

Adaptive Polynomial Approximation of Chromatographic Peaks

Yuri Kalambet, Ampersand International, Inc.,
Beachwood, OH

Sergey Maltsev, Dmitry Gladkov, Ampersand Ltd.,
Moscow, Russia;

Yuri Kozmin, Shemyakin Institute of Bioorganic
Chemistry, Moscow, Russia

[mailto: kalambet@ampersand-intl.com](mailto:kalambet@ampersand-intl.com)

Peak parameters to be “integrated”

Parameter	Method of evaluation
Area	Integration
Height	Approximation
Retention (apex X position)	Approximation
Width at half-height	Approximation
Width baseline	Approximation
Asymmetry (5% and 10%)	Approximation

Most of peak parameters are obtained by approximation, not integration.

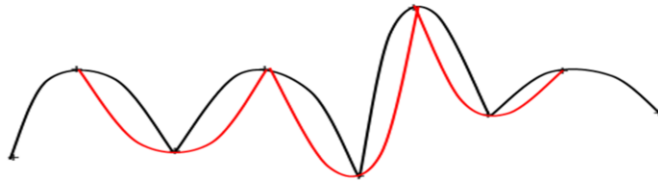
Area Integration

Integration method	Formula
Plain sum	$A = \sum R_i$
Trapezoidal rule	$A = \frac{1}{2} \sum (R_i + R_{i+1})$
Simpson's rule	$A = \frac{1}{6} \sum (R_i + 4R_{i+1} + R_{i+2})$
Simpson's rule $\frac{3}{8}$	$A = \frac{1}{8} \sum (R_i + 3R_{i+1} + 3R_{i+2} + R_{i+3})$

There are several integration formulas, but they originate mostly from interpolation (all available data points are known without error), not approximation (points have significant error).

About some myths concerning peak integration

- The results of integration by different integration methods (trapezoid rule, Simpson's rule, etc) are better for more complicated rules?
- Integration by "blocks" of N points – some points are better than other?
- Average of N integrations, each shifted one point makes points equally important and may improve results



Integration with averaging

- 1331 Simpson's $\frac{3}{8}$ rule
 - 1331133113311331 Unequal weights
 - 1331 Averaged weights
- 1331
--1331
---1331
----1331
-----1331
147**888**741

5

In the case when approximation scheme is applied with shift and results are averaged, only point weight at edges of the integrated region differ for different calibration schemes

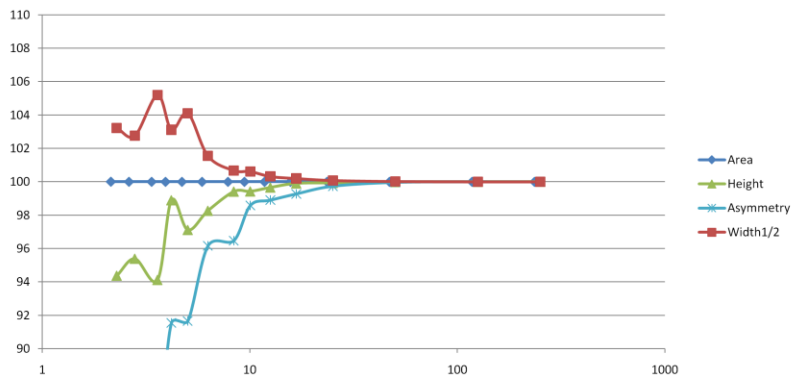
All methods are identical after averaging

- Peak boundaries are at the baseline – all values close to an end are (almost) zero!
Base-to-base peaks are integrated by mere summing up of all points
- Peak baseline drop separation using Simpson's formulas is not "WYSIWYG"
- The only "WYSIWYG" formula for baseline drop separation is trapezoidal rule

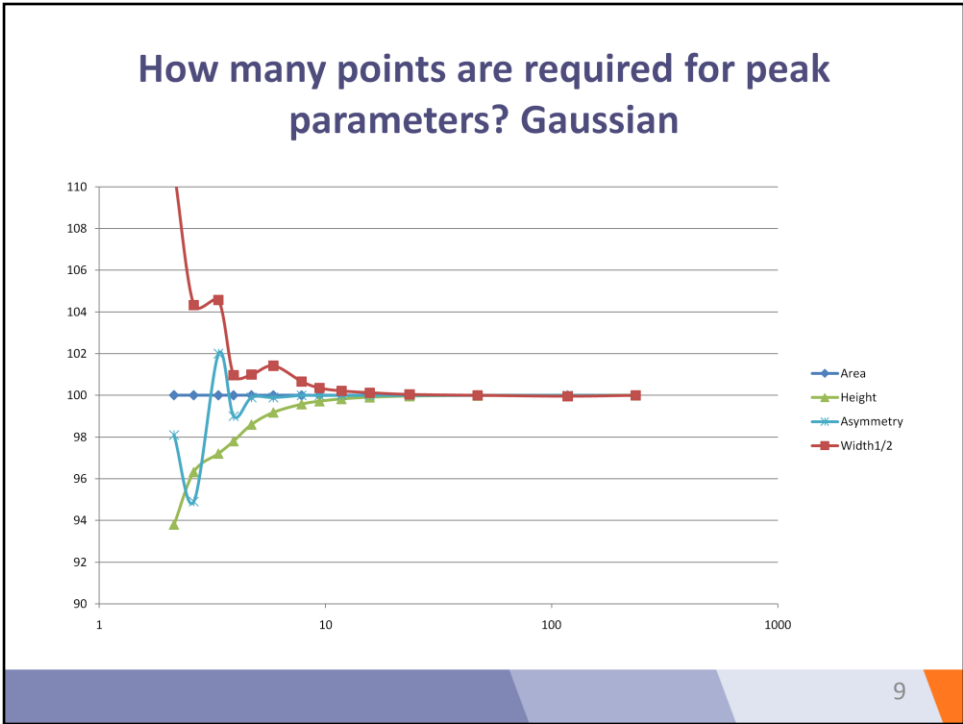
Why simple summation works so well?

- Integral of all derivatives over the peak region equals zero, as all derivatives are equal to zero at the end of the peak. All convex regions are exactly compensated by concave regions.

How many points are required for peak parameters? EMG, asymmetry = 3.46

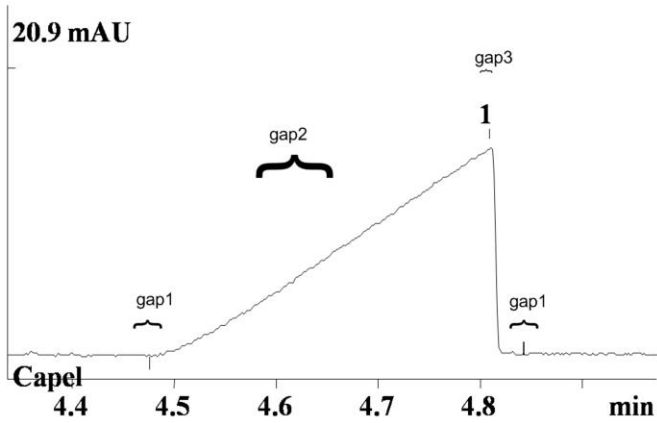


The most sensitive to the number of points per peak parameter is asymmetry. Area does not change at all.



Number of points per peak half-width depends on the required accuracy of the parameter being evaluated.

Peak approximation slits

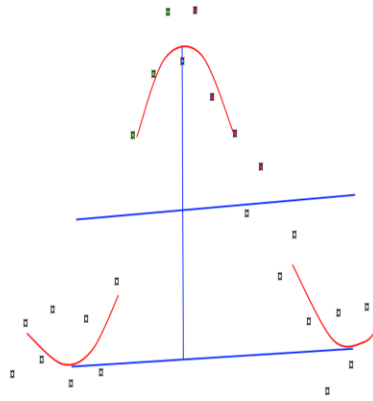


10

Approximation of the strongly asymmetric Capillary Electrophoresis peak. Different gaps may be used for approximation of different peak parts, providing maximum noise reduction without disturbing peak shape.

Rough slope width estimate

- Evaluate baseline using default gap (minimum peak width Integration parameter)
- Evaluate peak height using default gap
- Count all points from peak apex to slope end with height bigger than half-height of the peak. Count obtained is an estimate of the slope width.



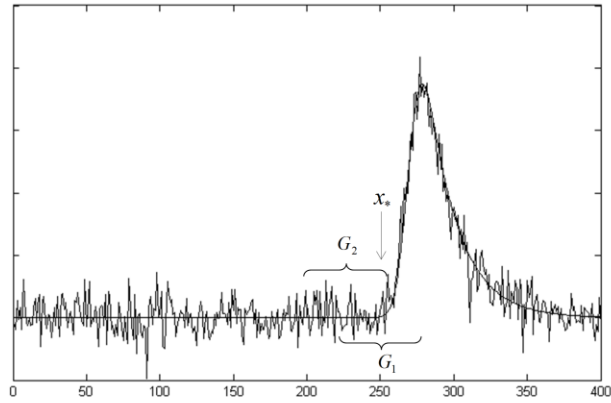
Description of the algorithm, used for rough slope width estimation.

Properties of adaptive peak approximation

- Good noise suppression at each slope
- Minimal peak shape disturbances
- All peak parameters are resistant to oversampling
- Baseline approximation may be poor – either noisy (small gap) or disturbed (large gap).
- No approximation outside of peaks
- **Baseline position is one of the most important sources of error**

Pros and cons of the adaptive peak approximation, described above

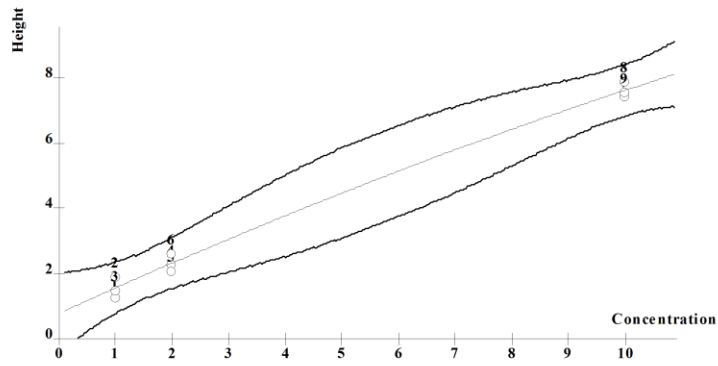
Improvement 1: Non-central approximation



13

New algorithm of noise filtering: some points, especially belonging to baseline, can be approximated by non-central approximation much better, than by central

Confidence intervals



Confidence interval can help in deciding, which approximation is better.

Confidence interval approximation

$$C_Y = t_{n-p}^{(1/2)\alpha} \cdot S \cdot \sqrt{u_*}$$

where

$$S^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \cdot (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-p} \quad u_* = \mathbf{x}'_* (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{x}_*$$

n - number of data points used for polynomial approximation (gap of the filter);

p - power of the polynomial;

\mathbf{X} - matrix of x power values on independent axis (time);

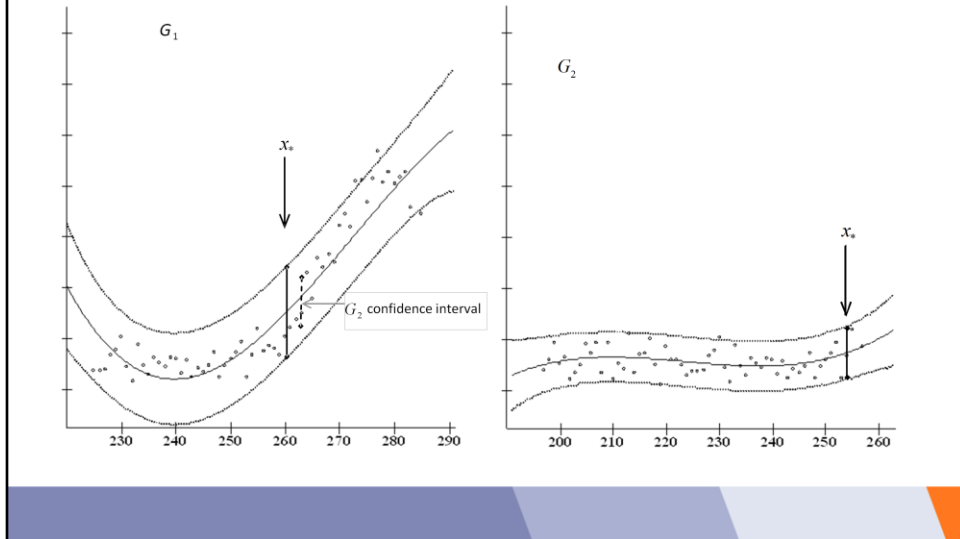
\mathbf{Y} - vector of detector response values;

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{X}' \cdot \mathbf{Y} \quad \mathbf{x}'_* = \{1, x_*, \dots, x_*^p\}$$

t_m^δ - Student's coefficient for confidence probability $(1-\delta)$ and m degrees of freedom

x_* - position at which smoothed (approximated) value is estimated.

Approximation using confidence intervals



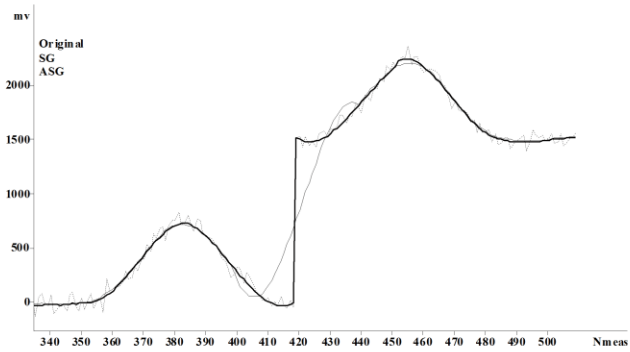
Non-central approximation of some points sometimes really gives better confidence interval, than central.

Algorithm of simple Confidence filter approximation

- Evaluate points and confidence intervals for new (shifted) window
- Compare new confidence interval with that for previously evaluated point. If the new one is smaller than previous, replace approximated point and its confidence interval.
- Computational complexity of Confidence filter is comparable with that of simple convolution, (e.g. Savitzky-Golay) and linearly depends on the product ***gap · (degree of the polynomial)***.

Outline of simplest Confidence filter algorithm.

Bonus #1: handling baseline steps



dotted – raw data; thick line – Confidence Filter;
thin line – Savitzky-Golay filter

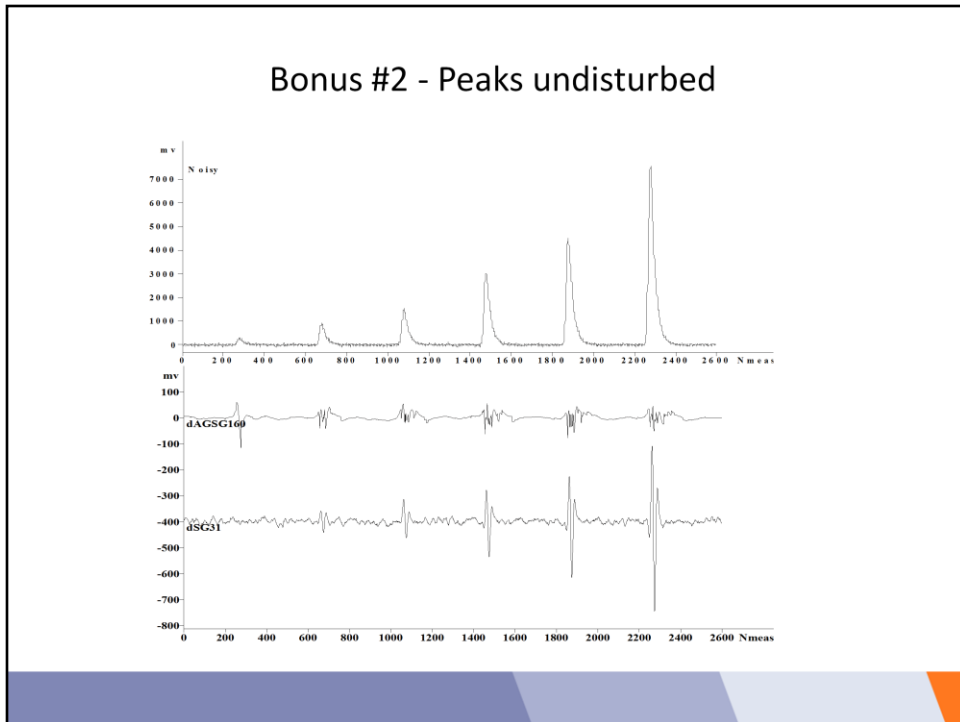
Advantage #1 – baseline steps are treated properly

Confidence filter algorithm improvement: Adaptive gap of the polynomial

- Repeat confidence filter algorithm for approximations with different windows (gaps)
- Computational complexity: **$degree \cdot gap \cdot (gap - 1) / 2$**
- Logarithmic step: next gap is k times smaller, than previous, e.g. $gap_2 = gap_1 / k$, $k > 1$;
Computational complexity: **$degree \cdot gap \cdot k / (k - 1)$**

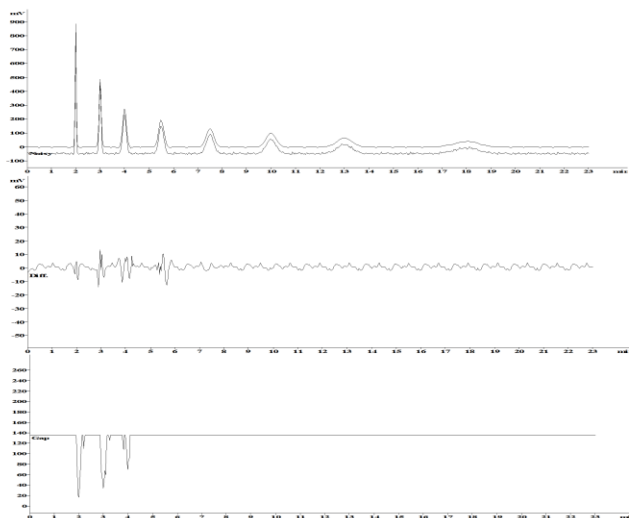
An obvious improvement allows varying gaps.

Bonus #2 - Peaks undisturbed



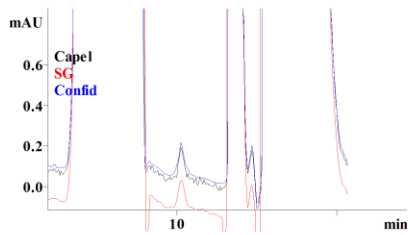
Quality of filtering does not degrade for large maximal gap. Upper picture presents the signal being filtered. Lower picture presents difference between initial noise-free and filtered signal for Savitzky-Golay method with gap 31 (lower curve) and Adaptive Confidence Filter gap 161 (upper curve). Peak disturbance almost does not depend on peak height for the Confidence filter and is proportional to the peak height for Savitzky-Golay one.

Example : Isocratic Chromatography

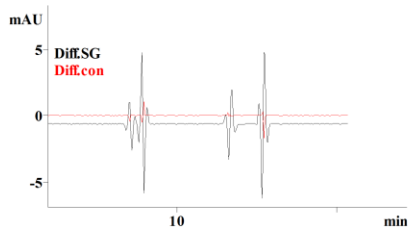


Approximation of the isocratic chromatogram. Upper graph shows model isocratic data, middle – difference between initial and filtered signal (10x amplified) and lower – gap, used for approximation of every point. Default value of 139 falls down to 31 near the top of narrow peaks.

Practical example: Capillary Electrophoresis



- Optimal improvement of sensitivity and detection limit does not lead to disturbance in large peaks



Aggressive noise reduction, used to suppress baseline noise for better quantification of the small peak, leads to rather small disturbance of the peak shape (compared to the case of Savitzky-Golay filtering)

Automatic selection of the gap:

- Small gaps: is pump pulsation a noise or a signal?
- Small gaps: accidental perfect fit
- Small gaps: confidence interval depends on confidence level
- Large gaps: treating small peaks as a noise

Questions behind the scene of noise filtering

Conclusions:

- Confidence filter is a natural extension of one of the best conventional noise filtering methods - Savitzky-Golay filter to the case of non-central approximation, varying gap and degree of the polynomial.
- Confidence filter is metrologically the best noise filtering method for the selected class of approximating functions.
- Image processing: object boundaries unchanged?

Patent pending