Artificial Neural Networks and Expert Systems '95
November 20-23, 1995, Dunedin, New Zealand

Deriving Baseline Detection Algorithms from Verbal Descriptions

Bärbel Herrnberger

Technical University of Ilmenau
Dept. of Computer Science and Automation
98684 Ilmenau, Germany
bahe@informatik.tu-ilmenau.de

Uwe R. Zimmer

University of Kaiserslautern
Computer Science Department
67663 Kaiserslautern, Germany
uzimmer@informatik.uni-kl.de

The presented strategy of automatic baseline detection in chromatograms combines fuzzy logic and neural network approaches. It is based on a verbal description of a baseline referring to a 2D image of a chromatogram instead of a data vector. Baselines are expected to touch data points on the lower border of the chromatogram forming a mainly horizontal and straight line. That description has been translated into a couple of algorithms forming a two–stage approach first proceeding on a local, and second, on a global level.

The first stage assigns a value regarded as the degree of baseline membership or significance to each data point; the second uses a global optimization strategy for coordinating these significances and for producing the final curve, simultaneously.

The statistical stability of the proposed approach is superior to known approaches, while keeping the computational effort low.

1 Motivation

Baseline definition. Chromatographic separation is a widely used technique for quantifying mixtures of substances. The mixture is decomposed into a sequence of their individual components resulting in an intensity vs. time profile – the chromatogram – as shown in figure 1. Substances can be distinguished by the positions of their corresponding peaks; peak area is a measure of the amount of a substance [2].

An ideal chromatogram consists of well–separated Gaussian peaks on a flat baseline. But usually, peak overlap, negative peaks and drifts of varying sign can occur, and even disturbances of the kind shown in figure 3 can be recorded. Thus, for correct peak measurement, and therefore, for correct quantification, a correcting signal, the baseline, is searched for ignoring peak overlap, separating positive peaks from negative ones, and following drifts and ruptures.

Other approaches. Most automatic baseline detection strategies (e.g. [3]) assume a mainly horizontal and straight curve. Following this, baseline candidates are those points having small or zero slope. Thus, minima between overlapping peaks and the tips of negative peaks will be marked by mistake. This problem is usually handled by introducing a threshold upon the drift: the line connecting two adjacent points marked as baseline members may not exceed a given slope. Though including additional data points inthe decision, the strategy remains local.

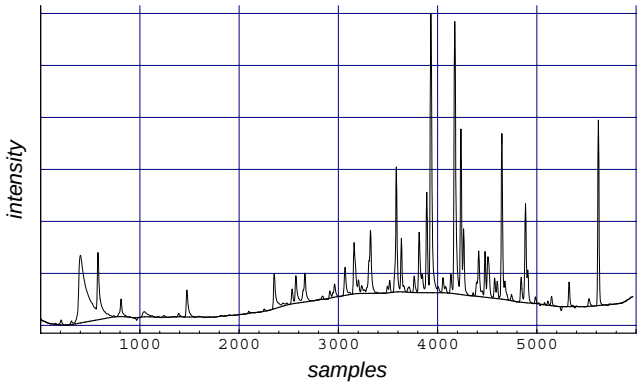

Figure 1: Chromatogram and manually assigned baseline

Threshold criteria on slope and drift are critical in multiple ways: first, a single local decision on baseline membership can change the further course of the baseline completely such that peak measurements of two similar chromatograms will possibly not be comparable; second, in general, a constant threshold will not be applicable to the whole chromatogram: with the threshold being too low, the baseline cannot follow the chromatogram, otherwise, it runs into peak groups. Besides, even data points with higher slope can be baseline members [4].

2 Concept

The proposed strategy is based on the following concepts:

Considering both local and global aspects. Local decisions on baseline membership without a close interplay with the course of the whole chromatogram does not seem reasonable. Considering this fact, a system of two stages is proposed. In

the first, local stage, each data point is assigned a continuous value considered its significance to the baseline. The second, global stage uses an optimization strategy for balancing these significances and for producing the final curve (section 5). The local stage consists of two parts, feature extraction and feature fusion. In the global stage, a nonlinear weighted regression method based on self–organizing maps was applied with weights given by the significances obtained in the local stage.

Extended feature set. Expecting a single feature was not sufficient for a decision on baseline membership, a couple of independent features were determined for each point (section 3).

Handling the chromatogram as a 2D image. This is the basis an expert evaluates a chromatogram. There, baseline can be described as a preferably horizontal signal touching data points at the lower border of the chromatogram, whereby, at a local level, differing orientations are allowed in principle. Using an image, positional relations between data points can be considered.

Feature fusion. The features assigned to each point are fused to a continuous value, its significance for the baseline. In order to avoid false decisions at this stage, threshold functions producing binary significances are not applied. In section 4, a rule describing the dependency of the significances from the local features will be given, from which appropriate membership functions for the corresponding linguistic terms are derived. The parameters of the membership functions and the type of inference is derived from a cost function upon a set of pre–classified data points and their features.

3 Local features

This section describes ideas for the identification of verbal baseline criteria with algorithms for the local stage. The features derived refer to a 2D image of a chromatogram. For the chromatogram being a binary image of some single–valued function, the computational effort from the employed filtering operations is relatively low.

3.1 Distance

The idea is to introduce some blurring into the line image: each point is dilated [5] to an area called the structuring element (figure 2).

The curve resulting from an erosion at the lower border of this chromatogram area (using the same structuring element) is employed as the reference for the following distance measurement, where a feature $m_c \in [0,1]$ describes the vertical distance to its corresponding point on that curve assigned to each data point[1].

1. All features are normalized to the feature extrema of the chromatogram the point belongs to – features are relative

Figure 2: Blurred line image

With line images as the chromatogram, there is no need to perform closing on the image *itself*, and therefore no need for scaling: with rectangular structure elements, dilatation is equivalent to a ‘running minimum’, and erosion with a ‘running maximum’, respectively.

The result of closing is mainly dependent on the size of the structuring element or the width of the 1D filter. With very small values, the resulting line runs into the peaks, otherwise, it will be too stiff.

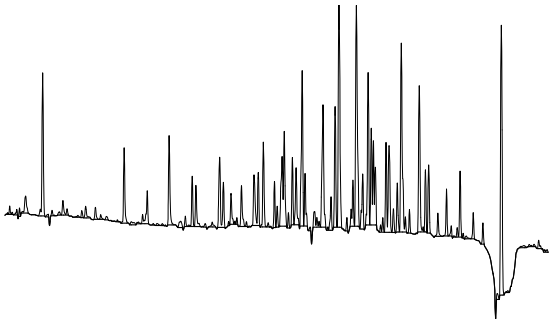

Figure 3: Closing by a rectangular structuring element. Both negative peaks and baseline disturbance are included.

3.2 Point density

Point density corresponds closely to the slope, but there are some significant differences. If the chromatogram is seen as a scatterplot, there is – in general – a higher point density at the baseline. Because a horizontal baseline course is preferred, an anisotropic Gaussian filter is applied to the width, larger than the height. The filtering result is shown in figure 4.

Here, scaling will be needed but it seems to be sufficient to adjust a significant peak to a given height/width–ratio and applying the resulting scaling factor to the whole chromatogram. Scaling is considered a less critical factor because of the continuous filter surface, i.e., small changes in the filter size will produce small density changes.

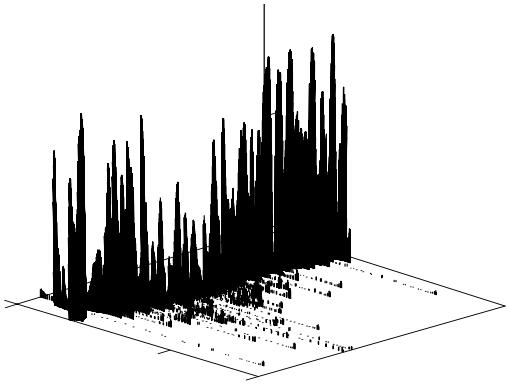

Figure 4: Local point density

4 Feature fusion

The significance of a data point for baseline consideration should be high, if there is a small slope m_s, a small distance m_c, and a high point density, m_d.
The feature analysis supporting this hypothesis is given in section 4.1. Then, feature fusion consists of the following steps:
1. Determining the principal form of the membership functions [6] for both the linguistic terms 'high' and 'small' – the functions $S()$ and $Z()$ were chosen, respectively.

$$S(x, a, \delta) = \begin{cases} 0 & ,x \le a - \delta \\ 2\left(\frac{x - a + \delta}{2\delta}\right)^2 & ,a - \delta < x \le a \\ 1 - 2\left(\frac{x - a + \delta}{2\delta}\right)^2 & ,a < x < a + \delta \\ 1 & ,x \ge a + \delta \end{cases} \quad (1)$$

$$Z(x, a, \delta) = 1 - S(x, a, \delta) \quad (2)$$

2. Optimizing the parameters a (inflection point) and δ (width) of these functions according to a cost function, using a training set of feature vectors $(m_s, m_d, m_c)^T$ together with their classification as baseline or non–baseline member.
3. Specifying an operator for fusing the individual conditions stated in the rule, again derived from a cost function.

4.1 Feature analysis

Figure 5 shows the histograms of the individual features for baseline and non–baseline points, separately. Most baseline points meet the conditions stated above, but there are even non–baseline points with the same features. The values are divided into h=10 classes with class i collecting the values from the interval $[(i\text{-}1)/h, i/h]$; i=1, …, h.

4.2 Optimizing membership functions

Optimization is intended to sharpen the discrimination between baseline and non–baseline points. For each of the

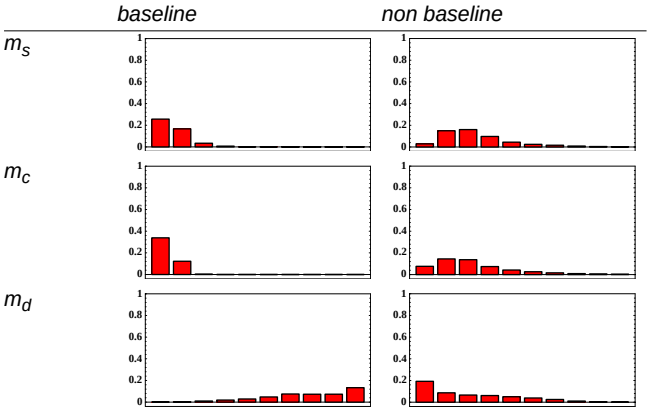

Figure 5: Feature histograms (nonlinear transformation of m_s and m_c into [0,1])

linguistic terms $M \in \{M_s, M_c, M_d\}$ bound to the individual features, the parameter set $\boldsymbol{p} = (a, \delta)^T$ was determined by minimizing the cost function

$$C(M) = \frac{1}{2K} \sum_{k=1}^{K} (1 - M_k)^2 b_k + M^2{}_k (1 - b_k)$$

$$\text{where } M = \mu(m, \boldsymbol{p}),\ M_k = \mu(m_k, \boldsymbol{p}) \quad (3)$$

upon a set of K (manually) pre–classified data points from various chromatograms of different forms. Here, $\mu(m, \boldsymbol{p})$ is the membership function describing M, and M_k is the value of μ applied to feature m of data point k, m_k. Baseline membership b_k is coded binary (b_k = 1 indicating baseline members, b_k = 0, otherwise). This function will penalize both non–baseline points with high membership values and baseline points having small ones.

The optimization result is shown in figure 6. Discrimination has been sharpened, but not all of the baseline points have yet got significantly higher membership values than non–baseline points.

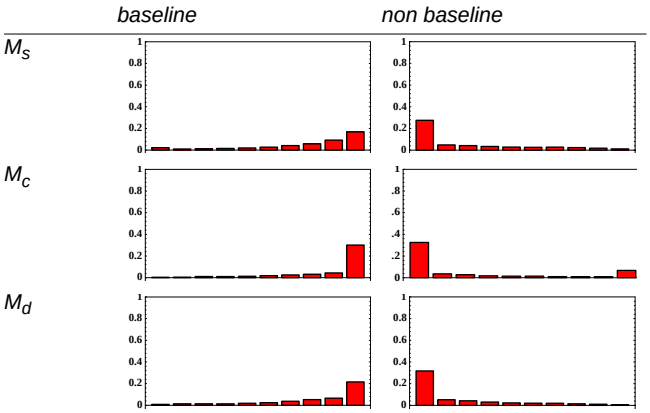

Figure 6: Histogram of fuzzified features

4.3 Fusion

Feature fusion is again derived from (3), with the argument M being replaced by an operator op (M_s, M_c, M_d) and M_k replaced by $s_k^{op} = op\ (M_{s,k}, M_{c,k}, M_{d,k})$, respectively. The value s_k^{op} will now be taken for the significance of point k for baseline.

From the set of non–parametric average operators op (minimum, geometric mean, arithmetic mean, dual–of–geometric–mean, maximum) chosen to cope with the inconsistencies within the features, the geometric mean was found to produce minimal costs. See figure 7 for results. But even feature fusion cannot produce clear decisions on baseline membership. To cope with this, a global optimization procedure is applied for coordinating the significances s_k.

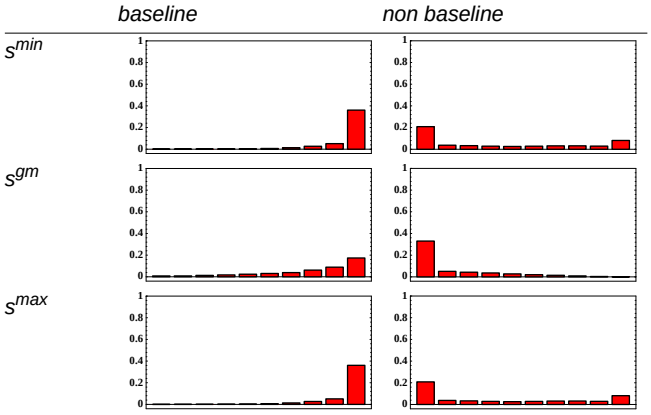

Figure 7: Histogram of significances obtained from applying the minimum (s^{min}), geometric mean (s^{gm}), and maximum (s^{max}) operator

5 Global optimization

Baseline is considered as a straight path through the data including significant points which is obtained here by nonparametric weighted regression on the basis of self–organizing maps for reasons of topology preserving and reflection of the input probability density. Here, probability is identified with significance. The network consists of a chain of nodes with M=2 weights each corresponding to the input data (x, y position of the sampled chromatogram points).

Constraint topological maps. For an approximation of single–valued functions of one or more variables, topology preservation [7] in the space of the independent variables is needed which is not a priori given in the original Kohonen algorithm. The constraints introduced by Cherkassky [1] are:

1. Initialize the weights such that a topological order in the space of independent variables is given. (Here, the initialization is such that the weights project onto a line in the input space connecting the first and the last chromatogram point.)

2. During the learning process, preserve the topological order by searching for the best–matching node in the space of independent variables only.

Thus, adaptation of the weights $\boldsymbol{w}_n$ for each of the N nodes can be found by:

1. Finding the best–matching node c

$$c = c(\boldsymbol{v}_k) = \arg\{min_{n=1}^{N} \|\hat{\boldsymbol{w}}_n - \hat{\boldsymbol{v}}_k\|\} \quad (4)$$

where $\boldsymbol{v}_k = (y_k, x_{k,1}, \ldots, x_{k,M-1})^T$, $\hat{\boldsymbol{v}}_k = (x_{k,1}, \ldots, x_{k,M-1})$

and $\boldsymbol{w}_n = (w_{n,1}, w_{n,2}, \ldots, w_{n,M})^T$, $\hat{\boldsymbol{w}}_n = (w_{n,2}, \ldots, w_{n,M})^T$

2. Adapting all weights including the dependent variable

$$\boldsymbol{w}_n(t) = \boldsymbol{w}_n(t\text{-}1) + \Delta_k \boldsymbol{w}_n(t) \quad (5)$$

$$\Delta_k \boldsymbol{w}_n(t) = s_k\, \eta(t)\, h_{cn}(t)\, (\boldsymbol{v}_k - \boldsymbol{w}_n(t\text{-}1)) \quad (6)$$

with $\eta(t)$ being a learning rate modified by the significances s_k, and h_{cn} being a monotonously decreasing neighborhood function, both being gradually shrunk. The smoothness of the resulting curve depends on the final width of h_{cn}.

Allowing the adaptation of all the weights, the knots of the regression function are found automatically, and, strictly speaking, there is correspondence to principal curves [8], [9].

Baseline construction. The result of weight adaptation is shown in figure 8**.** As the best matching node will be computed in only one dimension, again, here will be no need for scaling. For the regression technique is minimizing the distance of the individual points to the resulting curve, a qualitative course of a baseline is obtained.

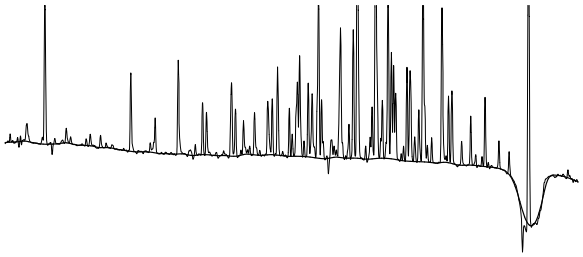

Figure 8: Result of the global stage. The negative peak is excluded from baseline, while the baseline disturbance is included.

6 Discussion

A strategy for baseline detection in chromatography combining fuzzy and neural network methods was proposed. Using a couple of independent features for a local estimation of baseline membership, a larger flexibility compared to local approaches was obtained.

Filter widths were the parameters of the local stage. Only a raw estimation seems to be necessary. Feature fusion parameters were determined in advance from minimizing a cost function.

Using a couple of features, however, does not necessarily lead to a clear decision on baseline membership in the local stage. The global stage, implemented as a self–organizing map, was expected to coordinate the significances obtained from feature fusion and to give an implicit correction of these significances. That is, the course of the resulting curve was adjusted to the majority of the points evaluated as significant.

Results were further improved with the parameters of feature fusion optimized on a training set containing data points of similar chromatogram forms – at the expense of losing generalization.

The proposed strategy was characterized by statistical stability in two ways: First, the resulting curve did not depend on the estimation of a single feature only. Second, identifying the global nature of baseline with a specific smoothing procedure ensures that no single decision on the local stage can change the further course of the baseline completely – the result was usable even under rather bad conditions. Though a couple of extensions were introduced (compared to local approaches), very little additional computational effort is required.

References

[1] V. Cherkassky and H. Lari-Najafi. *Constrained Topological Mapping for Nonparametric Regression Analysis.* Neural Networks, 4:27-40, 1991

[2] A. Braitwaite and F.J. Smith. *Chromatographic Methods.* Chapman & Hall, 1992

[3] *HP3350 User Reference Manual*, Hewlett Packard Company., 1987

[4] N. Dyson. *Chromatographic Integration Methods.* The Royal Society of Chemistry, 1990

[5] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision.* Addison-Wesley, 1993

[6] H.-J. Zimmermann. *Fuzzy set theory and its applications.* Kluwer Academic Publishers, 1991

[7] Th. Martinetz and K. Schulten. *Topology Representing Networks.* Neural Networks 7(3):507-522, 1994

[8] T. Hastie and W. Stuetzle. *Principal Curves.* Journal of the American Statistical Assocoation, 84(406):502-516, 1989

[9] H. Ritter, Th. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps.* Addison-Wesley, 1991