

A Comprehensive Approach to Two-Dimensional Alignment

Wim PH de Boer

ARTICLE INFO

Received Date: October 30, 2019

Accepted Date: December 06, 2019

Published Date: December 10, 2019

KEYWORDS

Curfit2D

Two-dimensional (2D) alignment

Warp function

Copyright: © 2019 Wim PH de Boer. Chromatography And Separation Techniques Journal. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation for this article: Wim PH de Boer. A Comprehensive Approach to Two-Dimensional Alignment. Chromatography And Separation Techniques Journal. 2019; 2(1):121

ABSTRACT

The two-dimensional alignment of chromatograms remains the subject of ongoing research. The shift of a Sample chromatogram compared to a Reference chromatogram is a challenge to calculate, particularly when the interest lies in smaller signal peaks. Previous attempts include a variety of different methods: peak selection approaches, two-dimensional correlation and the Smith-Waterman method taken from genetics. This manuscript shows the comprehensive approach of the theory that underlies computer program Curfit2D. In the first part, a newly developed approach to the underlying theory is presented. In the second part, the challenges of the implementation of the theory are discussed.

INTRODUCTION

First to decide is how to compare two chromatograms: the metric by which the shifts are measured directly sets the stage of how the theory evolves [1-5]. A natural measure of nearness is the familiar Lebesgue L2 integral of squared differences:

$$\iint [f(x, y) - g(x, y)]^2 dx dy,$$

where the Reference chromatogram $f(x, y)$ and the shifted Sample chromatogram $g(x, y)$ are compared on a shared two-dimensional retention time domain S . To model the shifts of the Sample chromatogram we introduce a continuous warp function $w(x, y)$ such that the measure of nearness becomes:

$$\iint [f(x, y) - g(w(x, y))]^2 dx dy,$$

which is to be minimized over the retention time domain S . Next, the warp function is taken to be a Cartesian product of two univariate spline functions $u(x)$ on the x -axis and $v(y)$ on the y -axis, respectively:

$$w(x, y) = (u \times v)(x, y) = (u(x), v(y)), \text{ with}$$

$$u(x) = x + \sum_{k=1}^K a_k B_k(x), \text{ and } v(y) = y + \sum_{l=1}^L b_l B_l(y), \text{ where the } B\text{'s stand for one-}$$

order or third-order splines [6]. Shifts are therefore modelled as corrections on linear time, and the warp function as a two-dimensional spline correction on two-dimensional retention time domain S :

$$w(x, y) = xy + \sum \sum [a_k b_l B_k(x) B_l(y)],$$

where the sum over the spline coefficients a_k runs from 1 to K and the sum over the spline coefficients b_l from 1 to L . It is worth pointing out that a spline is a local function on the time axis, and the associated spline coefficient determines its height. By

Corresponding author:

Wim PH de Boer,

Tel: +31 70 3996929;

Email: wdeboer@ziggo.nl

constructing a base of splines across the time axis interval, shifts across that interval are modelled as “waves” riding on top of linear time. Choosing the right number of splines K and L remains a topic for further study and will depend on the complexity of the shifts concerned.

The numerical problem to solve is then to minimize function

$$F(\{a, b\}) \equiv \iint [f(x, y) - g(w(x, y; \{a, b\}))]^2 dx dy,$$

as a function of spline coefficients: $a(1, \dots, K)$ and $b(1, \dots, L)$.

IMPLEMENTATION – AN ITERATIVE APPROACH

How to estimate the spline coefficients that minimize function F . An approach could be to devise an iterative approach by decomposing spline coefficients $Z = (a_1, \dots, a_K; b_1, \dots, b_L)$ into known and unknown parts $Z = Z + \Delta Z$ and start the iteration with $Z = C$; C could be the global shift of the Sample chromatogram with respect of the Reference chromatogram. Then take the first- and second-order derivatives of function F with respect of each of the spline coefficients and summarize the findings in “quadratic model” form:

$$G^T \Delta Z + \frac{1}{2} \Delta Z^T H \Delta Z,$$

with vector G and matrix H in terms of known quantities Z . One obtains then for vector G :

$$G_k = \partial F(\{a, b\}) / \partial a_k = 2 \iint \{(g(w(x, y)) - f(x, y)) (\partial g / \partial u) B_k(x)\} dx dy, \text{ for } k = 1, \dots, K$$

$$G_l = \partial F(\{a, b\}) / \partial b_l = 2 \iint \{(g(w(x, y)) - f(x, y)) (\partial g / \partial v) B_l(y)\} dx dy, \text{ for } l = 1, \dots, L.$$

Note that the derivatives of chromatogram $g(w(x, y))$ are with respect to the univariant spline functions u and v . Likewise matrix H can be obtained:

$$H_{kl} =$$

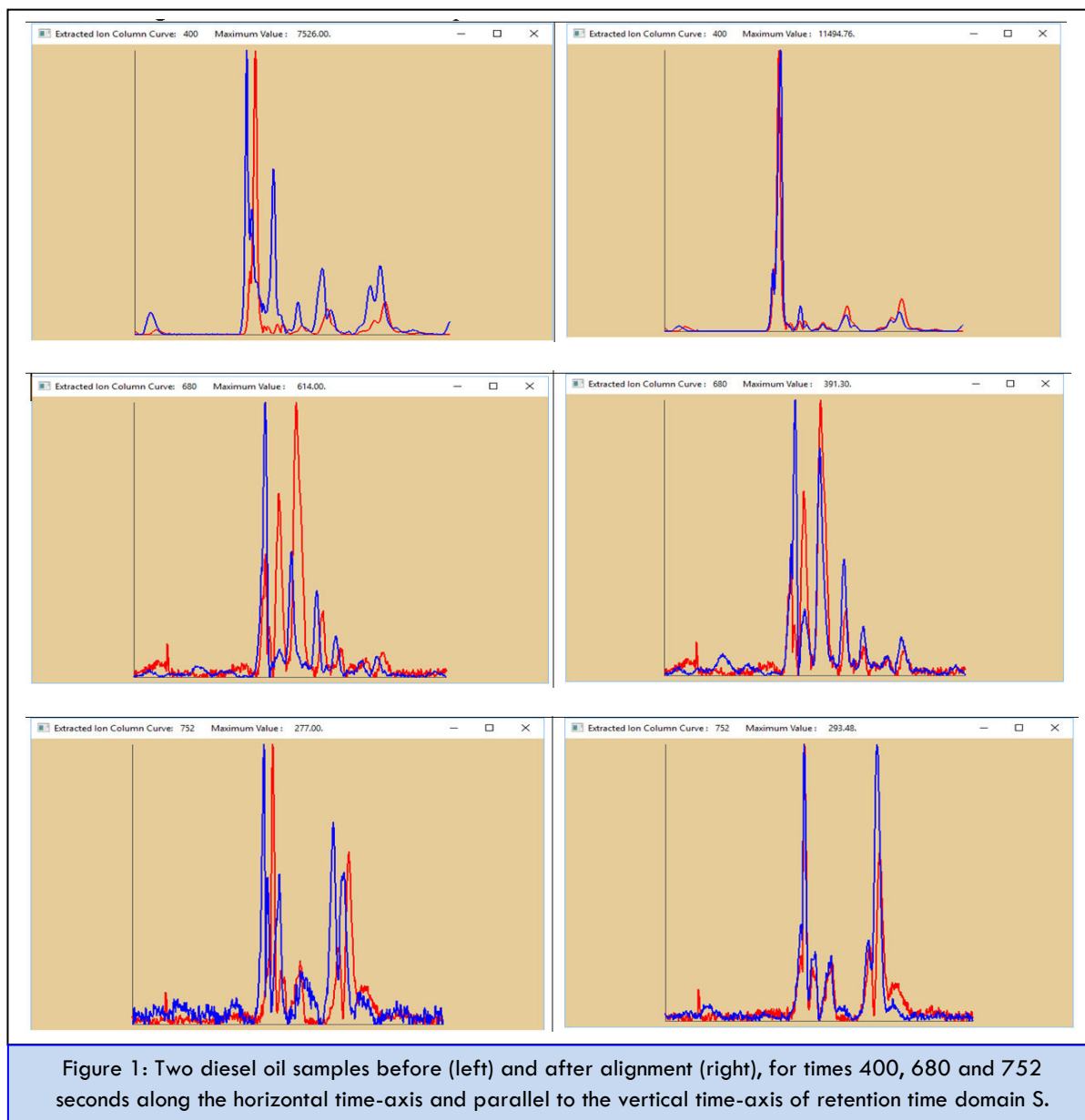
$\partial^2 F(\{a, b\}) / \partial a_k \partial a_l =$ $\iint \{(\partial g / \partial u)^2 B_k(x) B_l(x)\} dx dy +$ higher derivatives of g , for: $k, l = 1, \dots, K$.	$\partial^2 F(\{a, b\}) / \partial b_l \partial a_k =$ $\iint \{(\partial g / \partial v) (\partial g / \partial u) B_l(y) B_k(x)\} dx dy +$ higher derivatives of g , for: $l = 1, \dots, L; k = 1, \dots, K$.
$\partial^2 F(\{a, b\}) / \partial a_k \partial b_l =$ $\iint \{(\partial g / \partial u) (\partial g / \partial v) B_k(x) B_l(y)\} dx dy +$ higher derivatives of g , for: $k = 1, \dots, K; l = 1, \dots, L$.	$\partial^2 F(\{a, b\}) / \partial b_k \partial b_l =$ $\iint \{(\partial g / \partial v)^2 B_k(y) B_l(y)\} dx dy +$ higher derivatives of g , for: $k, l = 1, \dots, L$.

Tests show that the off-diagonal components of matrix H are important to obtain steady convergent iteration results. An example of the alignment achieved by this iterative approach is given in (Figure 1) for a set of diesel oil samples that proved to be difficult to align. The figure shows the quality of alignment for three arbitrary retention time points on the horizontal retention time-axis and parallel to the vertical retention time-axis of retention time domain S . Note that peaks after alignment appear to be higher than before alignment, but this makes sense because shifts occur in two dimensions and the peaks before alignment appear to be “out-of-focus”.

OUTLOOK

For a group of diesel oil files [5] the number of iterations run from 3 to close to the maximum of 50 iterations allowed in one run. To continue the alignment one has the option to start a new run with the outcome chromatogram of the previous run as input chromatogram, but run time depends, in order of importance, on the size of retention time domain S , the number of iteration loops needed to achieve convergence, and the number of splines chosen. Visualization of the development of the spline coefficients during the iteration process reveals that a fair number of coefficients reach a stable level soon in the iteration, while other coefficients keep on developing. This suggests a more dynamic minimalization approach in which the development of the spline coefficients could be followed and fixed to the limiting value that they reach. Visualization of the Chromatograms before alignment, combined with the knowledge that splines are locally defined, may help to decide which part of the Sample chromatogram needs most alignment and which spline coefficients are likely to be dominant.

Could it make sense to look for another faster minimalization approach? Other minimalization routines such as algorithm “TOLMIN” [7] from M.J.D. Powell may be worth looking at. Algorithm TOLMIN calculates the least value of a differential function of several variables subject to linear constraints on the values of the variables, using a method that is tolerant in handling those constraints while searching through variable space. Program Curfit2D uses linear constraints on the spline coefficients to ensure that retention time does not reverse because of the shift corrections. This makes the TOLMIN algorithm an interesting candidate to consider.



REFERENCES

1. Gros J, Nabi D, Dimitriou-Christidis P, Rutler R, Arey JS. (2012). Robust Algorithm for Aligning Two-Dimensional Chromatograms. *Anal Chem.* 84: 9033-9040.
2. Rempe D, Reichenbach S, Scott S, Spring (2016). Research Fair poster. University of Nebraska–Lincoln.
3. Suits F, Lepre J, Du P, Bischoff R, Horvatovich P. (2008). Two-Dimensional Method for Time Aligning Liquid Chromatography–Mass Spectrometry Data. *Anal Chem.* 80: 3095-3104.
4. Kim S, Koo I, Fang A, Zhang X. (2011). Smith-Waterman peak alignment for comprehensive two-dimensional gas chromatography-mass spectrometry. *BMC Bioinformatics.* 12: 235.
5. De Boer WPH, Lankelma J. (2014). Two-dimensional semi-parametric alignment of chromatograms. *Chromatography A.* 1345: 193-199.
6. Carl R de Boor. (1978). A practical guide to Splines. *Applied Math. Sciences* 27.
7. Powell MJD. (1989). A tolerant algorithm for linearly constrained optimization calculations. *Math Programming.* 45: 547-566.