

Peak Detection in Mass Spectra: An Application of Least-Squares Piecewise Monotonic Approximation

IOANNIS N. PERDIKAS¹, VASSILIOS KOUTOULIDIS², IOANNIS C. DEMETRIOU¹

¹Department of Economics

1 Sofokleous and Aristidou Street, 10559 Athens
National and Kapodistrian University of Athens
GREECE

²First Department of Radiology, Medical School-Areteion Hospital

76 Vassilisis Sofias, 11528, Athens
National and Kapodistrian University of Athens
GREECE

Abstract: Peak detection in mass spectrometry is a challenging signal-processing problem due to the presence of noise, overlapping peaks, and the intrinsic complexity of spectral data. This paper investigates the application of the least-squares piecewise monotonic approximation method to automatic peak detection in mass spectra. The method models the data as a sequence of alternating monotonic segments and identifies peaks through the optimal determination of turning points. The underlying optimization problem is combinatorial in nature, requiring to test a large number of possible breakpoint arrangements. However, because of a decomposition property of the approximation problem, the computation can be carried out efficiently through dynamic programming in quadratic time complexity. Within this approach, peak detection is formulated and solved as a globally optimized approximation problem under minimal assumptions. The method is applied to a complex mass spectrum from a publicly available MassBank record. Numerical results show that the method reliably identifies both dominant and subtle spectral features, while simultaneously capturing the underlying structure of the data stage by stage. The accuracy of the approximation improves consistently, remaining free of oscillatory artifacts, as the number of monotonic segments increases. The findings demonstrate that piecewise monotonic approximation gives a robust, accurate, and computationally efficient methodology for automatic peak detection in mass spectrometry.

Key-Words: data approximation, first differences, least squares, mass spectrometry, peak detection, piecewise monotonic

Received: May 7, 2026. Revised: June 9, 2026. Accepted: July 11, 2026. Published: October 6, 2026.

1 Introduction

Mass spectrometry (MS) is a powerful analytical technique widely used in chemistry, biology, medicine, environmental science, and related disciplines due to its ability to provide rapid and highly sensitive molecular information. By measuring the mass-to-charge ratio of ions, MS enables the identification and quantification of chemical compounds within complex samples, making it a valuable tool in both scientific research and industrial applications.

A fundamental task in mass spectrometry is peak detection (or peak finding), namely the identification of significant spectral features associated with ions and analytes of interest. Accurate peak detection is essential for the reliable interpretation and quantification of mass spectra. However, this task is intrinsically challenging due to the presence of noise, overlapping peaks, baseline fluctuations, instrumental limitations, and the overall complexity of spectral data. In particular, small variations in signal intensity may generate false positives, while

weak but meaningful peaks may remain undetected. Consequently, robust and computationally efficient peak detection remains an active area of research.

A variety of methodologies have been developed for automated peak detection in mass spectrometry, including smoothing procedures, local thresholding techniques, wavelet-based methods, and statistical signal-processing approaches. Representative contributions include the wavelet-based peak detection method of Du et al. [8], the comparative analysis of signal-to-noise and peak-processing techniques by Yang et al. [13], and the baseline correction and noise-reduction procedures proposed by Coombes et al. [2].

Although these methods have significantly improved the automation and reproducibility of mass spectrometry analysis, reliable peak detection in noisy and structurally complex spectra remains difficult. In particular, many existing approaches depend strongly on parameter selection, such as signal-to-noise thresholds, peak widths, or smoothing parameters. Poor choices may lead either to missed peaks or to the detection of spurious spectral features.

An efficient approach for peak detection would significantly enhance the analytical capabilities of mass spectrometry, particularly in the interpretation of complex and noisy spectra. Moreover, effective processing of large spectral datasets and reliable peak identification are essential for improved spectral resolution and consistent extraction of meaningful molecular information across mass spectra applications.

In this paper, it is investigated, for the first time in the context of mass spectrometry, the application of the piecewise monotonic approximation (PMA) method to automatic peak detection, which, as said, is an important and challenging practical problem. This method models spectral data as a sequence of alternating monotonic segments and identifies peaks through the optimal determination of turning points.

Unlike most existing peak-detection methods, which rely on local criteria, signal-processing heuristics, or transform-based representations, PMA determines the turning points of a globally optimized least-squares approximation under piecewise monotonicity constraints, with its maxima identifying the peaks and its minima the troughs. In particular, Fourier- and wavelet-based approaches (see, for example, Gasquet and Witomsky [10]) depend on the choice of basis functions, resolution scales, or thresholding parameters. An important advantage of PMA is that peak detection is incorporated directly into the globally optimized approximation, avoiding the need to balance accurate local representation of peaks against approximation quality elsewhere in the spectrum.

The PMA problem is combinatorial in nature, because the turning-point locations are unknown discrete variables that must be determined simultaneously with the optimal approximation. A direct exhaustive search over all turning-points rapidly becomes computationally infeasible as the number of turning-points increases. However, due to a decomposition property established by Demetriou and Powell [7] and Demetriou [5, 6], the problem can be solved quite efficiently through dynamic programming techniques. Specifically, depending on the implementation, the optimal approximation can be computed in at most $O(kn^2)$ operations, or alternatively in $O(n^2 + kn \log_2 n)$ operations, where $k - 1$ denotes the number of turning points of the fit.

In this study, the effectiveness of the PMA method is demonstrated through an application to a complex mass spectrum dataset consisting of 395 measurements obtained from the MassBank database [12]. The numerical results show that the method reliably identifies both dominant and subtle spectral features, while progressively improving the quality of the approximation as the number of monotonic

segments increases.

The remainder of the paper is organized as follows. Section 2 presents examples illustrating peak identification in mass spectrometry data. Section 3 outlines the PMA method, and its decomposition property. Section 4 presents the first application of PMA to peak detection in a real mass spectrum and examines its effectiveness in identifying progressively finer spectral features as the number of monotonic sections increases. The section concludes by stating the important separation property of PMA, which characterizes the relation between the extrema of optimal approximations with successive numbers of monotonic sections. Section 5 summarizes the main findings and outlines directions for further research. Detailed numerical results and plots and the are provided in Appendix A.

2 Numerical Examples in Peak Detection

This section presents three illustrative examples of peak detection in mass spectrometry data in order to introduce basic terminology, notation, and concepts used throughout the paper. The examples illustrate characteristic challenges encountered in spectral analysis, including overlapping peaks, noise interference, and the identification of dominant spectral features. In practical applications, such tasks are typically addressed through automated computational procedures.

Mass-to-Charge Ratio in Mass Spectrometry

Mass spectrometry operates by ionizing molecules, separating the resulting ions according to their mass-to-charge ratio (m/z), and recording their intensities to produce a spectrum. Here, m represents the total ion mass in Daltons (Da) and z is the dimensionless charge number. For singly charged ions ($z = 1$), typical of small molecules, the observed m/z value directly corresponds to the ion's mass. Conversely, larger biomolecules often form multiply charged species; in these cases, the true molecular mass must be reconstructed by accounting for the charge state z and the mass of the ionizing adducts (e.g., protons).

Example 1. Simple Mass Spectrum

Table 1 presents a simple mass spectrum in which significant peaks occur at $m/z = 101, 102$, and 104 . The dominant peak is located at $m/z = 104$ with intensity 600. Relative abundances normalized with respect to this peak are given by

$$\frac{350}{600} \times 100 = 58.3\%, \quad \frac{300}{600} \times 100 = 50.0\%,$$

$$\frac{600}{600} \times 100 = 100.0\%.$$

Table 1: Simple mass spectrum data

m/z (Da)	Intensity (Counts)
100	200
101	350
102	300
103	50
104	600
105	20

Example 2. Overlapping Peaks

Table 2 illustrates the presence of overlapping peaks in the region $m/z \approx 202$ – 203 . Peaks at $m/z = 202$ and $m/z = 203$ have intensities 1400 and 1300, respectively, while the intermediate value at $m/z = 202.5$ suggests partial overlap caused by fragmentation or isotopic effects. Such situations demonstrate the importance of reliable peak discrimination methods.

Table 2: Overlapping peaks

m/z (Da)	Intensity (Counts)
200	150
201	1200
202	1400
202.5	1000
203	1300
204	200

Example 3. Noise Interference

Table 3 demonstrates the effect of noise on peak identification. Applying a threshold of 50 counts suppresses low-intensity fluctuations and identifies significant peaks at $m/z = 301$, 304, and 306, with corresponding intensities 500, 550, and 900. Using the dominant peak at $m/z = 306$ as the reference, the relative abundances are

$$\frac{500}{900} \times 100 = 55.6\%, \quad \frac{550}{900} \times 100 = 61.1\%,$$

$$\frac{900}{900} \times 100 = 100.0\%.$$

Table 3: Noise interference

m/z (Da)	Intensity (Counts)
300	15
301	500
302	5
303	20
304	550
305	10
306	900
307	18

These examples demonstrate the complexity of peak identification in mass spectrometry, where closely spaced spectral components and noise can obscure the underlying structure of the data.

In the next section, the piecewise monotonic approximation method addresses these challenges through the systematic detection of peaks and troughs in noisy spectra, regardless of the complexity of the dataset.

3 PMA for Peak Detection

We formulate peak detection as a piecewise monotonic data approximation problem, in which the turning points of the optimal approximation provide the locations of the detected peaks and troughs.

Let $\{\phi_i : i = 1, 2, \dots, n\}$ denote observations of the values $\{f(x_i) : i = 1, 2, \dots, n\}$ of an unknown function $f(\cdot)$, at strictly increasing abscissae $\{x_i : i = 1, 2, \dots, n\}$. When the measurements are contaminated by noise or random errors, the sequence of first differences $\{\phi_{i+1} - \phi_i : i = 1, 2, \dots, n - 1\}$ generally exhibits substantially more sign changes than the corresponding sequence $\{f(x_{i+1}) - f(x_i) : i = 1, 2, \dots, n - 1\}$. Consequently, direct peak identification based solely on local extrema may become unreliable, since noise and small-scale fluctuations can generate numerous spurious extrema.

Motivated by this observation, for some prescribed integer k much smaller than n , we seek a sequence $\{y_i : i = 1, 2, \dots, n\}$ that provides a least-squares approximation to the data while restricting the number of sign changes in $\{y_{i+1} - y_i : i = 1, 2, \dots, n - 1\}$ to at most $k - 1$. We represent the data and the approximation as vectors ϕ and y , respectively, in $\mathbb{R}^n$. The constraints imposed on y generate at most k monotonic segments with alternating increasing and decreasing behavior. Without loss of generality, the first segment is assumed to be increasing.

We call *feasible* any vector $y \in \mathbb{R}^n$ satisfying the

piecewise monotonicity constraints

$$\left. \begin{aligned} y_{t_{j-1}} &\leq y_{t_{j-1}+1} \leq \dots \leq y_{t_j}, & j \text{ odd,} \\ y_{t_{j-1}} &\geq y_{t_{j-1}+1} \geq \dots \geq y_{t_j}, & j \text{ even,} \end{aligned} \right\} \quad (1)$$

where the indices $\{t_j : j = 1, 2, \dots, k-1\}$ satisfy

$$1 = t_0 \leq t_1 \leq \dots \leq t_k = n. \quad (2)$$

The objective is to determine a feasible vector $\mathbf{y}^*$ minimizing the functional of the squares of the residuals

$$\gamma(k, n) := \|\mathbf{y} - \boldsymbol{\phi}\|_2^2 = \sum_{i=1}^n (y_i - \phi_i)^2. \quad (3)$$

Such a vector $\mathbf{y}^*$ is called an *optimal piecewise monotonic approximation* of $\boldsymbol{\phi}$.

As already mentioned, the optimization problem is inherently combinatorial because the turning-point indices $\{t_j\}$ are themselves unknown variables. A direct exhaustive search over all possible breakpoint configurations requires consideration of $O(n^{k-1})$ possibilities and rapidly becomes computationally infeasible even for moderate values of k .

A key feature of the PMA method is that the global optimization problem admits a decomposition into smaller monotonic approximation subproblems over contiguous intervals. The structural properties established by Demetriou and Powell [7] are summarized below.

(i) If both $\mathbf{y}$ and the associated integer variables $\{t_j : j = 0, 1, \dots, k\}$ are optimal, and $\boldsymbol{\phi}$ is not feasible, then the indices $\{t_j\}$ are distinct.

(ii) For an optimal approximation $\mathbf{y} = \mathbf{y}^*$, the values at the turning points satisfy the interpolation conditions

$$y_{t_j} = \phi_{t_j}, \quad j = 1, 2, \dots, k-1. \quad (4)$$

(iii) If $\mathbf{y} = \mathbf{y}^*$ is an optimal approximation that contains a monotonic increasing segment on $[t_{j-1}, t_j]$, then the components $\{y_i : i = t_{j-1}, \dots, t_j\}$ minimize

$$\sum_{i=t_{j-1}}^{t_j} (y_i - \phi_i)^2$$

subject to the monotonicity constraints

$$y_i \leq y_{i+1}, \quad i = t_{j-1}, \dots, t_j - 1.$$

Similarly, for decreasing segments the inequalities are reversed. Hence, once the optimal turning points are determined, the approximation can be

constructed through independent subproblems on adjacent intervals.

To facilitate the analysis, define

$$\alpha(t_{j-1}, t_j) = \min_{y_{t_{j-1}} \leq \dots \leq y_{t_j}} \sum_{i=t_{j-1}}^{t_j} (y_i - \phi_i)^2, \quad j \text{ odd,} \quad (5)$$

and

$$\beta(t_{j-1}, t_j) = \min_{y_{t_{j-1}} \geq \dots \geq y_{t_j}} \sum_{i=t_{j-1}}^{t_j} (y_i - \phi_i)^2, \quad j \text{ even.} \quad (6)$$

The optimization problems in (5) and (6) are strictly convex quadratic programming problems with unique solutions. Specialized algorithms developed by [7] solve these subproblems substantially more efficiently than general-purpose quadratic programming methods (see, for example, Fletcher [9]). In particular, Algorithm 1 of [7] computes the corresponding monotonic approximation and all values $\alpha(t_{j-1}, i)$ for $i = t_{j-1}, \dots, t_j$ in only $O(t_j - t_{j-1})$ operations.

Further, the following theorem establishes the decomposition structure underlying the PMA method.

Theorem 1. (Demetriou [4]) Let $k > 1$, let $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots, \tau_k)$ satisfy conditions (2), after replacing t_j by τ_j , and let $\mathbf{y}(\boldsymbol{\tau})$ denote the vector $\mathbf{y}$ minimizing the sum of squares (3) subject to the condition that each segment $\{y(\boldsymbol{\tau})_i : i = \tau_{j-1}, \dots, \tau_j\}$ is monotonic (increasing for odd j , decreasing for even j). Then $\mathbf{y}(\boldsymbol{\tau})$ minimizes (3) over the set of feasible vectors if and only if

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^k \alpha(\tau_{j-1}, \tau_j) + \sum_{\substack{j=1 \\ j \text{ even}}}^k \beta(\tau_{j-1}, \tau_j) = \min_{1=s_0 \leq \dots \leq s_k=n} \left\{ \sum_{\substack{j=1 \\ j \text{ odd}}}^k \alpha(s_{j-1}, s_j) + \sum_{\substack{j=1 \\ j \text{ even}}}^k \beta(s_{j-1}, s_j) \right\}. \quad (7)$$

□

The importance of Theorem 1 lies in establishing that the piecewise monotonic optimization problem, despite its combinatorial nature, can be decomposed into tractable monotonic approximation subproblems over contiguous intervals, thus providing the structure required for an efficient dynamic programming solution. It remains to determine the optimal turning-point indices $\tau_1, \tau_2, \dots, \tau_{k-1}$ using this decomposition. To this end, in view of (3) and (7), we express $\gamma(k, n)$, the least sum of squares of

residuals, in the form

$$\gamma(k, n) = \min_{1=s_0 \leq \dots \leq s_k=n} \left\{ \sum_{\substack{j=1 \\ j \text{ odd}}}^k \alpha(s_{j-1}, s_j) + \sum_{\substack{j=1 \\ j \text{ even}}}^k \beta(s_{j-1}, s_j) \right\}. \quad (8)$$

This definition is useful because, Theorem 1 implies that if $\mathbf{y}(\boldsymbol{\tau})$, for k odd, is optimal, then τ_{k-1} satisfies the equation

$$\gamma(k-1, \tau_{k-1}) + \alpha(\tau_{k-1}, n) = \min_{1 \leq s \leq n} \left\{ \gamma(k-1, s) + \alpha(s, n) \right\}, \quad k \text{ odd}. \quad (9)$$

Therefore in order to calculate $\gamma(k, n)$, which is the least value of (3), we begin with the values $\gamma(1, t) = \alpha(1, t)$, for $t = 1, 2, \dots, n$ and proceed by applying the dynamic programming formulae

$$\gamma(m, t) = \begin{cases} \min_{1 \leq s \leq t} [\gamma(m-1, s) + \alpha(s, t)], & m \text{ odd} \\ \min_{1 \leq s \leq t} [\gamma(m-1, s) + \beta(s, t)], & m \text{ even}, \end{cases} \quad (10)$$

and storing $\sigma(m, t)$, which is the value of s that minimizes the right hand term of expression (10), for $t = 1, 2, \dots, n$, as $m = 2, 3, \dots, k$. Then $\gamma(k, n)$ can be found in $O(kn^2)$ computer operations. At the end of the calculation, $m = k$ occurs and the value $\sigma(k, n)$ is the integer τ_{k-1} that is required in equation (9) if k is odd and similarly if k is even.

Hence, we set $t_0 = 1$ and $t_k = n$, and we obtain the sequence of optimal values $\{t_j : j = 1, 2, \dots, k-1\}$ by the backward formula

$$t_{j-1} = \sigma(j, t_j), \text{ for } j = k, k-1, \dots, 2. \quad (11)$$

Thus, the decomposition of Theorem 1 establishes the fundamental connection to dynamic programming (see, Bellman [1]) and avoids the combinatorial explosion (see, for example, Korte and Vygen [11]) associated with exhaustive breakpoint enumeration. Consequently, the PMA method provides a computationally feasible and effective approach for global peak detection, independent of the level of noise, the structural complexity of the dataset, or underlying law.

Further developments of this calculation are presented in [5], while the corresponding Fortran software package is described in [3] and is available through the Collected Algorithms of the ACM Transactions on Mathematical Software (<http://calgo.acm.org/>).

More recently, Demetriou [6] established ordering and separation properties of the optimal turning points as the number of monotonic sections

varies. By exploiting these properties, the dynamic programming formulae (10) can be implemented through a tree structure, reducing the computational complexity to $O(n^2 + kn \log_2 n)$. Both algorithms compute the same optimal approximation; their difference lies in the organization and efficiency of the computation. One of the underlying separation properties is stated by the mass spectrometry application in the following section.

4 An Application: Peak Detection of a Mass Spectrum

In this section, we demonstrate the effectiveness of the piecewise monotonic approximation approach for automatic peak detection in mass spectrometry data. As stated in the introduction, peak detection in mass spectra is particularly challenging due to the presence of noise, overlapping spectral features, local fluctuations, and the complex structure of the underlying signals. These characteristics provide a demanding setting for evaluating the robustness and computational efficiency of optimization-based peak detection methods.

For the numerical experiments, we employed the mass spectrum of “5-Dehydroquinic acid” obtained from the publicly accessible MassBank database [12]. MassBank is a widely used repository containing spectral data for the identification of small chemical molecules in metabolomics, exposomics, and environmental science applications. The spectrum was acquired using a Shimadzu GCMS-2010 Plus single quadrupole gas chromatograph–mass spectrometer operated in positive ion mode using electron ionization (EI-B). The data cover the mass-to-charge interval 85–500 m/z.

The dataset consists of $n = 395$ observations in three-column format. The first column contains the mass-to-charge ratios, the second the corresponding intensities, and the third the relative intensities. In the present study, only the first two columns were used, yielding the data pairs

$$\{(x_i, \phi_i) : i = 1, 2, \dots, n\}.$$

Although the abscissae $\{x_i\}$ are not used in the optimization calculation, they determine the physical locations of the detected peaks. The raw spectrum contains 106 local maxima and 107 local minima, illustrating the difficulty of reliable peak identification based solely on local extremum criteria.

We applied the piecewise monotonic approximation method to this dataset for the values

$$k \in \{2, 4, 8, 16, 32, 48\}.$$

The computations were carried out using the L2WPMA software package [3]. Some numerical

results and plots are presented in the Appendix of the paper.

The detected turning points are summarized in Table 4 as follows. The left part of this table gives the turning points in the dataset obtained by a best fit with $k = 48$ monotonic sections. The right part gives t_j for $k \in \{2, 4, 8, 16, 32, 48\}$. Odd-indexed turning points (indices are in column j) correspond to peaks, whereas even-indexed turning points correspond to troughs of the optimal approximation. Fig. 1 illustrates the principal characteristics of the spectrum. The data contain several dominant peaks together with numerous lower-intensity fluctuations and small-scale oscillations.

The procedure was initially applied with $k = 2$, producing an approximation consisting of two monotonic segments and identifying the dominant peak of the spectrum at the data point index 62, with location

$$x_{62} = 147.15,$$

and corresponding intensity

$$\phi_{62} = 103,154.$$

These values are reported in Table 4 for $k = 2$, and the resulting fit is shown in Fig. 1.

The computation was then repeated for progressively larger values of k to identify additional spectral features. For $k = 4$, one additional peak is identified at the data point index 139, with location

$$x_{139} = 224.15,$$

and corresponding intensity

$$\phi_{139} = 87,944.$$

These values are reported in Table 4 for $k = 4$, and the resulting fit is shown in Fig. 2.

When $k = 8$, two additional peaks are identified, corresponding to the data point indices 48 and 111, with peak locations

$$x_{48} = 133.1, \quad x_{111} = 196.1,$$

and corresponding intensities

$$\phi_{48} = 33,150, \quad \phi_{111} = 40,018.$$

These values are reported in Table 4 for $k = 8$. The resulting approximation is shown in Fig. 3, where the additional peaks improve the fit in regions exhibiting intermediate-scale spectral variation.

Further refinement is obtained for $k = 16$, where the approximation introduces four additional peaks and four troughs. The detected peaks correspond to the data point indices 123, 215, 229, and 301, with peak locations

$$x_{123} = 208.05, \quad x_{215} = 300.2,$$

$$x_{229} = 314.2, \quad x_{301} = 390.25,$$

and corresponding intensities

$$\phi_{123} = 34,676, \quad \phi_{215} = 31,726,$$

$$\phi_{229} = 21,314, \quad \phi_{301} = 23,083.$$

These values are reported in Table 4 for $k = 16$. Fig. 4 illustrates the resulting fit, where the improvement can be seen in several regions of the spectrum.

For $k = 32$, the PMA method identifies sixteen additional turning points, eight of which are peaks that correspond to the data point indices 4, 58, 84, 106, 132, 211, 248, and 323, with peak locations

$$x_4 = 89.05, \quad x_{58} = 143.15,$$

$$x_{84} = 169.15, \quad x_{106} = 191.15,$$

$$x_{132} = 217.15, \quad x_{211} = 296.2,$$

$$x_{248} = 333.25, \quad x_{323} = 417.25,$$

and corresponding intensities

$$\phi_4 = 21,124, \quad \phi_{58} = 17,045,$$

$$\phi_{84} = 17,499, \quad \phi_{106} = 21,625,$$

$$\phi_{132} = 17,590, \quad \phi_{211} = 19,631,$$

$$\phi_{248} = 15,261, \quad \phi_{323} = 16,629.$$

These values are reported in Table 4 for $k = 32$, and the resulting fit is shown in Fig. 5.

Finally, the approximation with $k = 48$ produces sixteen additional turning points comprising eight peaks and eight troughs. The newly identified peaks correspond to the data point indices 183, 199, 257, 285, 297, 312, 375, and 390, with peak locations

$$x_{183} = 268.2, \quad x_{199} = 284.2,$$

$$x_{257} = 342.2, \quad x_{285} = 374.25,$$

$$x_{297} = 386.2, \quad x_{312} = 402.2,$$

$$x_{375} = 476.3, \quad x_{390} = 492.3,$$

and corresponding intensities

$$\phi_{183} = 10,626, \quad \phi_{199} = 7,511,$$

$$\phi_{257} = 7,721, \quad \phi_{285} = 7,355,$$

$$\phi_{297} = 18,732, \quad \phi_{312} = 12,713,$$

$$\phi_{375} = 10,143, \quad \phi_{390} = 9,833.$$

These values are reported in Table 4 for $k = 48$, and the resulting fit is shown in Fig. 6.

The numerical results clearly show that the PMA method effectively captures both dominant and subtle spectral features, while providing a globally optimized representation of the spectral variation.

To assess the quality of the approximation, two diagnostic measures are reported in Table 4. The first is the least-squares residual $\gamma(k, n)$, defined in expression (3), while the second is the maximum absolute residual

$$D = \max_{1 \leq i \leq n} |y_i - \phi_i|.$$

For example, when $k = 8$, the corresponding diagnostic values are

$$\gamma(8, n) = 6.56 \times 10^9, \quad D = 2.90 \times 10^4.$$

As shown in Table 4, the approximation accuracy improves consistently as the number of monotonic segments increases. In particular, the least-squares residual decreases from

$$1.63 \times 10^{10}$$

for $k = 2$ to

$$5.32 \times 10^8$$

for $k = 48$, while the maximum absolute residual decreases from

$$8.26 \times 10^4$$

to

$$6.13 \times 10^3.$$

These results confirm that the PMA method achieves a strictly progressive refinement across successive approximations in a mass spectrum, not only ensuring robust peak identification, but also remaining free from oscillatory artifacts like those found in wavelet or Fourier representations [10]. Furthermore, it is shown the accuracy of the PMA method at identifying peaks progressively according to their relative significance in the approximation at each stage of the computation.

A noteworthy feature of the results in Table 4 concerns the relation between the extrema obtained as the number of monotonic sections increases. The maxima (and similarly the minima) of an optimal approximation with $k - 2$ monotonic sections often appear to be retained in the optimal approximation with k monotonic sections. Although this behavior is frequently observed in practice, the theoretical result is more general. Specifically, if $t_i^{(k)}$ denotes the i th turning point of an optimal approximation with k monotonic sections, then

$$t_i^{(k)} \leq t_i^{(k-2)} \leq t_{i+2}^{(k)}, \quad i = 0, 1, \dots, k-2, \quad (12)$$

which is the separation property established in [7] and further developed in [6]. Thus, every extremum of the optimal approximation with $k - 2$ monotonic sections is separated by two corresponding extrema of the optimal approximation with k monotonic sections.

5 Conclusion

The application of the piecewise monotonic approximation method to automatic peak detection in a mass spectrometry data was investigated. Peak detection in mass spectra constitutes a challenging signal-processing problem due to the presence of noise, overlapping spectral features, local fluctuations, and the intrinsic complexity of the underlying signals. The numerical results demonstrate that PMA provides a reliable and effective methodology for identifying both dominant and subtle spectral features in complex and noisy spectra.

The PMA approach formulates peak detection as a constrained least-squares optimization problem in which the approximation consists of alternating monotonic segments. Although the underlying optimization problem is combinatorial in nature, the decomposition theorem stated in Section 3 enables the reduction of the global problem into tractable monotonic approximation subproblems over contiguous intervals. This decomposition provides the basis for dynamic programming algorithms that compute globally optimal approximations with quadratic-time complexity, replacing the exhaustive search over combinatorially many breakpoint configurations by an efficient recursive computation.

The methodology was applied to a real mass spectrum dataset consisting of $n = 395$ observations obtained from the MassBank database. The spectrum contains a substantial number of local extrema, making reliable peak identification particularly difficult. By progressively increasing the number of monotonic segments, the PMA approach identified additional peaks and systematically refined the approximation stage by stage. Both the least-squares residuals and the maximum absolute residuals decreased consistently as k increased, indicating progressively improved agreement between the approximation and the observed data. The computational cost is remarkably low. All the optimal approximations for $k \in \{2, 4, \dots, 48\}$ were obtained in only a few milliseconds on a standard personal computer.

The results demonstrated that piecewise monotonic approximation provides a robust, accurate, and computationally efficient approach for automatic peak detection in mass spectrometry. To the best of our knowledge, this is the first application of the PMA approach to this challenging signal-processing problem. The findings indicate that PMA offers a new optimization paradigm for peak detection in mass spectrometry.

The present study provides a proof of concept

for applying the deterministic PMA optimization approach to automatic peak detection in mass spectrometry. Future work will consider broader and more complex mass spectra, systematic comparisons with established peak-detection methods, and statistical and simulation studies under different noise models. Further developments will address automatic selection of the number of monotonic segments and extensions to large-scale and multidimensional spectral data, providing a broader assessment of the robustness and computational performance of PMA.

Acknowledgment:

We sincerely thank the reviewers for their evaluation and constructive feedback, which has helped improve the quality and presentation of the paper.

References:

- [1] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.
- [2] K. R. Coombes, S. Tsavachidis, J. S. Morris, K. A. Baggerly, M.-C. Hung and H. M. Kuerer, “Improved Peak Detection and Quantification of Mass Spectrometry Data Acquired from Surface-Enhanced Laser Desorption and Ionization by Denoising Spectra with the Undecimated Discrete Wavelet Transform,” *Proteomics*, vol. 5, pp. 4107–4117, 2005.
- [3] I. C. Demetriou, “Algorithm 863: L2WPMA, a Fortran 77 Package for Weighted Least-Squares Piecewise Monotonic Data Approximation,” *ACM Transactions on Mathematical Software*, vol. 33, no. 1, pp. 1–19, 2007.
- [4] I. C. Demetriou, “A Decomposition Theorem for the Least Squares Piecewise Monotonic Data Approximation Problem,” in *Approximation and Optimization: Algorithms, Complexity and Applications*, I. C. Demetriou and P. M. Pardalos, Eds., Springer Optimization and Its Applications, vol. 145, Springer, Cham, pp. 119–134, 2019.
- [5] I. C. Demetriou, “Discrete Piecewise Monotonic Approximation by a Strictly Convex Distance Function,” *Mathematics of Computation*, vol. 64, pp. 157–180, 1995.
- [6] I. C. Demetriou, “A Binary Search Algorithm for Best Piecewise Monotonic Data Approximation,” *Journal of Global Optimization*, vol. 82, no. 4, pp. 691–726, 2022. doi:10.1007/s10898-021-01042-x.
- [7] I. C. Demetriou and M. J. D. Powell, “Least Squares Smoothing of Univariate Data to Achieve Piecewise Monotonicity,” *IMA Journal of Numerical Analysis*, vol. 11, pp. 411–432, 1991.
- [8] P. Du, W. A. Kibbe and S. M. Lin, “Improved Peak Detection in Mass Spectrum by Incorporating Continuous Wavelet Transform-Based Pattern Matching,” *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, 2006.
- [9] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, UK, 2003.
- [10] C. Gasquet and P. Witomsky, *Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets*, Springer-Verlag, New York, 1999.
- [11] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Algorithms and Combinatorics, vol. 21, Springer, Berlin, 2018.
- [12] MassBank Consortium, “MassBank: High Quality Spectral Database for Life Sciences,” Available online: massbank.eu/MassBank/RecordDisplay?id=MSBNK-GL_Sciences_Inc-GLS00111, accessed on 15 October 2025.
- [13] C. Yang, Z. He and L. Yu, “Comparison of Public Peak Detection Algorithms for MALDI Mass Spectrometry Data Analysis,” *BMC Bioinformatics*, vol. 10, no. 4, 2009.

Contribution of Individual Authors to the Creation of the Scientific Article (Ghostwriting Policy)

All authors contributed equally to the present research at all stages, including the formulation of the problem, the development of the methodology, the numerical experiments, and the preparation of the manuscript.

Sources of Funding for Research Presented in the Scientific Article

No funding was received for conducting this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest relevant to the content of this article.

Creative Commons Attribution License 4.0 (CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0 https://creativecommons.org/licenses/by/4.0/deed.en_US

Appendix – Numerical Results and Plots

Table 4: Left four columns: Turning points in the 5-Dehydroquinic Acid mass spectrum by a best fit with $k = 48$ monotonic sections. Right six columns: The turning point indices t_j of the best fit for $k \in \{2, 4, 8, 16, 32, 48\}$.

j	t_j	x_{t_j}	Intensity (ϕ_{t_j})	$k =$	2	4	8	16	32	48
0	1	86.10	486		1	1	1	1	1	1
1	4	89.05	21,124						4	4
2	36	121.10	354						36	36
3	48	133.10	33,150				48	48	48	48
4	52	137.20	734				52	52	52	52
5	58	143.15	17,045						58	58
6	61	146.25	1,218						61	61
7	62	147.15	103,154		62	62	62	62	62	62
8	77	162.20	460						77	77
9	84	169.15	17,499						84	84
10	103	188.00	415			103	103	103	103	103
11	106	191.15	21,625						106	106
12	110	195.15	1,403						110	110
13	111	196.10	40,018				111	111	111	111
14	117	202.10	689				117	117	117	117
15	123	208.05	34,676					123	123	123
16	128	213.10	638					128	128	128
17	132	217.15	17,590						132	132
18	135	220.10	337						135	135
19	139	224.15	87,944			139	139	139	139	139
20	178	263.20	105							178
21	183	268.20	10,626							183
22	194	279.20	29							194
23	199	284.20	7,511							199
24	210	295.20	105					210	210	210
25	211	296.20	19,631						211	211
26	214	299.25	1,035						214	214
27	215	300.20	31,726					215	215	215
28	224	309.20	65					224	224	224
29	229	314.20	21,314					229	229	229
30	236	321.20	34						236	236
31	248	333.25	15,261						248	248
32	254	339.20	102							254
33	257	342.20	7,721							257
34	266	352.10	14					266	266	266
35	285	374.25	7,355							285
36	294	383.20	18							294
37	297	386.20	18,732							297
38	300	389.35	1,512							300
39	301	390.25	23,083					301	301	301
40	308	397.20	6						308	308
41	312	402.20	12,713							312
42	319	412.20	13							319
43	323	417.25	16,629						323	323
44	353	453.30	2							353
45	375	476.30	10,143							375
46	381	483.30	10							381
47	390	492.30	9,833							390
48	395	497.30	60		395	395	395	395	395	395
				$\gamma(k, n) =$	$1.63_{10}^{\dagger}$	9.07_9	6.56_9	3.44_9	1.40_9	5.32_8
				$D =$	8.26_4	3.43_4	2.90_4	1.86_4	1.07_4	6.13_3

[†] Subscripts indicate multiplication by 10 raised to the power of the subscript

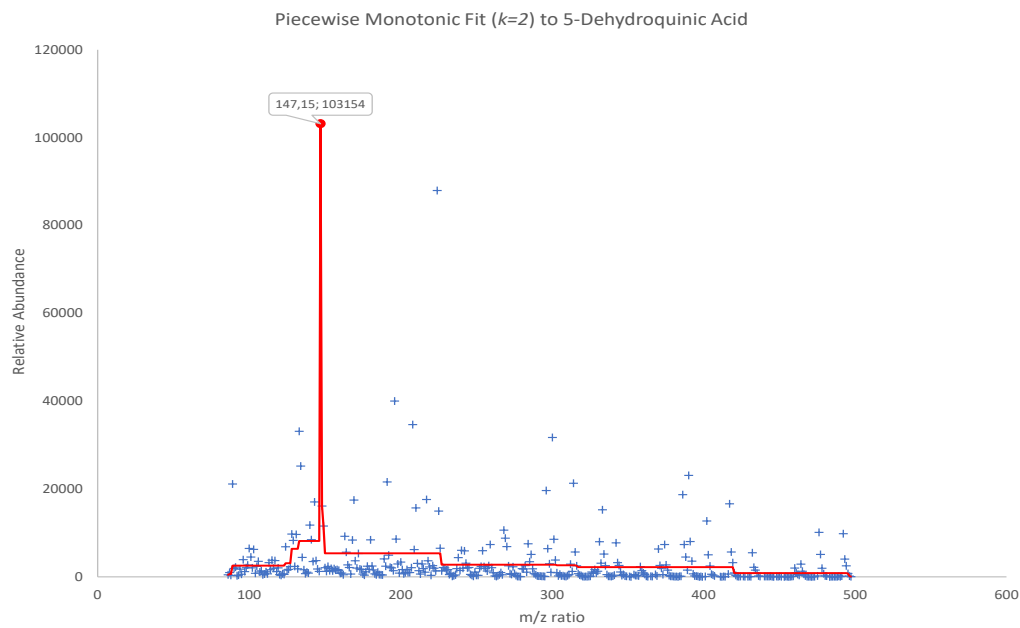


Figure 1: Detected peak (circle) by a best fit with $k = 2$ to the 5-Dehydroquinic Acid mass spectrum data (plus signs, $n = 395$). The solid line illustrates the best fit

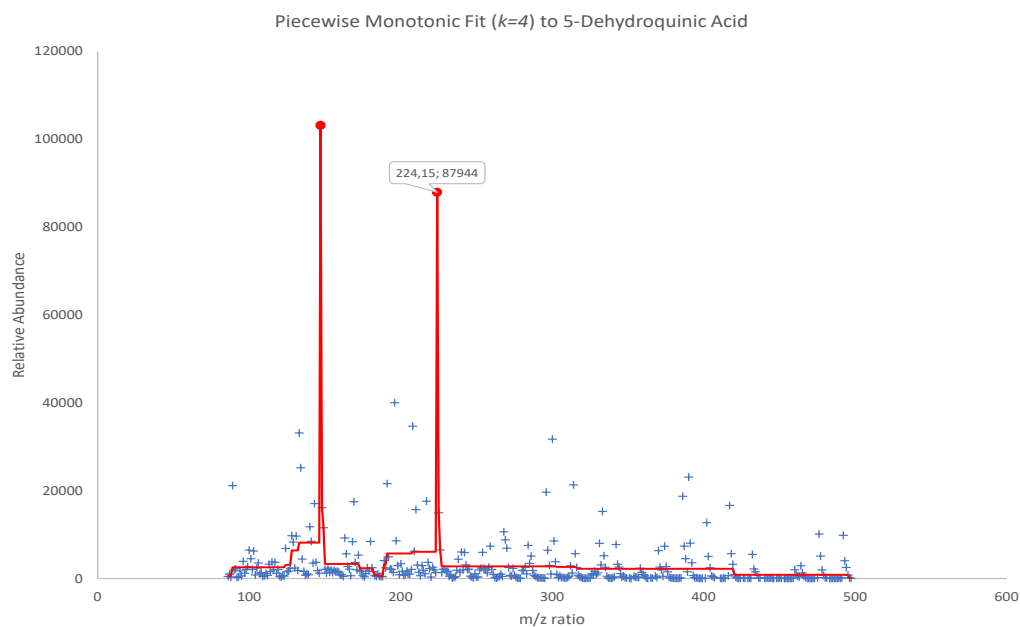


Figure 2: As in Fig. 1, but the detected turning points are found by a best fit with $k = 4$

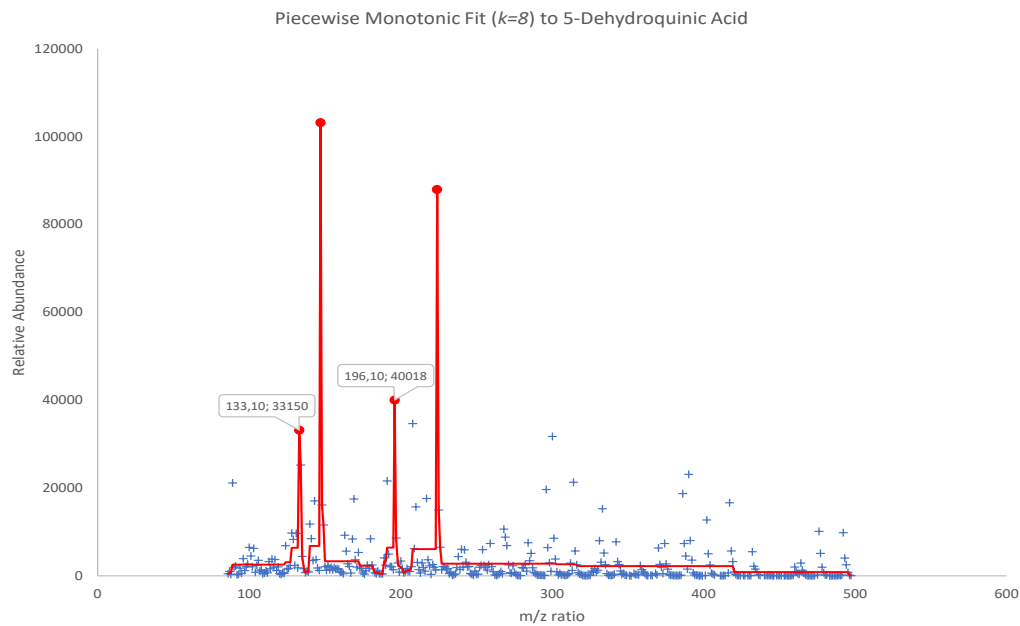


Figure 3: As in Fig. 1, but the detected turning points are found by a best fit with $k = 8$

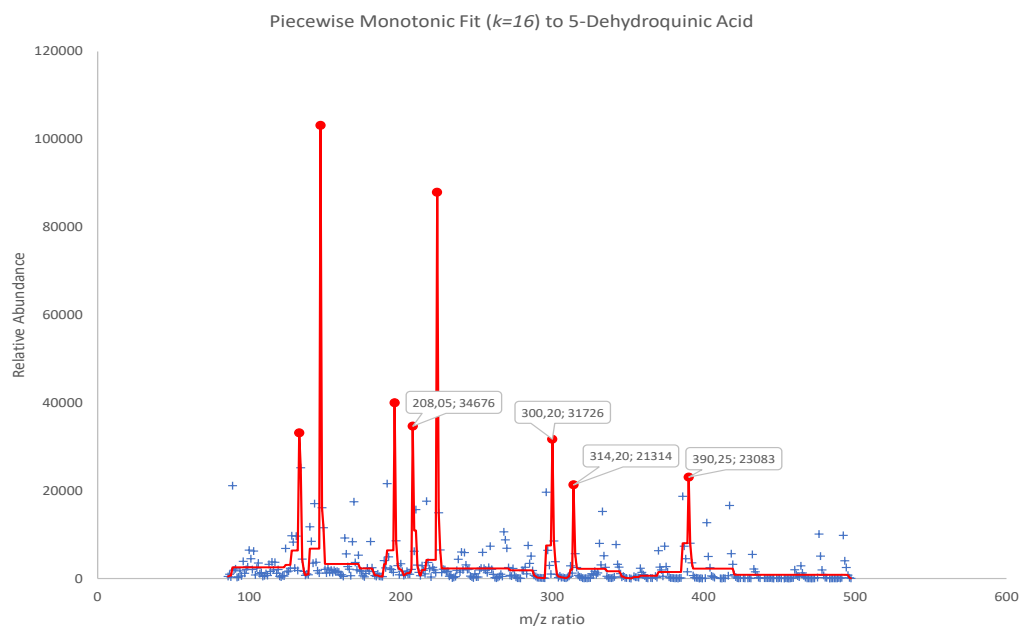


Figure 4: As in Fig. 1, but the detected turning points are found by a best fit with $k = 16$

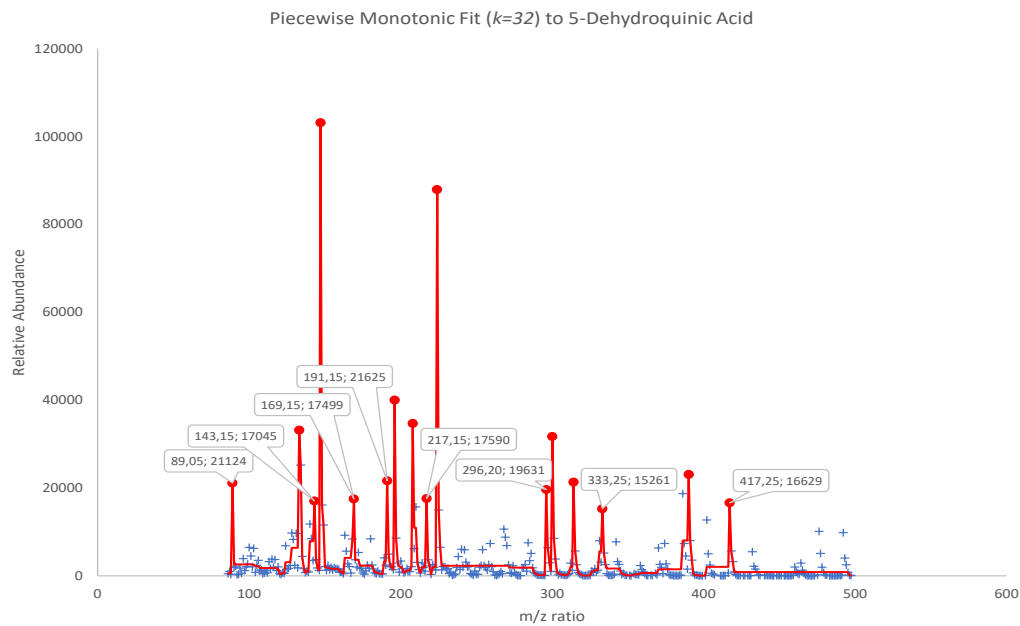


Figure 5: As in Fig. 1, but the detected turning points are found by a best fit with $k = 32$

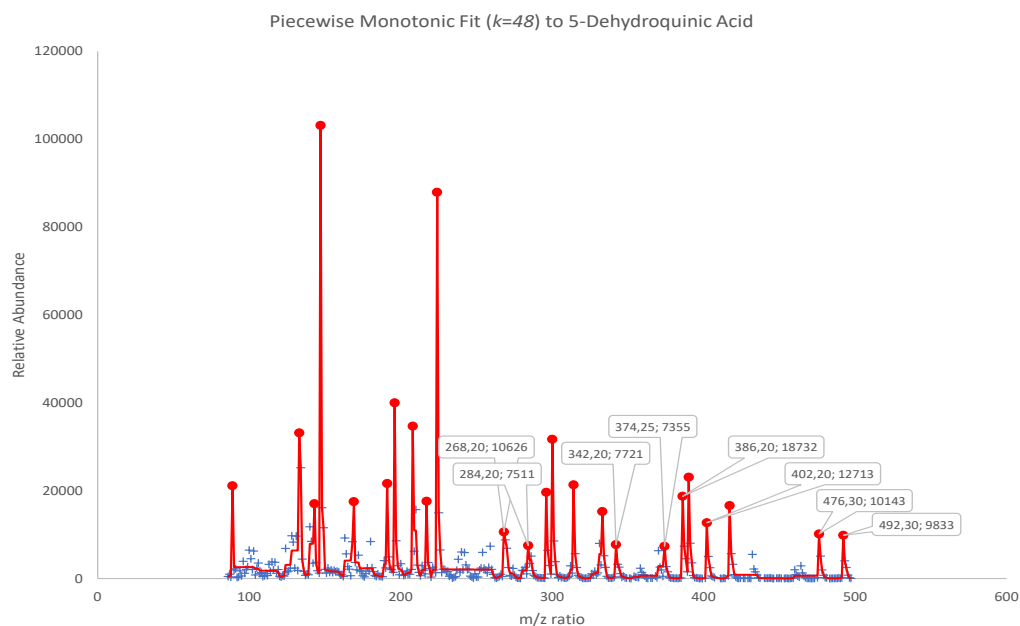


Figure 6: As in Fig. 1, but the detected turning points are found by a best fit with $k = 48$