Journal of Chromatography A, 1507 (2017) 1–10

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Assisted baseline subtraction in complex chromatograms using the BEADS algorithm

J.A. Navarro-Huerta[a], J.R. Torres-Lapasió[a,*], S. López-Ureña[b], M.C. García-Alvarez-Coque[a]

[a] Department of Analytical Chemistry, Faculty of Chemistry, Universitat de València, c/ Dr. Moliner 50, 46100 Burjassot, Spain
[b] Department of Mathematics, Faculty of Mathematics, Universitat de València, c/ Dr. Moliner 50, 46100 Burjassot, Spain

ARTICLE INFO

Article history:
Received 21 March 2017
Received in revised form 24 May 2017
Accepted 26 May 2017
Available online 29 May 2017

Keywords:
HPLC
Complex samples
Baseline subtraction
BEADS
Working parameters
Autocorrelation

ABSTRACT

The data processing step of complex signals in high-performance liquid chromatography may constitute a bottleneck to obtain significant information from chromatograms. Data pre-processing should be preferably done with little (or no) user supervision, for a maximal benefit and highest speed. In this work, a tool for the configuration of a state-of-the-art baseline subtraction algorithm, called BEADS (Baseline Estimation And Denoising using Sparsity) is developed and verified. A quality criterion based on the measurement of the autocorrelation level was designed to select the most suitable working parameters to obtain the best baseline. The use of a log transformation of the signal attenuated artifacts associated to a large disparity in signal size between sample constituents. Conventional BEADS makes use of trial-and-error strategies to set up the working parameters, which makes the process slow and inconsistent. This constitutes a major drawback in its successful application. In contrast, the assisted BEADS simplifies the setup, shortens the processing time and makes the baseline subtraction more reliable. The assisted algorithm was tested on several complex chromatograms corresponding to extracts of medicinal herbs analysed with acetonitrile-water gradients, and a mixture of sulphonamides eluted with acetonitrile gradients in the presence of the non-ionic surfactant Brij-35 under micellar conditions.

1. Introduction

Modern high-performance liquid chromatography (HPLC) instruments are able to provide highly complex signals in routine analysis, from which the relevant information should be extracted [1]. In these analyses, the data processing step constitutes a bottleneck, constraining sample throughput [2,3]. Problems such as noisy signals, coeluting peaks (sometimes highly overlapped), peak shifts and the presence of irregular baselines should be addressed. The operations to handle these problems should be done preferably with little (or no) user supervision for a maximal benefit and highest speed.

The aim of this work is to improve the baseline subtraction in chromatograms of high complexity, with complete drift suppression and little analyst supervision. Very recently, a new algorithm called "Baseline Estimation And Denoising using Sparsity" (BEADS) was proposed [4,5], which presents as novelty the capability of performing a full decomposition of chromatograms in net signal (i.e., the pure signals of the analytes and their accompanying compounds), baseline and noise. The baseline is modelled as a low frequency signal and the noise as a high frequency contribution, while the peaks of analytes are described as sparse signals, whose first and second derivatives are also sparse (a vector signal is classified as "sparse" when most of its elements are zero). For this purpose, BEADS requires that the user specify several parameters to ensure that the recovered signals have chemical meaning (e.g., positive signals for all analytes). It should be noted that most baseline subtraction algorithms also require some user inputs. This is the case of the mixture models based on splines proposed by Rooi and Eilers [6], the adaptive iteratively reweighted penalised least squares (airPLS) [7], and the backcor algorithm [8].

The authors of BEADS validated it in comparison with the airPLS and backcor algorithms [4]. The three methods yielded reasonable estimates of the baselines, but BEADS offered the best performance. Indeed, in our trials with a variety of chromatograms, BEADS was verified to provide excellent results in complex situations. However, we found some issues that make its routine application to real samples difficult, which should be addressed.

* Corresponding author.
E-mail address: jrtorres@uv.es (J.R. Torres-Lapasió).

http://dx.doi.org/10.1016/j.chroma.2017.05.057
0021-9673/

The triple decomposition of chromatograms in BEADS is done essentially by using highly efficient frequency filters, which makes the outline easier and the calculation faster. Moreover, the algorithmic framework is based on majorization-minimization [4], which converges quickly regardless of the set of values used in its initialization. The result of the combination of these techniques is a highly efficient algorithm that saves memory. Another advantage is that, in contrast to other baseline algorithms [8], the set of baselines obtained by BEADS is not described as a parametric family of functions. This feature confers BEADS an extreme flexibility to accommodate any baseline, whatever its complexity.

The limitations of BEADS can be classified in two categories. First, it requires a careful adjustment of the working parameters to properly process real signals of different origin. This operation may be difficult for highly complex signals, owing to the instability of the adjustment process (i.e., small changes in the parameters may lead to very different baselines). Secondly, chromatograms must fulfil some conditions (described in detail in Section 3.1), mandatory for the application of BEADS, but hardly fulfilled in practice with real chromatograms.

In this work, we analyse comprehensively the limitations of BEADS, and propose some solutions, which improve the results and reliability of this algorithm and contribute to make it more robust, faster and easier to apply to chromatograms of real highly complex samples of different origin, with little supervision.

2. Experimental

2.1. Reagents

In order to explore the correct subtraction of the baseline, several fingerprints of medicinal herbs were processed, corresponding to extracts in hot water of horsetail and decaffeinated teas obtained in the laboratory. For the chromatographic analysis, hydro-organic gradients were prepared with acetonitrile (Scharlab, HPLC grade, Barcelona, Spain) and water. This was buffered at pH 3 with 0.01 M sodium dihydrogen phosphate (Sigma, Roedermark, Germany) and a suitable amount of 0.01 M HCl (Scharlab). The chromatographic signals of extracts of red peony root, taken from Ref. [7], were also processed.

The influence of negative peaks associated with refractometric void volume signals was studied using chromatograms for a mixture of 15 sulphonamides: sulphaguanidine, sulphanilamide, sulphacetamide, sulphadiazine, sulphathiazole, sulphapyridine, sulphamera-zine, sulphamethazine, sulphamethizole, sulphamonomethoxine, sulphachloropyridazine, sulphamethoxazole, sulphisoxazole, sulphadimethoxine and sulphaquinoxaline, eluted with an acetonitrile gradient in the presence of Brij-35 (Sigma, St. Louis, MO, USA), buffered at pH 3 with 0.01 M sodium dihydrogen phosphate. All solutions were filtered through 0.45 μm Nylon membranes from Micron Separations (Westboro, MA, USA), before their injection into the chromatographic system.

2.2. Preparation of extracts of medicinal herbs

The extracts of horsetail and decaffeinated teas were obtained following the recommendations of Dumarey et al. [9]. For this purpose, 20 ml of nanopure water was added to 0.2 g of ground sample, and boiled in the absence of light. The extracts were filtered through 0.2 μm membrane filters from Pall Gelman Laboratory (Karlstein/Main, Germany), to finally fill 2 ml vials for chromatographic analysis.

2.3. Apparatus, columns and software

An Agilent modular instrument (HP 1100, Waldbronn, Germany) was used, consisting of quaternary pump, automatic injector, temperature controller, and variable wavelength UV-visible detector. The chromatograms of the medicinal herbs and mixtures of sulphonamides were detected at 210 and 254 nm, respectively. The column temperature was fixed at 25 °C. The injection volume was 10 μl, and the flow rate was kept constant at 1 ml/min, in all instances.

An OpenLAB CDS LC ChemStation (Agilent, B.04.03 revision) was used for the acquisition of chromatographic signals. Raw chromatograms were processed without any correction by the ChemStation software, unless those associated to the default working parameters, such as autobalance in the pre-run, 5% zero offset, or attenuation to 1000 mAU. Matlab 2016b (The MathWorks Inc., Natick, MA, USA) was applied for data treatment. The Matlab function [5] (which is included in the Supplementary material of Ref. [4]) was used for the conventional application of BEADS.

3. Results and discussion

3.1. Limitations of BEADS

As indicated, BEADS makes the simultaneous decomposition of a signal **y** in three contributions:

$$\mathbf{y} = [y_1, y_2, .., y_n] = \mathbf{c} + \mathbf{b} + \mathbf{e} \quad (1)$$

where **c**, **b** and **e** make reference to the sparse chromatogram, baseline and noise vectors computed by BEADs, which depend on a set of working parameters **p**. The working parameters are the cutoff frequency (f_c, which constitutes the boundary between the baseline and the rest of contributions), asymmetry (r, which penalizes the negative values) and regularization parameters (λ_0, λ_1 and λ_2, which control the sparsity of vector **c**). An additional parameter is the amplitude (A), which multiplies the regularization parameters; thus, the regularization parameters are actually $A \times \lambda_i$, which makes the ratios among the λ_i parameters independent of their magnitude.

The adaptability of BEADS to real baselines is noteworthy, but its application has the following limitations, especially severe for complex chromatograms:

(i) Requirement of the same signal intensity for the first and last points in the chromatogram (i.e., periodicity of the signal).
(ii) Abnormal risings of the baseline under major signals in chromatograms where the analytes exhibit extreme variations in signal size. The overall appearance of the computed baseline is wavy (see figures discussed in Section 3.2), instead of having a smooth trend at large scale.
(iii) Problematic processing of chromatograms containing sporadic negative peaks, such as those corresponding to refractometric signals, or those observed in chromatograms obtained using indirect UV–visible detection. This forces a careful adjustment of the working parameters for each sample.
(iv) Dependence among the working parameters. The baseline is particularly susceptible to the selected cutoff frequency at low frequencies, which results in an unstable adjustment process. This situation is worsened by the wide range of values to be explored, which in some cases comprises several orders of magnitude (a typical chromatogram composed of 10,000 points can involve exploring cutoff frequencies over 4 orders of magnitude).
(v) Need for each chromatogram of a particular adaptation of the working parameters (i.e., each set of parameters is translated