Article

The Cascade Hilbert-Zero Decomposition: A Novel Method for Peaks Resolution and Its Application to Raman Spectra

Eugene B. Postnikov ^{1,*} , Elena A. Lebedeva ² , Andrey Yu. Zyubin ³ and Anastasia I. Lavrova ^{4,5} ¹ Department of Theoretical Physics, Kursk State University, 305000 Kursk, Russia² Mathematics and Mechanics Faculty, Saint Petersburg State University, 198504 Saint Petersburg, Russia; ealebedeva2004@gmail.com³ Research and Educational Center “Fundamental and Applied Photonics. Nanophotonics”, Immanuel Kant Baltic Federal University, 236041 Kaliningrad, Russia; azubin@mail.ru⁴ Saint-Petersburg Research Institute of Phthisiopulmonology, 194064 Saint Petersburg, Russia; aurebours@googlemail.com⁵ Medical Faculty, Saint-Petersburg State University, 199034 Saint Petersburg, Russia

* Correspondence: postnikov@kursksu.ru

Abstract: Raman spectra of biological objects are sufficiently complex since they are comprised of wide diffusive spectral peaks over a noisy background. This makes the resolution of individual closely positioned components a complicated task. Here we propose a method for constructing an approximation of such systems by a series, respectively, to shifts of the Gaussian functions with different adjustable dispersions. It is based on the coordination of the Gaussian peaks' location with the zeros of the signal's Hilbert transform. The resolution of overlapping peaks is achieved by applying this procedure in a hierarchical cascade way, subsequently excluding peaks of each level of decomposition. Both the mathematical rationale for the localization of intervals, where the zero crossing of the Hilbert-transformed uni- and multimodal mixtures of Gaussians occurs, and the step-by-step outline of the numerical algorithm are provided and discussed. As a practical case study, we analyze results of the processing of a complicated Raman spectrum obtained from a strain of *Mycobacterium tuberculosis*. However, the proposed method can be applied to signals of different origins formed by overlapped localized pulses too.

Keywords: multiresolution; Hilbert transform; peak resolution; Raman spectra**MSC:** 44A15; 49M27; 78M22; 92-08

Citation: Postnikov, E.B.; Lebedeva, E.A.; Zyubin, A.Y.; Lavrova, A.I. The Cascade Hilbert-Zero Decomposition: A Novel Method for Spectral Peak Resolution and Its Application to Raman Spectra. *Mathematics* **2021**, *9*, 2802. <https://doi.org/10.3390/math9212802>

Academic Editors: Piotr Porwik and Hongyu Liu

Received: 2 September 2021

Accepted: 3 November 2021

Published: 4 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Raman spectroscopy is an effective method for studying the molecular structure of a substance; it can be used in a wide range of scientific applications for organic [1,2] and inorganic [3,4] matter investigations. Biological molecules have a complex nature and structure [5–7]. Due to this fact, the spectrum of Raman light scattering is complex and consists of many components. Moreover, their typical property is a significant widening of peaks, which originates not only from the instrumental response function, but also from complex biochemical structure and interactions [8–10]. Such complex spectra often belong to heterogeneous systems, such as cells, tissues, bacteria. This makes the task of analyzing such spectra similar to the task of regressing continuous (albeit noisy) functions formed by a superposition of wide bell-shaped curves (their width under different biophysical conditions also carries meaningful information) rather than revealing positions of thin well-localized peaks in Raman spectra of pure chemical components for which effective methods of the discrete spectrum reconstructions are developed; see, for example, [11–13]. For the successful identification, classification and quantification of desired features, it is not enough to obtain a high-resolution raw spectrum, but to perform complex pre-processing

procedures using commercial or laboratory-made software and algorithms. Minor or no spectral variation can be observed in raw spectra describing different structures in biological objects. In that case, to obtain meaningful spectral information, we need to process and analyze the data in detail. Univariate and multivariate methods can be applied for it. Univariate methods can include first- and second-order derivatives [14], curve fitting [15], component-wise spectral analysis [16], various bands intensity, peak broadening [17], change in intensities, etc. The multivariate methods are efficient for huge datasets [18] and investigations of subtle changes in spectral data. These methods include principal component analysis (PCA) [19], linear discriminant analysis (LDA) [20], multiple linear regression (MLR) [21], cluster analysis (CA) [22], partial least squares regression (PLS) [23] and can be performed for Raman spectroscopy spectral analysis [24]. On the other hand, applying standard univariate methods for the analysis of subtle spectral differences can be a problem for multicomponent spectra investigations [25]; the multivariate methods need a large amount of data [26], adversely affecting the speed of analysis in the case of practical applications.

As spectral pre-processing methods are successfully used for solving combined applied biophysical and medical problems [27,28], there is a demand for the development of fast and accurate new mathematical methods for Raman spectral analysis, due to the problem of resolution of overlapping peaks. That is especially important in the case of spectral analysis, which draws active attention recently from the point of view of integral transforms, e.g., the wavelet transform [29,30]. This kind of multiscale approach uses possibilities of an extended 2D resolution over multiple scales of a signal's smoothing that improves the accuracy of peak detection from the zero crossings and ridges of the smoothed signal's derivatives. However, in addition to the complicated procedure of 2D image processing, its applicability is limited in the cases of completely merged peaks, i.e., when several peaks form asymmetric monomodal ones as well as for highly noisy data typical for Raman spectra of biological objects.

Additionally, there are some methods that were applied for peak resolution in signals of other origin. The first attempt at using the Hilbert transform for the localization of peaks in signals was proposed in the work [31] within the context of determining the maxima of cross-correlations of signals. However, its author noted the difficulties strongly affecting the results in the case of multiple overlapping peaks and wide peaks of asymmetric shape. On the other hand, such a methodology was successfully applied to detecting R-peaks in electrocardiogram signals, where these peaks are significantly narrow and separated by sufficiently long time intervals [32]. Further, this approach gained popularity and various improvements were proposed, e.g., [33,34].

The idea to expand a function into a series with respect to shifts of the Gaussian function (note that fitting a set of Gaussian functions to a complex spectrum of chemical substance was discussed [35,36] within the context of optimization algorithms, allowing the resolution of close peaks and the analysis of the corresponding residuals) was mathematically justified in [37], where so-called "approximate approximation" by means of various shift-invariant systems was developed. However, our construction is not based on a shift-invariant system. It is adaptive to the form of the function and allows for different dispersion of the Gaussian functions.

Thus, the work is structured as follows: we introduce a mathematical justification for adapting Hilbert's approach to recognize overlapping peaks, then demonstrate it on spectra obtained experimentally on bacteria *Mycobacterium tuberculosis* causing tuberculosis.

2. Zeros of the Hilbert Transform from a Sum of Gaussian Functions

2.1. Basic Properties: A Single Gaussian Function

Let us consider the following Gaussian function:

$$g(x) = e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (1)$$

normalized for the illustrative clarity in such a way that it has the unit amplitude maximum located at the point x_0 and the dispersion σ .

Its Hilbert transform is defined as the following:

$$H[g](x) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{+\infty} \frac{g(x')}{x - x'} dx' \quad (2)$$

and results in the Dawson function as follows:

$$H[g](x) = \frac{2}{\sqrt{\pi}} D\left(\frac{x - x_0}{\sigma\sqrt{2}}\right) \equiv \frac{2}{\sqrt{\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \int_0^{\frac{x-x_0}{\sigma\sqrt{2}}} e^{u^2} du, \quad (3)$$

which has zero value at the point $x = x_0$. See an example in Figure 1 for Equation (1) with $x_0 = 5$ and $\sigma = 1.5$.

Figure 1. The Gaussian function and its Hilbert transform. The vertical dotted lines bound the interval $[x_0 - \sigma, x_0 + \sigma]$.

Thus, the zero crossing of the function $H[g](x)$ can be used to localize the Gaussian's maximum. Taking into account that the Gaussian function is fast decaying, i.e., 60.6% – 13.5% – 1.1 % of the magnitude maximum for $|x - x_0|$ equal to $\sigma - 2\sigma - 3\sigma$, Gaussians separated by a distance exceeding several dispersion values can be considered practically independent, and distances between their peaks can be found by searching for the distances between zeros of the Hilbert transform of such a sum of Gaussian functions as was proposed in Ref. [32].

Figure 2A illustrates such a case for the combined function as follows:

$$f(x) = A_1 e^{-\frac{(x-x_{01})^2}{2\sigma_1^2}} + A_2 e^{-\frac{(x-x_{02})^2}{2\sigma_2^2}} \quad (4)$$

with the amplitudes $A_1 = 1$, $A_2 = 0.5$, the dispersions $\sigma_1 = 1$, $\sigma_2 = 0.5$, and the peak positions $x_{01} = 5$, $x_{02} = 15$.

However, when peak positions are close together, i.e., Gaussian component functions are overlapping, the situation is much more non-trivial and requires a special consideration.

2.2. Zero Crossing for the Hilbert Transform of Overlapping Gaussian Components

As it is seen in Figure 2B,C the closeness of the Gaussian components leads to vanishing separated uni-modal components accompanied with the reduction in the Hilbert

transform's zero crossing to one point, which is clearly separated from the positions of any peak of the individual Gaussian components. More precisely, let us consider the sum of two Gaussians f (4), where for simplicity we put $A_1 = 1$, $x_{01} = 0$, $\sigma_1 = 1$, $A_2 = A$, $x_{02} = x_0$, $\sigma_2 = \sigma$. Without loss of generality, we assume A , x_0 , $\sigma > 0$. We take the Hilbert transform of f and claim that for sufficiently small shift x_0 , the Hilbert transform of f has a unique root.

Figure 2. The functions (solid lines) combined from two Gaussians with different parameters, (A): $A_1 = 1$, $A_2 = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $x_{01} = 5$, $x_{02} = 15$, (B): $A_1 = 1$, $A_2 = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $x_{01} = 5$, $x_{02} = 7$, (C): $A_1 = 1$, $A_2 = 0.25$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $x_{01} = 5$, $x_{02} = 6.5$, and their Hilbert transforms (dashed lines). Vertical dash-dotted lines indicate peak positions for individual components.

Theorem 1. If $x_0 < \min\{q\sqrt{2}, q\sigma\sqrt{2}\}$, where q is the maximum point of the Dawson function, $q = \text{Argmax}D \approx 0.92$, then $H[f](x)$ has a unique root.

Proof. Taking the Hilbert transform of f , we obtain the sum of the Dawson functions as follows:

$$H[f](x) = D\left(\frac{x}{\sqrt{2}}\right) + AD\left(\frac{x-x_0}{\sqrt{2}\sigma}\right).$$

Since $D\left(\frac{x}{\sqrt{2}}\right) > (<)0$ as $x > (<)0$, it follows that $H[f](x) > 0$ as $x \geq x_0$ and $H[f](x) < 0$ as $x \leq 0$. Therefore, the zeros of $H[f](x)$ can only be in the interval $(0, x_0)$. Let us check that $H[f](x)$ is monotone on $(0, x_0)$. To this end, we prove that $H[f]'(x) > 0$ on $(0, x_0)$.

$$H[f]'(x) = \frac{1}{\sqrt{2}}D'\left(\frac{x}{\sqrt{2}}\right) + \frac{A}{\sqrt{2}\sigma}D'\left(\frac{x-x_0}{\sqrt{2}\sigma}\right).$$

It follows from the properties of the Dawson function that $D'(x) > 0$ iff $x \in (-q, q)$. Therefore, $H[f]'(x) > 0$ for $x \in E := (-q\sqrt{2}, q\sqrt{2}) \cap (-q\sigma\sqrt{2} + x_0, q\sigma\sqrt{2} + x_0)$. It remains to note that the inequality $x_0 < \min\{q\sqrt{2}, q\sigma\sqrt{2}\}$ implies the inclusion $(0, x_0) \subset E$. $\square$

For an arbitrary function with some strongly marked clusters of peaks and without any symmetry with respect to the vertical axis, the situation is much more complicated. However, it is possible to justify the following statement (Theorem 2 below): if a function decreases (increases) in some large enough neighborhood of x_0 , then $H[f](x_0) > (<)0$. The size of the neighborhood is defined by a certain inequality involving global properties of the function. This statement allows us to look for a local maximum of f in some neighborhood of the zero of the function $H[f]$; more precisely, at this point, $H[f](x)$ crosses zero from

negative to positive values (see step (II) of the proposed method for the peaks resolution discussed in the next section). Since the Hilbert transform is shift invariant, it is sufficient to examine the point $x_0 = 0$.

Theorem 2. Let f be a positive differentiable function with a compact support $\text{supp } f = [-a, b]$. Suppose there exists a positive number $u > 0$ such that f decreases in $(-u, u)$, and u satisfies the following inequality:

$$2cu + m_a \log \frac{a}{u} - M_b \log \frac{b}{u} > 0, \quad (5)$$

where $c := \min_{x \in (-u, u)} (-f'(x))$, $M_b := \max_{[u, b]} f(x)$, and $m_a := \min_{[-a, -u]} f(x)$, then $H[f](0) > 0$.

Proof. It follows from ([38], Theorem 106) that $H[f](x)$ satisfies the Hölder condition $|H[f](x+t) - H[f](x)| < C|t|^\beta$, $0 < \beta < 1$. As a result, $H[f]$ is well defined and continuous.

By the definition of the Hilbert transform, we obtain the following:

$$H[f](0) = -\frac{1}{\pi} \int_0^\infty \frac{f(t) - f(-t)}{t} dt = -\frac{1}{\pi} \left(\int_0^u + \int_u^\infty \right) \frac{f(t) - f(-t)}{t} dt.$$

By the Lagrange theorem, we obtain the following:

$$-\int_0^u \frac{f(t) - f(-t)}{t} dt = \int_0^u \frac{-2tf'(\tilde{t})}{t} dt \geq 2uc.$$

Since f decreases in $(-u, u)$ it follows that $c \geq 0$. We estimate the remaining integrals as follows:

$$\begin{aligned} \int_u^\infty \frac{f(-t)}{t} dt &\geq m_a \int_u^a \frac{dt}{t} = m_a \log \frac{a}{u}, \\ \int_u^\infty \frac{f(t)}{t} dt &\leq M_b \int_u^b \frac{dt}{t} = M_b \log \frac{b}{u}. \end{aligned}$$

Collecting all the estimates together, we conclude that the inequality (5) completes the proof. $\square$

Remark 1. Analogously, it can be checked that if f increases in $(-u, u)$, and u satisfies the inequality

$$2\tilde{c}u + \tilde{m}_b \log \frac{b}{u} - \tilde{M}_a \log \frac{a}{u} > 0,$$

where $\tilde{c} := \min_{x \in (-u, u)} (f'(x))$, $\tilde{m}_b := \min_{[u, b]} f(x)$, and $\tilde{M}_a := \max_{[-a, -u]} f(x)$, then $H[f](0) < 0$.

3. Outline of the Proposed Method for the Peaks Resolution

3.1. Resolution of Peaks at One Level of Resolution

The method which we propose for the decomposition of a multi-modal function $f(x)$ consists of the following steps.

(I) Compute the Hilbert transform of the function $f(x)$, i.e., $H[f](x)$.

This can be easily carried out for a discrete data sample either straightforwardly applying the fast Fourier transform or by another appropriate numerical method [39] depending on the signal's complexity.

(II) Find all zeros of the function $H[f](x)$ found at the previous step such that $H[f](x)$ crosses zero from negative to positive values.

Numerically, the simplest way to do this is to check the zero crossing when one of the subsequent data points is negative and another is positive. The latter condition is based on Theorem 2 and Remark 1: if the Hilbert transform intersects the axis, i.e., $H[f](x) = 0$, then in a sufficiently large neighborhood of such a point on the left-hand side, the original function cannot increase, and on the right-hand side, it cannot decrease.

(III) Around each appropriate j -th zero crossing point, find an interval within which the Hilbert transform with 98 % correlation corresponds to a linear fit on this interval, and fit the processed function on this interval by the following Gaussian:

$$G_j(x) = B_j e^{-\frac{(x-x_{0j})^2}{2\sigma_j^2}}$$

by a non-linear regression that gives its amplitude B_j , variance σ_j^2 , and the corrected point of maximum x_{0j} .

The necessity of such an approach is again based on the important fact noted above: the zero crossing point does not coincide with the true peak position if the signal is asymmetric, e.g., as a consequence of merging several close positioned wide elementary components. Thus, the optimization procedure shifts x_{0j} to the actual position. The choice of the interval for this optimization procedure is based on the properties of the Hilbert transform of the Gaussian (1) given by the Dawson function (3) as follows:

$$D(z) = \sum_{k=0}^{\infty} \frac{(-1)^k 2^k}{(2k+1)!!} z^{2k+1} = z - \frac{2}{3} z^3 + \dots,$$

$$H[g(x)] \approx \frac{2}{\sqrt{\pi}} \frac{x - x_0}{\sigma \sqrt{2}} =: y(x),$$

$$0 < y(x) - H[g(x)] < \frac{2}{3\sqrt{2\pi}} \frac{(x - x_0)^3}{\sigma^3}.$$

One can see in Figure 1 that this interval covers the peak's vicinity practically up to the boundaries of the peak's dispersion, i.e., the optimization procedure for searching the parameters of the nearest Gaussian function closest to the found point of the Hilbert transform's zero crossing acts within the principal part $[x_0 - \sigma, x_0 + \sigma]$ of the desired Gaussian function.

(IV) Sum up all the Gaussian components found as $f_1(x) = \sum G_j(x)$ and compare the obtained approximation with the processed function.

Figure 3 illustrates the results of applying the steps described above to three test signals shown in Figure 2. The case of Figure 3A consists of two sufficiently separated Gaussians that result in two zero crossings of its Hilbert transform; see Figure 2A. Both these components are detected and the processed signal is recovered perfectly. The parameters (B_j, x_{0j}, σ_j) found for the first and the second component are (1.02, 5.02, 1.02) and (0.49, 7.00, 0.50), respectively.

For two more complicated cases, Figure 3B,C, only one zero crossing, and, moreover, at the point that does not coincide with any of the maxima of individual components, occurs; see Figure 2B,C. As a result, only one Gaussian component can be recovered for these cases. At the same time, it is clearly seen that the procedure of the non-linear fitting-based optimization of the recovered peak's position accurately shifts the revealed single Gaussian component to its true position. When both Gaussian components overlap but two peaks are still visible, Figure 2B, the reconstruction of the left one quite accurately follows the processed signal, where an influence of the right Gaussian is negligible. When both peaks are so close that the resulting signal is a monomodal asymmetric peak, Figure 2C, the left component differs from the shape of the overall curve.

Thus, the next stage required is to supply this primary analysis with additional steps aimed at revealing the rest of the components.

Figure 3. Blue solid lines correspond to the functions combined from two Gaussians with different parameters. The functions plotted in subpanels (A–C) are the same as in Figure 2 denoted by the respective letters. Dashed black lines depict results of a primary run of the features extraction algorithm described in Section 3.1 (note that original and recovered curves in the subpanel (A) coincide practically perfectly). Vertical dash-dotted lines indicate exact peak positions for individual components.

3.2. The Cascade Revealing Hidden Peaks

As seen in Figure 3, the part where a component (or several components) associated with zero crossings of the original signal's Hilbert transform dominates is reproduced with high accuracy. Therefore, the task of reconstructing the rest of the components can be realized as a cascade of the following steps.

(V) Find the difference between the original signals and the signal reconstructed during the steps (I)–(IV), $\Delta y(x) = f(x) - f_1(x)$, and consider this the new function to which the steps (I)–(IV) should be applied again.

Figure 4 illustrates this step. One can see that the difference between the initial function and the actually found Gaussian has now the form of a bell-shaped curve. Respectively, the Hilbert transform taken from this residual again exhibits zero crossing. For the case of more separated Gaussians forming the initial signal, Figure 4A, this residual looks like an accurate Gaussian, has a unique zero crossing of its Hilbert transform, and the procedure of the component's fitting is simple and straightforward as above.

Figure 4B illustrates a more complicated situation of more closely positioned components in the initial signal. Here, one can see a more complicated shape of the Hilbert transform, which has two zero crossings. However, as it is argued in the description of step (II), the first one should not be considered since it is the zero crossing by the decaying part of the Hilbert transform. Note also that the plot of the residual around this point demonstrates $\Delta y < 0$, i.e., one cannot consider this “hump” as a hidden peak because we are interested in the positive peaks only for our particular class of problems.

Figure 4. The residuals between the processed signals and their primary found components shown in Figure 3 (solid lines), and the reconstruction of the Gaussian components within these residuals (dashed lines). The subpanels (A,B) of this figure correspond to signals presented in subpanels (B,C) in Figure 3. Vertical dash-dotted lines indicate exact peak positions for individual components in the initial signal, and the dotted line is the Hilbert transform of the residual.

The second zero crossing goes in the required direction, and the non-linear fitting procedure restores the desired second Gaussian component of the signal.

Finally, the full combination of revealed peaks can be constructed. However, since the second peaks' magnitude can significantly differ from zero at the point of location of primary peaks, we need the step of internalization of amplitudes for the peaks found during previous steps, similar to the correction step of the lifting scheme for constructing wavelets [40]:

(VI) $B_j^{(k)} \rightarrow B_j^{(k)} - \sum_{j'} B_{j'}^{(k+1)} e^{-\frac{(x_{0j}^{(k)} - x_{0j'}^{(k+1)})^2}{2(\sigma_{j'}^{(k+1)})^2}}$, where superscripts (k) and $(k+1)$ indicate the iterations of the cascade of calculations (I)–(V), which can be repeated up to the moment, when the current residual function indicates the absence of further peaks of interest.

Figure 5 illustrates the full reconstructed signal in comparison with the original one. The parameters of individual Gaussian components are found as $(B_1^{(1)}, x_{01}^{(1)}, \sigma_1^{(1)}) = (1.00, 5.02, 1.09)$ and $(B_1^{(2)}, x_{01}^{(2)}, \sigma_1^{(2)}) = (0.18, 6.51, 0.45)$, where superscripts indicate the required number of iterations in the cascade algorithm for revealing the particular component. The amplitude of the first component is rescaled according to step (VI) after finding the second component. As it is seen, the proposed method allows for reasonable reproduction of signals formed by overlapping peaks even in the case when the resulting function looks monomodal, and extracting the respective hidden constituents.

Figure 5. Original signals corresponding to Figure 3B,C (here the subpanels (A,B), respectively) shown as blue solid lines, their fitting by the series of Gaussian functions (black dashed lines) and the respective individual Gaussian components (magenta dashed lines). The green dotted curves close to zero demonstrate the residual between the original curve and its regression. Vertical dash-dotted lines indicate exact peak positions for individual components in the initial signal.

4. A Practical Case Study: Decomposition of Raman Spectra of *Mycobacterium tuberculosis*

4.1. Materials, Methods, and Data

As an example, we consider the experimental profile for an XDR strain of *M. tuberculosis* from the work [6], where details of data origin and experimental procedure are described.

4.2. Resolution of Peaks

Figure 6 shows the analyzed real Raman spectrum as a function of the wavenumber κ and indicates Gaussian components as follows:

$$f(\kappa) = \sum_j B_j e^{-\frac{(\kappa - \kappa_{0j})^2}{2\sigma_j^2}} \quad (6)$$

found by the proposed method within three iterations. The signal was normalized to unity instead of 100 used in the work [41] for the sake of convenience. Here, we used the numerical factor 6 for comparison of the interval of local fitting and the obtained dispersion of the Gaussian functions (see the description of step (III) above) for rejecting suspiciously too wide peaks. In addition, peaks with amplitudes less than 0.05 were also rejected, comparable with the noisy background and deviations originated from imperfect reproduction of the constituent peaks' tails.

One can see that the envelope (dashed black curve) representing Equation (6) accurately reproduces the shape of the whole sequence of Raman peaks with amplitudes larger than 0.05, including ones which have a complicated asymmetric shape with “wide shoulders”. The picture of individual components demonstrating such peaks can be considered to comprise several close-positioned Gaussians. The full set of parameters for the individual components revealed is listed in Table 1.

These components are found during three iterations illustrated in Figure 7. Note that their amplitudes shown in this figure correspond to the non-linear fitting at the current

iteration (step (III) in the description above), while in Figure 6 and in Table 1 they are rescaled to keep the magnitude of the whole signal (step (VI)).

Figure 6. Normalized Raman spectrum of *M. tuberculosis* pulmonary strain (green curve) and its decomposition into the series of Gaussian components via the cascade Hilbert-Zero method proposed in this work. Magenta, blue, and red dotted bell-shaped curves show components identified during the first, the second, and the third iterations, respectively. The black dashed curve is the sum of all components resembling main course of principal features revealed.

Table 1. The components of the expansion of the Raman spectrum into a series of Gaussian peaks, where κ_{0j} is the maximum's wavenumber (cm^{-1}), B_j and σ_j are the amplitude and the dispersion (cm^{-1}) of the j -th component. In addition, the numbers (k) of the cascade algorithm's iteration and biochemical comments are provided in the corresponding columns.

κ_{0j}	B_j	σ_j	(k)	Comment
642	0.07	1.7	3	Skeletal (C – C) (peptidoglycan) [42]
673	0.19	11.0	1	Tyrosine [43]
781	0.08	5.0	3	Cytosine, Uracil [42]
819	0.36	7.7	2	Ring-breathing mode of tyrosine with the CaDPA carboxylate stretching mode [44]
850	0.66	12.3	1	ν CC) ring breathing (peptidoglycan) [42]
922	0.29	12.1	1	ν (CHR ₂), C~C skel. str where R – CH ₃ (phospholipids) [45]
974	0.13	7.7	3	L-Histidine [46]
1002	0.18	1.9	3	ν CC) Aromatic ring breathing (phenylalanine) [47]
1029	0.13	5.9	3	L-Valine [48]
1042	0.49	8.0	2	L-Glutamate [48]
1061	0.60	8.2	1	C~C, C~N (possible MA profile) [49]
1088	0.59	20.8	2	D-(C)-Fucose [48]
1119	0.27	10.7	3	D-glucose [48]
1248	0.19	20.0	3	Amide III [42]
1295	0.34	2.6	1	(possible MA profile) [50]
1302	0.14	3.4	3	ω triolein and trilinolenin [48]
1436	0.30	6.5	3	δ (CH ₂) (saturated lipids) (MA profile) [50]
1460	0.99	15.0	1	CH ₂ /CH ₃ deformation mode of proteins and lipids [51]
1665	0.26	16.2	1	Amide I [51]

Figure 7. Three iterations of the cascade Hilbert-Zero decomposition applied to the Raman spectra shown in Figure 6. Blue solid curves show the initial signal processed during the first iteration (A) and two sequential residuals (B,C), and the red dash-dotted curves are the respecting Hilbert transform curves. Black dotted bell-shaped curves are the Gaussian components for each iteration.

5. Discussion

We have developed a new approach that allows more accurate spectral analysis in comparison with those available earlier. We confirmed this by applying the approach to already published Raman spectra aimed at the investigation of the cell wall of *Mycobacterium tuberculosis* [6]. The attempt described in the cited paper to apply the commercial KnowItAll Vibrational Spectroscopy Edition (BioRad, Hercules, CA, USA) software, which is de facto standard in biophysical Raman-based studies, to the experimentally obtained curves resulted in revealing 15 identified peaks. The results of our new approach lists them also in the set given in Table 1 with the same accuracy of resolution in wavelengths that validates the new method. However, also Table 1 provides information on four additional peaks, which correspond to actually existing biochemical structures; see below.

We have revealed new spectral vibrations in the composition of the complex spectrum of mycobacterium: 1029 cm^{-1} for L-valine; 1042 cm^{-1} for L-glucamate; 1088 cm^{-1} for D-(C)-fucose; 1119 cm^{-1} for D-glucose; 1302 cm^{-1} for triolein/trilinolenin. This information can be used for detailed analysis of the cell wall and life cycle of mycobacteria. Thus, the amino acid L-valine and its spectral characteristics can be used for the analysis of enzymes associated with it [52], the analysis of the mechanisms of their synthesis [53]. D-glucose is also an element of the cell cycle and it can be used as a biomarker of its description [54]. Glutamate is one of amino acids that act as cellular nitrogen donors for the synthesis of biomolecules inside the cell. In mycobacteria, the assimilation of synthetic nitrogen and its conversion to glutamate is carried out by glutamate synthetase. The cell wall of pathogenic mycobacteria is known to possess a poly-L-glutamine layer. Poly-L-glutamine layer synthesis has been directly linked to the glutamine synthetase enzyme [55]. Due to this fact, glutamate spectral characteristics can be used as a biomarker in cell wall changes. D-fucose can be found in both Gram-negative and Gram-positive bacteria, in which it is a constituent of the cell wall and capsule structures [56]; this fact can be used for strain discrimination. The triolein and lipolytic protein have roles in the pathogenesis and survival of *M. tuberculosis* [57].

It should be noted that the fast and reliable identification and discrimination of *Mycobacterium tuberculosis* (MbT) bacteria species and its drug resistance is a serious global problem today. The successful spectral determination of a specific MbT strain and its intra-strain differences is vital for the appointment of anti-tuberculosis therapy [58].

Note that the developed approach may lead to inadequately wide Gaussians as outputs in the case when a processed signal contains strong noise in the vicinity of a narrow peak. This problem can be avoided by the control of output peaks' width by their comparison with the interval of the fitting. As it is seen in Figure 1, the fitting interval $[x_0 - \sigma, x_0 + \sigma]$, where the Hilbert transform of the function is almost linear, is comparable in width with a true Gaussian's dispersion. This assures the robustness of fitting as defined by the classical criteria of linear regression. On the contrary, if a revealed Gaussian has a dispersion much wider than the fitting interval, then it should be rejected. As well, two peaks cannot be resolved if the distance between their maxima is either less than the minimum of two σ or comparable with the recorded instrumental resolution.

Nevertheless, the developed approach can be convenient in revealing hidden peaks in the spectra of biological objects, which is extremely useful in revealing subtle differences between cells or cell populations. Moreover, an application or program created on the basis of the approach can be made available to a wide range of users working in experimental spectroscopy.

6. Conclusions

While the vast majority of peak localization algorithms solve the detection problem for relatively well-expressed maxima in signals, we have developed a new approach using the Hilbert transform, which allows us to determine with high accuracy the hidden peaks too. This problem is especially important for spectra obtained from biological samples. We verified this method using Raman spectra obtained on bacterial cells, which made it possible to identify additional chemical substances in the composition of the cell wall. It could be helpful for further interpretation of the spectra in detail to find differences between the strains more precisely.

Finally, the knowledge of not only the separated peaks' localization, but also their amplitudes and characteristic width can be useful for further data interpretation by modern methods, which use machine learning, requiring a larger set of data characteristics.

Author Contributions: Conceptualization, A.I.L. and E.B.P.; methodology, E.B.P. and E.A.L.; software, E.B.P.; validation, A.Y.Z. and A.I.L.; formal analysis, E.B.P., E.A.L., A.Y.Z. and A.I.L.; investigation, E.B.P., E.A.L., A.Y.Z. and A.I.L.; resources, A.I.L.; data curation, A.Y.Z. and A.I.L.; writing—original draft preparation, E.B.P., E.A.L., A.Y.Z. and A.I.L.; writing—review and editing, E.B.P., E.A.L., A.Y.Z. and A.I.L.; visualization, E.B.P.; supervision, E.B.P. and A.I.L.; project administration, E.B.P. and A.I.L.; funding acquisition, E.A.L. and A.Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: E.A.L. is supported by the Russian Science Foundation, grant number 18-11-00055; A.Y.Z. and A.I.L. are supported by the Ministry of Science and Higher Education, project number 075-02-2021-1748.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Rygula, A.; Majzner, K.; Marzec, K.M.; Kaczor, A.; Pilarczyk, M.; Baranska, M. Raman spectroscopy of proteins: A review. *J. Raman Spectrosc.* **2013**, *44*, 1061–1076. [\[CrossRef\]](#)
2. Howell, N.K.; Arteaga, G.; Nakai, S.; Li-Chan, E.C.Y. Raman spectral analysis in the C-H stretching region of proteins and amino acids for investigation of hydrophobic interactions. *J. Agric. Food Chem.* **1999**, *47*, 924–933. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Lenz, R.; Enders, K.; Stedmon, C.A.; Mackenzie, D.M.A.; Nielsen, T.G. A critical assessment of visual identification of marine microplastic using Raman spectroscopy for analysis improvement. *Mar. Pollut. Bull.* **2015**, *100*, 82–91. [\[CrossRef\]](#)
4. Zobkov, M.B.; Esiukova, E.E.; Zyubin, A.Y.; Samusev, I.G. Microplastic content variation in water column: The observations employing a novel sampling tool in stratified Baltic Sea. *Mar. Pollut. Bull.* **2019**, *138*, 193–205. [\[CrossRef\]](#)
5. Ma, P.; Xu, L.; Wang, L.; Chen, N.; Zhang, X.; Chen, H.; Li, J. Molecular detection of cordycepin-induced HeLa cell apoptosis with surface-enhanced Raman spectroscopy. *Appl. Sci.* **2019**, *9*, 3990. [\[CrossRef\]](#)

6. Zyubin, A.; Lavrova, A.; Manicheva, O.; Dogonadze, M.; Belik, V.; Samusev, I. Raman spectroscopy reveals M. tuberculosis strains with different antibiotic susceptibility. *Laser Phys. Lett.* **2019**, *16*, 085602. [\[CrossRef\]](#)
7. Onogi, C.; Motoyama, M.; Hamaguchi, H. High concentration trans form unsaturated lipids detected in a HeLa cell by Raman microspectroscopy. *J. Raman Spectrosc.* **2008**, *39*, 555–556. [\[CrossRef\]](#)
8. Kneipp, K.; Kneipp, H.; Itzkan, I.; Dasari, R.R.; Feld, M.S. Surface-enhanced Raman scattering and biophysics. *J. Phys. Condens. Matter* **2002**, *14*, R597. [\[CrossRef\]](#)
9. Atkins, C.G.; Buckley, K.; Blades, M.W.; Turner, R.F.B. Raman spectroscopy of blood and blood components. *Appl. Spectrosc.* **2017**, *71*, 767–793. [\[CrossRef\]](#)
10. Zyubin, A.; Rafalskiy, V.; Tcibulnikova, A.; Moiseeva, E.; Matveeva, K.; Tsapkova, A.; Lyatun, I.; Medvedskaya, P.; Samusev, I.; Demin, M. Surface-enhanced Raman spectroscopy for antiplatelet therapy effectiveness assessment. *Laser Phys. Lett.* **2020**, *17*, 045601. [\[CrossRef\]](#)
11. Kwiatkowski, A.; Gnyba, M.; Smulko, J.; Wierzbza, P. Algorithms of chemicals detection using Raman spectra. *Metrol. Meas. Syst.* **2010**, *17*, 549–560. [\[CrossRef\]](#)
12. Sizikov, V.; Sidorov, D. Discrete spectrum reconstruction using integral approximation algorithm. *Appl. Spectrosc.* **2017**, *71*, 1640–1651. [\[CrossRef\]](#)
13. Aleksandrova, A.A.; Sizikov, V.S. Recovery of discrete spectra radiated by substance in deep vacuum using integral approximation algorithm. *Sci. Tech. J. Inf. Technol.* **2020**, *20*, 353–363. [\[CrossRef\]](#)
14. Yacoby, Y.; Kruhl, W.W.; Just, S. Study of soft modes by temperature-derivative first-and second-order Raman spectroscopy. *Phys. Rev. B* **1976**, *13*, 4132–4140. [\[CrossRef\]](#)
15. Mysen, B.O.; Finger, L.W.; Virgo, D.; Seifert, F.A. Curve-fitting of Raman spectra of silicate glasses. *Am. Mineral.* **1982**, *67*, 686–695.
16. Zyubin, A.Y.; Rafalskiy, V.V.; Tcibulnikova, A.V.; Matveeva, K.I.; Moiseeva, E.M.; Kolosova, V.V.; Misuk, V.O.; Tsapkova, A.A.; Samusev, I.G. Single human platelet study using surface-enhanced Raman spectroscopy as a perspective tool for antiplatelet therapy effectiveness prediction. *Proc. SPIE* **2019**, *11190*, 111900K. [\[CrossRef\]](#)
17. Winkler, W.; Musso, M.; Kirchner, E.C. Fourier transform Raman spectroscopic data on the fossil resin siegburgite. *J. Raman Spectrosc.* **2003**, *34*, 157–162. [\[CrossRef\]](#)
18. Jobson, J.D. *Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods*; Springer Science & Business Media: New York, NY, USA, 2012.
19. Shinzawa, H.; Awa, K.; Kanematsu, W.; Ozaki, Y. Multivariate data analysis for Raman spectroscopic imaging. *J. Raman Spectrosc.* **2009**, *40*, 1720–1725. [\[CrossRef\]](#)
20. Krafft, C.; Steiner, G.; Beleites, C.; Salzer, R. Disease recognition by infrared and Raman spectroscopy. *J. Biophotonics* **2009**, *2*, 13–28. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Rajalahti, T.; Kvalheim, O.M. Multivariate data analysis in pharmaceuticals: A tutorial review. *Int. J. Pharm.* **2011**, *417*, 280–290. [\[CrossRef\]](#)
22. Santos, R.S.; Sidaoui, H.A.; Silveira, L.; Pasqualucci, C.A.G.; Pacheco, M.T.T. Classification system of Raman spectra using cluster analysis to diagnose coronary artery lesions. *Instrum. Sci. Technol.* **2009**, *37*, 327–344. [\[CrossRef\]](#)
23. Zhang, L.; Li, Q.; Tao, W.; Yu, B.; Du, Y. Quantitative analysis of thymine with surface-enhanced Raman spectroscopy and partial least squares (PLS) regression. *Anal. Bioanal. Chem.* **2010**, *398*, 1827–1832. [\[CrossRef\]](#)
24. Barmapalexis, P.; Karagianni, A.; Nikolakakis, I.; Kachrimanis, K. Artificial neural networks (ANNs) and partial least squares (PLS) regression in the quantitative analysis of cocrystal formulations by Raman and ATR-FTIR spectroscopy. *J. Pharm. Biomed. Anal.* **2018**, *158*, 214–224. [\[CrossRef\]](#)
25. Filzmoser, P.; Hron, K.; Reimann, C. Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Sci. Total. Environ.* **2009**, *407*, 6100–6108. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Camacho, J.; Magán-Carrión, R.; García-Teodoro, P.; Treinen, J.J. Networkmetrics: Multivariate big data analysis in the context of the internet. *J. Chemom.* **2016**, *30*, 488–505. [\[CrossRef\]](#)
27. Geetha, R.; Sivasubramanian, S.; Kaliappan, M.; Vimal, S.; Annamalai, S. Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier. *J. Med. Syst.* **2019**, *43*, 286. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Kumar, V.; Sachdeva, J.; Gupta, I.; Khandelwal, N.; Ahuja, C.K. Classification of brain tumors using PCA-ANN. In Proceedings of the 2011 World Wongress on Information and Communication Technologies, Mumbai, India, 11–14 December 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1079–1083. [\[CrossRef\]](#)
29. Zhang, Z.M.; Tong, X.; Peng, Y.; Ma, P.; Zhang, M.J.; Lu, H.M.; Chen, X.Q.; Liang, Y.Z. Multiscale peak detection in wavelet space. *Analyst* **2015**, *140*, 7955–7964. [\[CrossRef\]](#)
30. Liu, M.; Dong, Z.; Xin, G.; Sun, Y.; Qu, R. An improved method based on a new wavelet transform for overlapped peak detection on spectrum obtained by portable Raman system. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 1–8. [\[CrossRef\]](#)
31. Cabot, R. A note on the application of the Hilbert transform to time delay estimation. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 607–609. [\[CrossRef\]](#)
32. Manikandan, M.S.; Soman, K.P. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed. Signal Process. Control* **2012**, *7*, 118–128. [\[CrossRef\]](#)
33. Sun, S.; Jiang, Z.; Wang, H.; Fang, Y. Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform. *Comput. Methods Programs Biomed.* **2014**, *114*, 219–230. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Rakshit, M.; Das, S. An efficient wavelet-based automated R-peaks detection method using Hilbert transform. *Biocybern. Biomed. Eng.* **2017**, *37*, 566–577. [[CrossRef](#)]
35. Griffiths, T.R.; Nerukh, D.A.; Eremenko, S.A. The application of theoretical models of complex shape to the fitting of experimental spectra having closely overlapping bands. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3199–3208. [[CrossRef](#)]
36. Nerukh, D.; Griffiths, T.R. Complex vibrational correlation functions extracted from the resolved ν 2 band of liquid acetonitrile. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1799–1805. [[CrossRef](#)]
37. Maz'ya, V.; Schmidt, G. *Approximate Approximations*; American Mathematical Society: Providence, RI, USA, 2007.
38. Titchmarsh, E.C. *Introduction to the Theory of Fourier Integrals*; Clarendon Press: Oxford, UK, 1948.
39. King, F.W. *Hilbert Transforms*; Cambridge University Press: Cambridge, UK, 2009; Volume I.
40. Daubechies, I.; Sweldens, W. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.* **1998**, *4*, 247–269. [[CrossRef](#)]
41. Zyubin, A.; Lavrova, A.; Manicheva, O.; Dogonadze, M.; Belik, V.; Samusev, I. Raman spectroscopy for glutathione measurements in Mycobacterium tuberculosis strains with different antibiotic resistance. *J. Raman Spectrosc.* **2021**, *52*, 1661–1666. [[CrossRef](#)]
42. Kahraman, M.; Keseroğlu, K.; Çulha, M. On sample preparation for surface-enhanced Raman scattering (SERS) of bacteria and the source of spectral features of the spectra. *Appl. Spectrosc.* **2011**, *65*, 500–506. [[CrossRef](#)]
43. Xie, C.; Goodman, C.; Dinno, M.A.; Li, Y.Q. Real-time Raman spectroscopy of optically trapped living cells and organelles. *Opt. Express* **2004**, *12*, 6208–6214. [[CrossRef](#)]
44. De Gelder, J. Raman Spectroscopy as a Tool for Studying Bacterial Cell Compounds. Ph.D. Thesis, Ghent University, Ghent, Belgium, 2008.
45. Rivera-Betancourt, O.E.; Karls, R.; Grosse-Siestrup, B.; Helms, S.; Quinn, F.; Dluhy, R.A. Identification of mycobacteria based on spectroscopic analyses of mycolic acid profiles. *Analyst* **2013**, *138*, 6774–6785. [[CrossRef](#)]
46. De Gelder, J.; De Gussem, K.; Vandenabeele, P.; De Vos, P.; Moens, L. Methods for extracting biochemical information from bacterial Raman spectra: An explorative study on *Cupriavidus metallidurans*. *Anal. Chim. Acta* **2007**, *585*, 234–240. [[CrossRef](#)]
47. Harz, M.; Rösch, P.; Peschke, K.D.; Ronneberger, O.; Burkhardt, H.; Popp, J. Micro-Raman spectroscopic identification of bacterial cells of the genus *Staphylococcus* and dependence on their cultivation conditions. *Analyst* **2005**, *130*, 1543–1550. [[CrossRef](#)] [[PubMed](#)]
48. De Gelder, J.; De Gussem, K.; Vandenabeele, P.; Moens, L. Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* **2007**, *38*, 1133–1147. [[CrossRef](#)]
49. Boyaci, I.H.; Temiz, H.T.; Geniş, H.E.; Soykut, E.A.; Yazgan, N.N.; Güven, B.; Uysal, R.S.; Bozkurt, A.G.; İlaslan, K.; Torun, O.; et al. Dispersive and FT-Raman spectroscopic methods in food analysis. *RSC Adv.* **2015**, *5*, 56606–56624. [[CrossRef](#)]
50. Liu, T.Y.; Chen, Y.; Wang, H.H.; Huang, Y.L.; Chao, Y.C.; Tsai, K.T.; Cheng, W.C.; Chuang, C.Y.; Tsai, Y.H.; Huang, C.Y.; et al. Differentiation of bacteria cell wall using Raman scattering enhanced by nanoparticle array. *J. Nanosci. Nanotechnol.* **2012**, *12*, 5004–5008. [[CrossRef](#)]
51. Stöckel, S.; Meisel, S.; Elschner, M.; Rösch, P.; Popp, J. Raman spectroscopic detection of anthrax endospores in powder samples. *Angew. Chem. Int. Ed.* **2012**, *51*, 5339–5342. [[CrossRef](#)] [[PubMed](#)]
52. Allaudeen, H.S.; Ramakrishnan, T. Biosynthesis of isoleucine and valine in *Mycobacterium tuberculosis* H37 Rv. *Arch. Biochem. Biophys.* **1968**, *125*, 199–209. [[CrossRef](#)]
53. Amorim Franco, T.M.; Hegde, S.; Blanchard, J.S. Chemical mechanism of the branched-chain Aminotransferase IlvE from *mycobacterium tuberculosis*. *Biochemistry* **2016**, *55*, 6295–6303. [[CrossRef](#)]
54. O'Barr, T.P.; Rothlauf, M.V. Metabolism of D-glucose by *Mycobacterium tuberculosis*. *Am. Rev. Respir. Dis.* **1970**, *101*, 964–966.
55. Tripathi, D.; Chandra, H.; Bhatnagar, R. Poly-L-glutamate/glutamine synthesis in the cell wall of *Mycobacterium bovis* is regulated in response to nitrogen availability. *BMC Microbiol.* **2013**, *13*, 226. [[CrossRef](#)]
56. Mäki, M.; Renkonen, R. Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology* **2004**, *14*, 1R–15R. [[CrossRef](#)]
57. Monu; Meena, L.S. Roles of triolein and lipolytic protein in the pathogenesis and survival of *Mycobacterium tuberculosis*: A novel therapeutic approach. *Appl. Biochem. Biotechnol.* **2016**, *178*, 1377–1389. [[CrossRef](#)] [[PubMed](#)]
58. Denholm, J.T.; McBryde, E.S. The use of anti-tuberculosis therapy for latent TB infection. *Infect. Drug Resist.* **2010**, *3*, 63–72. [[CrossRef](#)] [[PubMed](#)]