoising spectra with lower SNR. From the average rRMSE data we can see that the performance of NCD-WE is relatively consistent
because its average rRMSE spans a smaller range from SNR 1 to SNR 4 compared to the SG method. Comparing NCD-WE (GN) and NCD-WE (PN), the performance of NCD-WE (PN) is slightly better for the phantom sample and leukemia cells in higher SNR cases (e.g. P1, P2 and L2) but worse in lower SNR cases (e.g. P3, P4, F3 and L4).

Figure 5 presents the best denoised results of Raman spectra measured from the phantom sample using the NCD-WE (GN) method. Graphs from left to right correspond to P1 – P4 respectively. In the upper panel, the blue line shows the low-SNR test spectrum, and the red dot line shows the denoised spectrum. In the lower panel, the grey line and the red dot line display the reference spectrum and the denoised spectrum in both of which fluorescence background has been removed using a wavelet algorithm [20] for easy comparison. It can be observed that the spectral shape and the positions of main peaks are accurately recovered using the NCD-WE method, even in the very low SNR case (P4), where the peaks are overwhelmed by noise in the original test spectrum. Another observation is that the major peak intensity values are slightly underestimated, but this problem could be ameliorated by performing normalization on the denoised spectrum. In addition, some artificial peaks are observed in the relatively flat regions in cases of P3 and P4.

![Fig. 5. Best denoised results of Raman spectra measured from the phantom sample using the NCD-WE (GN) method. Graphs from left to right correspond to P1 – P4, respectively.](image)

The best denoised results of Raman spectra measured from human fingernail and leukemia cells are not shown here due to limited space. Generally, the spectral shape and the main peaks of the spectrum are nicely reconstructed, except the very narrow peak at 1000 cm\(^{-1}\) peak in the spectra of leukemia cells. Only some small spectral features are not accurately recovered.

4. Discussion

In this study, we develop a denoising method using Wiener estimation based on a calibration dataset created numerically rather than experimentally, to overcome the disadvantages of the latter approach in cost, complexity and applicability because experimental measurements of real samples can be costly and the Wiener matrix obtained in this manner is specific to calibration samples. The clean spectra in the numerical calibration dataset is composed of a series of basic component spectra each with a single Gaussian peak possessing a different width. The noisy spectra in the numerical calibration dataset are obtained by adding random noise to the clean spectra. Two types of commonly used noises are taken into consideration, i.e. Gaussian noise (GN) and Poisson noise (PN). An actual spectrum with multiple peaks is treated as the linear summation of several basic components spectra each with a single Gaussian peak after proper shifting in the central location. A similar treatment in which the Raman spectrum of a sample
is the linear summation of the spectra of all basic components has been applied earlier in cell and tissue spectra studies [18,22,23]. Furthermore, we prove that the Wiener matrix obtained from calibration spectra with multiple peaks is equivalent to the linear summation of all Wiener matrices obtained from those calibration spectra each with a single peak after proper peak shifting along the spectral dimension. This finding significantly reduces the complexity and time of generating Wiener matrices.

4.1. Guideline on the parameter selection

Our method involves two parameters that may affect the denoising performance. One is the number of principal components in PCA, and the other is the window size used in the reconstruction stage. The number of principal components determines the percentage of variance explained by the components. Since PCA plays a role in not only dimension reduction but also denoising, the components that capture more variance do not necessarily yield better denoising performance. Similarly, an appropriate choice of the window length would ensure the denoising performance while preventing weak peaks from being smoothed out.

We systematically analysed the denoising performance in terms of rRMSE by varying the two parameters one at a time. The number of PCs ranging from 3 to 9 with an increment of 1 was utilized to test our method. Since the set of noisy spectra in the numerical calibration dataset contains 10 spectra, including five basic components each with a different Gaussian bandwidth and two variants for every basic component, the maximum number of PCs obtained from the PCA of the calibration dataset is 9. The smallest number of PCs was set to be 3 because the first three components account for more than 80% of the total variance. Taking the spectra of the phantom sample reconstructed using GN based calibration dataset as an example, the upper panel of Fig. 6 demonstrates the effect of PCs number on the denoising performance for each SNR case, where the error bar indicates the standard deviation of rRMSE when various window size is used in the reconstruction. It can be seen that performance of 3 PCs is not as good as that of other number of PCs, which may be due to the loss of useful spectral information in the principal component analysis. In general, 5–9 PCs yield equally good denoising performance regardless of the SNR of Raman spectra, so that the number of PCs can be selected freely in this range for denoising Raman spectra with unknown noise level. The same trend can be observed from the data for human fingernail and leukemia cells (results now shown for conciseness).

For selected number of PCs, a series of window length varied from 1 to 20 were used in the reconstruction stage. The lower panel of Fig. 6 demonstrates the effect of window length on the denoising performance for each SNR case, where the error bar indicates the standard deviation of rRMSE when various number of PCs are used. It can be seen that a window length of 5–10 is more suitable in high SNR cases, e.g. P1 and P2, while the denoising performance is less dependent on the selection of window length for low SNR cases generally. Overall, a window length within 5–16 can provide satisfied results for all SNR (P1 – P4) without significant difference. This rule of thumb is also applicable to human fingernail and leukemia cells.

One implication regarding the selection of two parameters is that a user does not have to make a perfect choice on the parameters of the NCD-WE method to obtain average denoising performance for a noisy spectrum with unknown SNR, which establishes a great advantage of our method over existing approaches, such as SG.

4.2. Comparison between GN and PN model

Many studies have been carried out on noise modelling in Raman spectral data. While Gaussian noise model has been widely used [1,24–26], Sompel D Van de et al. reported in an earlier study [16] that noise in Raman spectra follows closer to Poisson distribution than to Gaussian distribution. They developed a new version of ‘hybrid reference spectrum and principle component analysis’ algorithm by incorporating a Poisson noise model. Results indicated that both noise models yield
Effects of number of PCs and window length on denoising performance for spectra with various SNR when applying the NCD-WE (GN) method to Raman spectra from the phantom sample. For each PC number, the $r$RMSE values at all window sizes are analysed to yield the mean and standard deviation (error bar) as shown in the upper panel. For each window length, the $r$RMSE values at all number of PCs are analysed to yield the mean and standard deviation (error bar) as shown in the lower panel.

comparable quantification performance for measured data, nevertheless, the Poisson noise model outperformed the Gaussian noise model consistently for the simulated data. For comparison, we take into account both the Gaussian noise and Poisson noise models when creating the calibration dataset in this study. In Table 3, no significant difference was observed between the two noise models in terms of either the best or the average $r$RMSE values, except that the Poisson model often slightly outperformed in higher SNR cases.

4.3. Comparison between NCD-WE and UCD-WE methods

In addition to modelling a Raman spectrum as the summation of many single-peak spectra as shown in this study, it is also feasible to model it as the linear summation of multiple component spectra assuming that the underlying biochemical components are known, and their Raman spectra have been experimentally measured a priori. These component spectra constitute a universal calibration dataset [15] for Wiener estimation (UCD-WE). We applied the UCD-WE method to denoising and compared the performance of denoising the Raman spectra of leukemia cells between the proposed NCD-WE method and the UCD-WE method (data not shown due to limited space). It was found that the NCD-WE method yields superior performance compared to the UCD-WE method, but the reconstruction of the narrow peak at 1000 cm$^{-1}$ is more accurate using the UCD-WE method due to the involvement of the experimental measurement of albumin and actin in the calibration dataset. Furthermore, the denoising performance of the UCD-WE method is highly dependent on the complete knowledge of the constituent components, which determines its applicability to different types of samples.

4.4. Comparison of the NCD-WE method with two common denoising methods

Comparison is carried out among NCD-WE, SG and MA methods from the perspectives of denoising performance and spectral shape recovery. The best denoised spectra (according to $r$RMSE) reconstructed from the measurements of the phantom sample and leukemia cells using NCD-WE, SG and MA methods are displayed in Fig. 7 and Fig. 8, respectively, where (A) – (D) correspond to P1 – P4 and L1 – L4. In each graph, the black line shows the noisy Raman measurement without any treatment; the red bold line is the reference spectrum; the magenta line
and blue line represent NCD-WE with the GN model and PN model-based calibration datasets, respectively; the orange line and green line are denoised spectra reconstructed using SG and MA methods, respectively. It should be mentioned that the fluorescence background of the reference spectrum and denoised spectra has been removed using a wavelet algorithm [20] for easy comparison. It can be observed that in the high SNR cases, all methods yield good results and the performance degrades as the SNR of the original spectrum decreases. In low SNR cases (P3, P4 and L3, L4), NCD-WE and SG are still capable of recovering the main peaks of the original spectrum accurately, although noises are not completely removed in some relatively flat regions. In contrast, MA works very poorly in these cases; and some spectral regions with large noises are mistakenly retained as peaks, more importantly, there are artificial shifts in the main peaks between the recovered spectrum and the reference spectrum, despite the fact that MA shows advantage over the other two methods in recovering very narrow peak, e.g. the peak of leukemia cells spectra at 1000 cm$^{-1}$, in high SNR cases (L1 and L2).

Fig. 7. Comparison of the best denoised spectra reconstructed from measurements of the phantom sample using NCD-WE, SG and MA methods. (A) – (D) correspond to P1 – P4, respectively. The test spectrum refers to the noisy spectrum with various SNR without fluorescence background removal. The reference spectrum refers to the clean spectrum with fluorescence background removed, which is meant to be compared to the denoised result using each method.

Fig. 8. Comparison of the best denoised spectra reconstructed from measurements of leukemia cells using NCD-WE, SG and MA methods. (A) – (D) correspond to L1 – L4, respectively. The test spectrum refers to the noisy spectrum with various SNR without fluorescence background removal. The reference spectrum refers to the clean spectrum with fluorescence background removed, which is meant to be compared to the denoised result using each method.

The computational time values for denoising a single spectrum using the NCD-WE, SG and MA methods are 0.82s, 2.8×10$^{-3}$s and 7×10$^{-4}$s, respectively. The computational efficiency of the NCD-WE method could be improved by performing the reconstruction using the Wiener matrix (after spectral shifting) with an adaptive shifting increment according to the bandwidth of Raman peaks.

In the future, we believe it is possible to improve the proposed algorithm performance using a multi-scale approach, which is similar to wavelet reconstruction. In this approach, multiple calibration datasets will be created each with a different peak width in the initial single-peak spectrum. The reconstruction will be performed in multiple steps and the calibration datasets
associated with progressively smaller peak widths will be sequentially used in these steps to reconstruct the details of the target spectrum in a series of spectral scales. This approach is expected to yield more accurate details at the cost of longer computation.

5. Conclusion

In this study, we proposed a method based on the principle of spectral integration followed by Wiener estimation using a numerical calibration dataset, which eliminates the need of experimental measurements for calibration as required in the previous Wiener estimation based denoising method. This method was tested on three types of samples, including a phantom sample, human fingernail, and leukemia cells. Compared to two commonly used denoising methods, i.e., moving-average filtering and Savitzky-Golay filtering, the performance of the proposed method is significantly less sensitive to the choices of parameters. Moreover, this method provides comparable or even better denoising performance in the cases with low signal-to-noise ratios.

Funding

Ministry of Education - Singapore (MOE2015-T2-2-112, MOE2017-T2-2-057); Nanyang Technological University (CG - 01/16, NAM/15004); Agency for Science, Technology and Research (H17/01/a0/008, H17/01/a0/0F9).

Acknowledgements

The authors would like to appreciate help from Dr. Clint Perlaki for providing advice on fluorescence background removal using the wavelet algorithm and Dr. Shuo Chen for providing guidance on coding the Wiener estimation in MATLAB.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. S. J. Barton, T. E. Ward, and B. M. Hennelly, “Algorithm for optimal denoising of Raman spectra,” Anal. Methods 10(30), 3759–3769 (2018).
2. N. Tarcea and J. Popp, “Raman data analysis,” in Raman Spectroscopy Applied to Earth Sciences and Cultural Heritage, J. Dubessy, M. C. Caumon, and F. Rull, eds. (Mineralogical Society of Great Britain and Ireland, 2012), pp. 193–226.
3. A. Kwiatkowski, M. Gnyba, J. Smulko, and P. Wierza, “Algorithms of Chemicals Detection Using Raman Spectra,” Metro. Meas. Syst. 17(4), 549–559 (2010).
4. M. Villarroel, P. Barreiro, P. Kettlewell, M. Farish, and M. Mitchell, “Time derivatives in air temperature and enthalpy as non-invasive welfare indicators during long distance animal transport,” Biosyst. Eng. 110(3), 253–260 (2011).
5. A. Zhao, X. Tang, Z. Zhang, and J. Liu, “The parameters optimization selection of Savitzky-Golay filter and its application in smoothing pretreatment for FTIR spectra,” in 2014 9th IEEE Conference on Industrial Electronics and Applications, 2014), 516–521.
6. A. R. M. Radzol, K. Y. Lee, W. Mansor, and A. Azman, “Optimization of Savitzky-Golay smoothing filter for salivary surface enhanced Raman spectra of non structural protein 1,” in TENCON 2014 - 2014 IEEE Region 10 Conference, 2014), 1–6.
7. M. a. Awal, S. Mostafa, and M. Ahmad, “Performance Analysis of Savitzky-Golay Smoothing Filter Using ECG Signal,” International Journal of Computer and Information Technology 1(2), 24 (2011).
8. P. Lasch, “Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging,” Chemom. Intell. Lab. Syst. 117, 100–114 (2012).
9. H. Chen, W. Xu, N. Broderick, and J. Han, “An adaptive denoising method for Raman spectroscopy based on lifting wavelet transform,” J. Raman Spectrosc. 39(9), 1529–1539 (2018).
10. F. Ehrentreich and L. Stümchen, “Spike Removal and Denoising of Raman Spectra by Wavelet Transform Methods,” Anal. Chem. 73(17), 4364–4373 (2001).
11. P. M. Ramos and J. Raisínchez, “Noise and background removal in Raman spectra of ancient pigments using wavelet transform,” J. Raman Spectrosc. 36(9), 848–856 (2005).
12. A. E. Villanueva-Luna, J. Castro-Ramos, S. Vazquez-Montiel, A. Flores-Gil, J. A. Delgado-Atencio, and E. E. Orozco-Guillen, “Fluorescence and noise subtraction from Raman spectra by using wavelets,” Opt. Mem. Neural Netw. 19(4), 310–317 (2010).

13. C. A. Lieber and A. Mahadevan-Jansen, “Automated method for subtraction of fluorescence from biological Raman spectra,” Appl. Spectrosc. 57(11), 1363–1367 (2003).

14. S. Chen, X. Lin, C. Yuen, S. Padmanabhan, R. W. Beuerman, and Q. Liu, “Recovery of Raman spectra with low signal-to-noise ratio using Wiener estimation,” Opt. Express 22(10), 12102–12114 (2014).

15. S. Chen, Y. H. Ong, and Q. Liu, “A Method to Create a Universal Calibration Dataset for Raman Reconstruction Based on Wiener Estimation,” IEEE J. Sel. Top. Quantum Electron. 22(3), 164–170 (2016).

16. D. V. d. Sompel, E. Garai, C. Zavaleta, and S. S. Gambhir, “Comparison of Gaussian and Poisson noise models in a hybrid reference spectrum and principal component analysis algorithm for Raman spectroscopy,” Proc. SPIE 8590, 85900F (2013).

17. H.-L. Shen and J. H. Xin, “Spectral characterization of a color scanner based on optimized adaptive estimation,” J. Opt. Soc. Am. A 23(7), 1566–1569 (2006).

18. Y. H. Ong, M. Lim, and Q. Liu, “Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells,” Opt. Express 20(20), 22158–22171 (2012).

19. G. J. Puppels, J. H. F. Olminkhof, G. M. J. Segers-Nolten, C. Otto, F. F. M. de Mul, and J. Greve, “Laser irradiation and Raman spectroscopy of single living cells and chromosomes: Sample degradation occurs with 514.5 nm but not with 660 nm laser light,” Exp. Cell Res. 195(2), 361–367 (1991).

20. Z.-M. Zhang, S. Chen, Y.-Z. Liang, Z.-X. Liu, Q.-M. Zhang, L.-X. Ding, F. Ye, and H. Zhou, “An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy,” J. Raman Spectrosc. 41(6), 659–669 (2009).

21. N. Tarcea, “Raman data analysis,” (2012).

22. A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, “Diagnosing breast cancer by using Raman spectroscopy,” Proc. Natl. Acad. Sci. U. S. A. 102(35), 12371–12376 (2005).

23. K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R. R. Dasari, and M. S. Feld, “Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo,” J. Raman Spectrosc. 33(7), 552–563 (2002).

24. T. Bocklitz, A. Walter, K. Hartmann, P. Rosch, and J. Popp, “How to pre-process Raman spectra for reliable and stable models?” Anal. Chim. Acta 704(1-2), 47–56 (2011).

25. D. Semrau, R. J. Killey, and P. Bayvel, “The Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering,” J. Lightwave Technol. 36(14), 3046–3055 (2018).

26. D. Tian, X. Lv, J. Mo, and C. Chen, “Raman spectroscopy denoising based on smoothing filter combined with EEMD algorithm,” in Tenth International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2017) (SPIE, 2018), 1060704.