Progress in Nuclear Magnetic Resonance Spectroscopy 59 (2011) 271–292

Contents lists available at ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

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

Analysis of non-uniformly sampled spectra with multi-dimensional decomposition

Vladislav Yu. Orekhov [a,*], Victor A. Jaravine [b]

[a] Swedish NMR Centre, University of Gothenburg, Box 465, 40530 Gothenburg, Sweden
[b] Center for Biomolecular Magnetic Resonance, Institute of Biophysical Chemistry, Goethe-University, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 4 January 2011
Accepted 21 February 2011
Available online 24 February 2011

Keywords:
PARAFAC, MDD, TWD
Non-linear sampling
Hyper-dimensional spectroscopy
Targeted acquisition

Contents

Abbreviations: 3DD, 4DD, three or four-dimensional versions of MDD, respectively; Co-MDD, co-processing of several spectra by MDD; DFT, discrete Fourier transform; E.COSY, exclusive correlation spectroscopy; FID, free-induction decay, typically refers to 1D spectrum in time domain; GFT, G-matrix Fourier transform spectroscopy; ICA, independent component analysis; HD-MDD, hyper-dimensional MDD, synonym of Co-MDD; HD, hyper-dimensional (spectrum); HMQC, heteronuclear multiple quantum coherence (experiment); HNCO, A 3D experiment for correlating sequential protein backbone amide group (^{15}N–^{1}H) of one residue with the carbonyl (^{13}CO) of the preceding residue; HSQC, heteronuclear single quantum coherence; MDD, multi-dimensional decomposition; MSG, malate synthase G; NUS, non-uniform sampling; NOESY, nuclear Overhauser effect spectroscopy; NOE, nuclear Overhauser effect; PARAFAC, parallel factor analysis; PCA, principal component analysis; R-MDD, recursive multi-dimensional decomposition; TA, targeted acquisition; TROSY, transverse relaxation optimized spectroscopy; SVD, singular value decomposition; USF3, unified spectra format for compressed spectra representations; VDAC, voltage-dependent anion-selective channel protein.

* Corresponding author. Tel.: +46 31 7863886; fax: +46 31 7863880.
E-mail address: orov@nmr.gu.se (V.Yu. Orekhov).

0079-6565/$ - see front matter
doi:10.1016/j.pnmrs.2011.02.002

1. Introduction

Multidimensional NMR spectroscopy has been established as an indispensible tool for studying structure, dynamics, and interactions of biopolymers. When used in the frame of a contemporary structural molecular biology project with streamlined efficient target selection and protein expression, NMR often represents a major time bottleneck [9]. In order to determine spatial structure of a protein, one needs to spend weeks on data collection followed by data analysis, which is at least as lengthy and is usually performed manually. The NMR community has devoted significant attention to the need to save spectrometer time and to automate the analysis steps [10–12]. This review is devoted to one of the approaches developed to address these challenges and thus to increase the effectiveness of biomolecular NMR. It gives an account of the work performed in our group over the last several years on non-uniform sampling (NUS) and multi-dimensional decomposition (MDD).

In traditional spectroscopy, which relies on the uniform sampling of the signal in the time domain, high resolution in the indirect spectral dimensions comes at the expense of long measurement time and compromised sensitivity. Novel sparse sampling schemes and corresponding processing methods aim to avoid these problems. The duration of a multidimensional NMR experiment is determined by the time needed for measurement of one data point and the number of these. Both factors are targeted in ongoing efforts to speed up the experiments. Recording individual data points can be accelerated by reducing the delay between consecutive measurements [13] or by gradient encoding of the indirect dimensions in a single scan [14]. Sparse sampling schemes reduce the number of data points without losing essential information. In particular, in *GFT* [15] and *projection reconstruction* [16,17], only specific spectral projections of lower dimensionality are used to obtain information present in complete multidimensional spectra. As an alternative, NUS, which is also referred to as non-linear or sparse sampling, allows reconstruction of a complete spectrum from only a small number of optimally selected experimental data points [18]. By taking into account prior information about signal properties, a NUS schedule can be optimized for maximum spectral sensitivity and resolution [18]. The approach requires nontraditional signal processing schemes such as non-linear Fourier transform [19–21], maximum entropy [22–24], or multi-dimensional decomposition (MDD) [1,5–7]. The later method and its applications is the focus of this review.

2. Theory

2.1. The MDD model

For most multidimensional NMR spectra it is reasonable to assume that a peak in the spectrum is completely described by its one-dimensional line-shapes in all spectral dimensions [25,26]. Thus, the model of multi-dimensional decomposition (MDD) looks for an approximation of a M-dimensional spectral matrix by the sum of a small number of tensor products of one-dimensional vectors:

$$\boldsymbol{S}_{MDD} = \Sigma_{\beta} {}^{\beta}a \, {}^{\beta}F^{1} \otimes \ldots {}^{\beta}F^{M-1} \otimes {}^{\beta}F^{M} \quad (1)$$

where the model spectrum $\boldsymbol{S}_{MDD}$ is the sum of fixed number of components N_c enumerated by index $\beta = 1, \ldots, N_c$. Each component is given by the product of normalized vectors ${}^{\beta}F^{m}$ for every spectral dimension $m = 1, \ldots, M$, referred to below as shapes, and the component amplitude ${}^{\beta}a$. The term shape is introduced here in relation to the spectral line shape; its several synonyms are present in the literature, i.e. loads, modes, factors, etc. Symbol $\otimes$ denotes the outer product operation, which produces M-dimensional matrix from M one-dimensional shapes.

MDD has been used since the early seventies in the variety of fields under various names, such as parallel factor analysis (PARAFAC) [27], canonical decomposition [28], and three-way decomposition [29]. The method was first applied in psychometrics and has been successfully used for signal processing and data analysis in functional magnetic resonance imaging (fMRI) [30,31], electro encephalography (EEG) [2,32–36], and other fields [37,38].

2.2. Components and shapes

A simple approach is to think about a component as the representation of a cross peak in a multi-dimensional spectrum. The shapes then are traditional line-shapes of the peak in all dimensions. The actual situation, however, is more complex, since the components do not always have a one-to-one correspondence to peaks. In general, a peak showing complex structure, e.g. in an E.COSY spectrum, may require several components for its description. It also can be the other way around, as in the NOESY spectrum – one component may accommodate several cross peaks. It is important to emphasize that the MDD model does not make any assumptions about the shape vectors ${}^{\beta}F^{m}$. Thus it can be equally well applied to data in the time or frequency domains, as well as