Journal of Chromatography A, xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Orthogonality measurements for multidimensional chromatography in three and higher dimensional separations

Mark R. Schure[a,*], Joe M. Davis[b,**]

[a] Theoretical Separation Science Laboratory, Kroungold Analytical, Inc., 1299 Butler Pike, Blue Bell, PA 19422 USA
[b] Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, Carbondale, IL 62901-4409 USA

ARTICLE INFO

Article history:
Received 27 April 2017
Received in revised form 11 June 2017
Accepted 12 June 2017
Available online xxx

Keywords:
Orthogonality
Orthogonality metric
Multidimensional chromatography
Three-dimensional chromatography
Information theory

ABSTRACT

Orthogonality metrics (OMs) for three and higher dimensional separations are proposed as extensions of previously developed OMs, which were used to evaluate the zone utilization of two-dimensional (2D) separations. These OMs include correlation coefficients, dimensionality, information theory metrics and convex-hull metrics. In a number of these cases, lower dimensional subspace metrics exist and can be readily calculated. The metrics are used to interpret previously generated experimental data. The experimental datasets are derived from Gilar's peptide data, now modified to be three dimensional (3D), and a comprehensive 3D chromatogram from Moore and Jorgenson.

The Moore and Jorgenson chromatogram, which has 25 identifiable 3D volume elements or peaks, displayed good orthogonality values over all dimensions. However, OMs based on discretization of the 3D space changed substantially with changes in binning parameters. This example highlights the importance in higher dimensions of having an abundant number of retention times as data points, especially for methods that use discretization. The Gilar data, which in a previous study produced 21 2D datasets by the pairing of 7 one-dimensional separations, was reinterpreted to produce 35 3D datasets. These datasets show a number of interesting properties, one of which is that geometric and harmonic means of lower dimensional subspace (i.e., 2D) OMs correlate well with the higher dimensional (i.e., 3D) OMs. The space utilization of the Gilar 3D datasets was ranked using OMs, with the retention times of the datasets having the largest and smallest OMs presented as graphs. A discussion concerning the orthogonality of higher dimensional techniques is given with emphasis on molecular diversity in chromatographic separations.

In the information theory work, an inconsistency is found in previous studies of orthogonality using the 2D metric often identified as $\%O$. A new choice of metric is proposed, extended to higher dimensions, characterized by mixes of ordered and random retention times, and applied to the experimental datasets. In 2D, the new metric always equals or exceeds the original one. However, results from both the original and new methods are given.

1. Introduction

The techniques known collectively as multidimensional liquid chromatography are practiced typically with two columns, whereby the effluent from the first column is provided to the second column for further separation based on molecular structure. These are referred to as two-dimensional (2D) separations [1]. While 2D gas chromatography has been widely available for a number of years, instrumentation for 2D liquid chromatography (2DLC) is now becoming widely available, and the ability to resolve compounds with similar molecular structure has been invaluable in studies of complex materials, most notably biomolecules.

Three-dimensional (3D) and higher dimensional separations have been discussed for years, but only one report exists on comprehensive 3DLC instrumentation able to use three different columns [2]. However, with increasing needs of speed and resolution for demanding biomolecule characterization, more attention is being given to using n chromatographic dimensions and a number of mass spectrometry (MS) dimensions, which might include ion mobility mass spectrometry, MS/MS or a combination of other techniques.

Orthogonality metrics (OMs) have been used to measure space occupancy and the uniformity of the spreading of components within these spaces [3–9]. This has become important in one-

* Corresponding author.
** Corresponding author. Current address: 733 Schloss Street, Wrightsville Beach, NC 28480 USA.
E-mail addresses: Mark.Schure@GMail.com (M.R. Schure), Chimicajmd@gmail.com (J.M. Davis).

http://dx.doi.org/10.1016/j.chroma.2017.06.036
0021-9673/

dimensional (1D) separations in determining and optimizing the resolution of specific components [6]. In 2D chromatography, OMs [3–5,7–9] measure the utilization of the separation space with occupancy metrics, peak spacing metrics and uniformity metrics, among others. These metrics can be utilized to measure the effectiveness of the separation and to optimize the separation.

Higher dimensional separations are indeed rare. One of these, published by Moore and Jorgenson [2], is the separation of peptides using 3D chromatography. In this experiment, the first dimension column is a size exclusion chromatography (SEC) column, followed by reversed-phase liquid chromatography (RPLC) column, followed by a very fast optically-gated capillary zone electrophoresis (CZE) experiment. As is a rather common practice, the columns and techniques were chosen to put the slower separation in the first dimension followed by faster and faster steps in succeeding dimensions. This is due to the sampling requirements [1,10,11], which dictate that the slower separations are typically put in the first dimension and the fastest steps are performed in the last dimension.

In this paper, we explore some of the aspects of 3D and higher dimensional OMs. The OMs described here are extensions of known 2D OMs and include correlation coefficients (CCs), dimensionality, information theory metrics and convex-hull metrics. For the dimensionality and information theory metrics, which employ discretization of retention times into bins, recommendations are given to the number of bins and the implications of binning in higher dimensions.

Two applications of 3D OMs to experimental systems are discussed. First, a 3D dataset is reconstructed from the published 3D comprehensive separation of Moore and Jorgenson [2]. Second, we utilize the Gilar peptide data [7], which in a previous study [9] produced 21 unique 2D datasets by the pairing of 7 1D separations. This data now is reinterpreted to produce 35 unique 3D datasets. In all cases, datasets are m x 3 matrices, where the ith column vector contains m retention times for the ith dimension. The evaluation of these datasets using 3D OMs produces many interesting insights that are discussed in detail and point out many aspects and difficulties in analyzing and optimizing 3D separations.

With the exception of CCs, a global OM in 3D is available, along with lower dimensional subspace OMs. For 3D systems, there are 3 2D subspace metrics based on paired dimensions. They will be shown to be quite useful. This scheme can be extended to nD (n–dimensional) systems, such that for $n \geq 3$ all of the lower dimensional subspace OMs exist.

This work uncovers a problem with the previously derived 2D OM from information theory known as %O [8]. We will show that this OM is not symmetrical with respect to the input data, and we will derive a more universal equation able to be employed in any number of dimensions. The analysis of the datasets shows the 2D OMs from the original and new methods are nearly equal, but we prove the new method always gives the same or a greater OM than the original one.

2. Mathematical development

2.1. Notation and terminology

The letter "D" in 1D, 2D, 3D, nD, etc., henceforth means both "dimension" and "dimensional", depending on context. Its dual interpretation provides simplicity. The symbol D is the traditional symbol for the OM, dimensionality; it differs from these abbreviations in the absence of a number preceding it.

Various notations for column vectors of retention times are used, depending on circumstances. Vector groupings are referred to as pairs in 2D and triplets in 3D. In 2D, the individual vectors are labeled X (first dimension) and Y (second dimension); in 3D, they are usually X, Y, and Z (third dimension). In nD, they are X_1, X_2,. . .X_n. Depending on circumstances, pairs and triplets are represented by letter or number combinations, e.g., XZ, XYZ, and 246 (the numbers are explained later). Individual coordinates in any vector are represented by lower case letters, e.g., x, y, and z, and represent retention times of peaks. The notation style used here for the most part is found in information theory texts, for example, in the well-known book of Cover and Thomas [12].

Generic symbols for OMs have no subscripts, e.g., dimensionality D. Subscripts are used to identify specific types, e.g., D_{2D} for 2D dimensionality.

The information theory OM has the traditional name, orthogonality. The word "orthogonality" also is used in its general sense to indicate a good spreading of zones over a separation. We have tried to avoid confusion between the two meanings through context.

Zones in the SEC and CZE dimensions of the Moore and Jorgenson separation have elution times, not retention times, because retentive chromatography was not used. However, for consistency we describe all such times as retention times and describe this 3D separation as a 3D chromatogram. We also use the older abbreviation, CZE, to conform to the Moore and Jorgenson chromatogram reproduced here.

2.2. General properties of OMs

To be useful, OMs must have some essential properties. The first is the scale. The scale should be defined between 0 and 1, or some other values that convey characteristics of zone ordering or zone coverage. The CC analyses scale from 0 to 1 after a suitable normalization. The information theory results are typically scaled between 0 and 100%, although the 0–1 scale is more natural, and we will use that scaling here. The fractal dimension scale is between 0 and D_C, where D_C is the number of chromatographic instrumental dimensions, i.e., the number of columns that are used for successive fractionation. The convex hull relative area, volume, or hypervolume in any number of dimensions varies between 0 and 1. These scalings, with the exception of the fractal dimension, can be applied so that an OM can be suitably normalized between 0 and 1 no matter how many dimensions are utilized as the separation space.

Another important property is that OMs should give the same answer upon swapping the column data vectors. In other words, the OMs should preserve the data symmetry and give the same answer for any order of the vectors. This is not to say that the order in which the different dimensions of a multidimensional separation are developed is not important; obviously it is. However, the vector order does not affect the intervals between these times, and therefore should not affect the OM. For example, in 2D, OM(X,Y) should equal OM(Y,X), where OM is the specific orthogonality metric calculation. In 3D, it should be true that OM(X, Y, Z) = OM(X, Z, Y) = OM(Y, X, Z) = OM(Y, Z, X) = OM(Z, X, Y) = OM(Z, Y, X).

A few of these OMs do not have this property of data symmetry. For example, the Pearson and Kendall CCs have this property while the Spearman CC does not, and this suggests that the Spearman CC is not suitable for use as an OM. Furthermore, the %O OM [8] from information theory [12] does not have this property. However, on further inspection, a similar OM can be derived so that this symmetry property is preserved. We derive such a metric below.

OMs can be classified as non-discretized and discretized [9]. The non-discretized metrics do not require the binning or scaling of retention times, although scaling does not change the OM value. Discretized OMs require both binning and scaling.

The details of binning are introduced as needed. For consistency, all retention times are scaled here over the range as 0–1 using the maximum (max) and minimum (min) times of each dimension. For

Please cite this article in press as: M.R. Schure, J.M. Davis, Orthogonality measurements for multidimensional chromatography in three and higher dimensional separations, J. Chromatogr. A (2017), http://dx.doi.org/10.1016/j.chroma.2017.06.036