Three-Way Data Reduction Based on Essential Information

Raffaele Vitale, A. Azizi, M. Ghaffari, N. Omidikia, Cyril Ruckebusch

Journal of Chemometrics

WILEY

Journal of CHEMOMETRICS

RESEARCH ARTICLE OPEN ACCESS

Three-Way Data Reduction Based on Essential Information

Raffaele Vitale[1] | Azar Azizi[2] | Mahdiyeh Ghaffari[3] | Nematollah Omidikia[2] | Cyril Ruckebusch[1]

[1]Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, LASIRE (UMR 8516), CNRS, Univ. Lille, Lille, France | [2]Department of Chemistry, Faculty of Science, University of Sistan and Baluchestan, Zahedan, Iran | [3]Department of Analytical Chemistry, Institute for Molecules and Materials (IMM), Radboud University, Nijmegen, The Netherlands

Correspondence: Raffaele Vitale (raffaele.vitale@univ-lille.fr)

Received: 28 February 2024 | **Revised:** 28 August 2024 | **Accepted:** 29 August 2024

Keywords: convex polytopes | essential columns (ECs) | essential rows (ERs) | essential tubes (ETs) | parallel factor analysis–alternating least squares (PARAFAC-ALS) | three-mode factor analysis

ABSTRACT

In this article, the idea of essential information-based compression is extended to trilinear datasets. This basically boils down to identifying and labelling the essential rows (ERs), columns (ECs) and tubes (ETs) of such three-dimensional datasets that allow by themselves to reconstruct in a linear way the entire space of the original measurements. ERs, ECs and ETs can be determined by exploiting convex geometry computational approaches such as convex hull or convex polytope estimations and can be used to generate a reduced version of the data at hand. These compressed data and their uncompressed counterpart share the same multilinear properties and their factorisation (carried out by means of, for example, parallel factor analysis–alternating least squares [PARAFAC-ALS]) yield, in principle, indistinguishable results. More in detail, an algorithm for the assessment and extraction of the essential information encoded in trilinear data structures is here proposed. Its performance was evaluated in both real-world and simulated scenarios which permitted to highlight the benefits that this novel data reduction strategy can bring in domains like multiway fluorescence spectroscopy and imaging.

1 | Introduction

Trilinear curve resolution resting on methodologies like parallel factor analysis–alternating least squares (PARAFAC-ALS [1–3]) can be extremely demanding from a computational point of view especially when high-dimensional datasets as those resulting from multimodal imaging modalities [4–10] are coped with. In order to overcome this particular limitation and speed up calculations, users and practitioners have lately made significant efforts aimed at two distinctive goals: (i) optimising the decomposition algorithms typically resorted to for the trilinear factorisation of the data arrays to be analysed and (ii) providing compressed versions of these data arrays while retaining the portion of their original information content that is essential for carrying out such a factorisation. On the one hand, algorithm acceleration, although proven effective in a wide variety of application scenarios [11–14], does not permit to directly circumvent the intrinsic burden that even modern computing systems face when it comes to loading and keeping in memory massive amounts of measurement observations. On the other hand, compression approaches can ideally yield data structures significantly reduced in size that might be more easily stored, handled and processed also via more standard commercial devices [15].

The need for three- and, more generally speaking, multimodal data compression prior to trilinear or multilinear factorisations has been extensively discussed in literature. In the domain of chemometrics, several computational strategies relying on the principles of bilinear (e.g., principal component analysis [PCA] [16, 17]) and multilinear (e.g., Tucker1 and Tucker3 [18–23]) decomposition techniques have been proposed over the last decades for this particular purpose [15, 24–26]. However, most, if

Raffaele Vitale and Azar Azizi are equally contributing authors.

not all, these methods share a common fundamental shortcoming: the meaningfulness of the information captured during the compression stage for the sake of the final factorisation process is usually neglected. In this regard, in the context of bilinear curve resolution and unmixing, it has recently been shown that a more sensible compression can be achieved through the identification of the so-called essential information (EI) that the data at hand encode—that is, their essential samples (ESs) and/or their essential variables (EVs) [27–29]. ESs and EVs are the most linearly dissimilar row and column vectors of a generic two-dimensional dataset, enable the lossless reproduction of such a dataset in a convex linear way and (under specific normalisation constraints) can be found laying on the multidimensional convex hulls of its row and column space, respectively. By the assessment and extraction of ESs and EVs, one can attain compressions that preserve only those chunks of the data structures under study characterised by a minimal overlap among the components or factors to be retrieved and that, thus, keep the local rank[1] properties of these data structures unchanged, a key aspect when dealing with resolution or unmixing problems. EI-based data pruning has, indeed, been satisfactorily exploited for challenging analytical applications—for example, minor species detection and recovery—mainly in the field of hyperspectral imaging [31–39]. Given, therefore, the considerable benefits that can inherently result from a similar approach, the main aim of this article is to extend the idea of EI selection for the rational reduction of three-way datasets and their faster trilinear factorisation. In this respect, it is worth noticing that any preliminary compression operation conducted on the investigated recordings should always ensure that (i) their local rank properties are rigorously preserved since they can be as crucial as when bilinear decompositions are to be performed [40] and (ii) all the sources of variability underlying them are adequately represented in their reduced counterparts. This cannot be guaranteed by random sampling methods [41–43] such as Monte Carlo experiments [44] and CUR matrix decompositions [45], but strictly requires the use of targeted procedures like the one described before.

2 | Methods

2.1 | Essential Information-Based Reduction of Three-Way Data

Let $\underline{\mathbf{D}}$ be a generic three-way data array (tensor) of dimensions N rows $\times$ J columns $\times$ K tubes and effective rank F (i.e., F is the actual amount of components/factors underlying $\underline{\mathbf{D}}$). The EI-based compression of $\underline{\mathbf{D}}$ basically boils down to recognising and isolating the rows, columns, and tubes of this tensor that allow by themselves to reconstruct the entire space of the original measurements. The number of these so-called essential rows (ERs), essential columns (ECs) and essential tubes (ETs) is obviously lower compared to N, J and K, respectively, but the reduced version of $\underline{\mathbf{D}}$—say $\underline{\mathbf{D}}_r$ of size $N_r \times J_r \times K_r$—shares with it the same trilinear properties. Subsequently, the trilinear factorisations of $\underline{\mathbf{D}}$ and $\underline{\mathbf{D}}_r$ (carried out by means of, e.g., PARAFAC-ALS) yield, in principle, indistinguishable results from the perspective of self-modelling curve resolution. Operationally speaking, the ERs, ECs and ETs of $\underline{\mathbf{D}}$ can be retrieved through the following 5-step computational procedure:

1. $\underline{\mathbf{D}}$ is subjected to a higher order singular value decomposition (HOSVD, [18–20, 46]) that can be written as

$$\underline{\mathbf{D}} = \underline{\mathbf{S}} \times_1 \mathcal{U} \times_2 \mathcal{V} \times_3 \mathcal{W} \tag{1}$$

where $\underline{\mathbf{S}}$ is a $(N \times J \times K)$ -dimensional core tensor, $\mathcal{U}$ $(N \times N)$, $\mathcal{V}$ $(J \times J)$ and $\mathcal{W}$ $(K \times K)$ carry the row, column and tube space factors of $\underline{\mathbf{D}}$ and $\times_1$, $\times_2$ and $\times_3$ connote the mode-1, mode-2 and mode-3 product [47], respectively.[2] In other words, HOSVD returns a specific Tucker3 model encompassing all possible extractable components and characterised by fully orthogonal factor profiles for each individual data mode;

2. $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ are then truncated retaining only their first F columns;

3. Row, column and tube space HOSVD scores are calculated as

$$\mathbf{X} = \mathcal{U}^* \boldsymbol{\Sigma}_1 \tag{2}$$

$$\mathbf{Y} = \mathcal{V}^* \boldsymbol{\Sigma}_2 \tag{3}$$

$$\mathbf{Z} = \mathcal{W}^* \boldsymbol{\Sigma}_3 \tag{4}$$

with $\mathcal{U}^*$ $(N \times F)$, $\mathcal{V}^*$ $(J \times F)$ and $\mathcal{W}^*$ $(K \times F)$ deriving from the truncation of $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$ being the $F \times F$ diagonal matrices containing the first F singular values of the two-way arrays resulting from the row-wise, column-wise and tube-wise unfolding of $\underline{\mathbf{D}}$;

4. $\mathbf{X}$ $(N \times F)$, $\mathbf{Y}$ $(J \times F)$ and $\mathbf{Z}$ $(K \times F)$ are normalised [48, 49] such that

$$\tilde{\mathbf{X}} = \mathbf{X} \oslash \mathbf{x}_1 \mathbf{1}^T \tag{5}$$

$$\tilde{\mathbf{Y}} = \mathbf{Y} \oslash \mathbf{y}_1 \mathbf{1}^T \tag{6}$$

$$\tilde{\mathbf{Z}} = \mathbf{Z} \oslash \mathbf{z}_1 \mathbf{1}^T \tag{7}$$

where $\mathbf{x}_1$, $\mathbf{y}_1$ and $\mathbf{z}_1$ denote the first column of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, respectively, $\oslash$ represents the element-wise (Hadamard) division operator and $\mathbf{1}$ is a vector of ones of dimensions $F \times 1$. After normalisation, the first column of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$ features only unit entries and is therefore removed before further processing. For this reason, from now on, the number of columns of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$ will be considered equal to $F - 1$;

5. ERs, ECs and ETs are finally obtained by identifying the subsets of rows of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$ (say $\tilde{\mathbf{X}}_r$ - $N_r \times F - 1$; $\tilde{\mathbf{Y}}_r$: $J_r \times F - 1$; and $\tilde{\mathbf{Z}}_r : K_r \times F - 1$) supporting their corresponding $(F - 1)$ -dimensional convex hulls. This implies that all the original row vectors of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$ can be expressed as convex linear combinations of those gathered into $\tilde{\mathbf{X}}_r$, $\tilde{\mathbf{Y}}_r$ and $\tilde{\mathbf{Z}}_r$, respectively:

$$\begin{aligned} &\tilde{\mathbf{x}}_n^T = \sum_{n_r=1}^{N_r} \alpha_{n,n_r} \tilde{\mathbf{x}}_{r_{n_r}}^T \text{ s.t. } \alpha_{n,n_r} \geq 0 \\ &\text{and } \sum_{n_r=1}^{N_r} \alpha_{n,n_r} = 1 \ \forall n = 1, \ \ldots, N \text{ and} \\ &\forall n_r = 1, \ \ldots, N_r \end{aligned} \tag{8}$$

$$\tilde{\mathbf{y}}_j^{\mathrm{T}} = \sum_{j_{\mathrm{r}}=1}^{J_{\mathrm{r}}} \beta_{j,j_{\mathrm{r}}} \tilde{\mathbf{y}}_{\mathrm{r}_{j_{\mathrm{r}}}}^{\mathrm{T}} \text{ s.t. } \beta_{j,j_{\mathrm{r}}} \geq 0$$
$$\text{and } \sum_{j_{\mathrm{r}}=1}^{J_{\mathrm{r}}} \beta_{j,j_{\mathrm{r}}} = 1 \;\; \forall j = 1, \ldots, J \quad (9)$$
$$\text{and } \forall j_{\mathrm{r}} = 1, \ldots, J_{\mathrm{r}}$$

$$\tilde{\mathbf{z}}_k^{\mathrm{T}} = \sum_{k_{\mathrm{r}}=1}^{K_{\mathrm{r}}} \gamma_{k,k_{\mathrm{r}}} \tilde{\mathbf{z}}_{\mathrm{r}_{k_{\mathrm{r}}}}^{\mathrm{T}} \text{ s.t. } \gamma_{k,k_{\mathrm{r}}} \geq 0$$
$$\text{and } \sum_{k_{\mathrm{r}}=1}^{K_{\mathrm{r}}} \gamma_{k,k_{\mathrm{r}}} = 1 \;\; \forall k = 1, \ldots, K \quad (10)$$
$$\text{and } \forall k_{\mathrm{r}} = 1, \ldots, K_{\mathrm{r}}$$

Here, $\tilde{\mathbf{x}}_n^{\mathrm{T}}$, $\tilde{\mathbf{y}}_j^{\mathrm{T}}$ and $\tilde{\mathbf{z}}_k^{\mathrm{T}}$ denote the n-th, the j-th and the k-th rows of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$, $\alpha_{n,n_{\mathrm{r}}}$, $\beta_{j,j_{\mathrm{r}}}$ and $\gamma_{k,k_{\mathrm{r}}}$ are positive scalar coefficients and $\tilde{\mathbf{x}}_{\mathrm{r}_{n_{\mathrm{r}}}}^{\mathrm{T}}$, $\tilde{\mathbf{y}}_{\mathrm{r}_{j_{\mathrm{r}}}}^{\mathrm{T}}$ and $\tilde{\mathbf{z}}_{\mathrm{r}_{k_{\mathrm{r}}}}^{\mathrm{T}}$ represent the n_{r}-th, the j_{r}-th and the k_{r}-th rows of $\tilde{\mathbf{X}}_{\mathrm{r}}$, $\tilde{\mathbf{Y}}_{\mathrm{r}}$ and $\tilde{\mathbf{Z}}_{\mathrm{r}}$.

At this point, $\mathbf{D}_{\mathrm{r}}$ can be generated by reducing $\underline{\mathbf{D}}$ to a tensor consisting only of the N_{r} ERs, the J_{r} ECs and the K_{r} ETs identified as described before.

2.2 | Full- Versus Reduced-Size Data Trilinear Decomposition

The main benefit that the three-way data reduction procedure outlined in the previous section could bring to practitioners and operators regards the fact that the trilinear factorisation of $\underline{\mathbf{D}}_{\mathrm{r}}$ should in essence provide the same outcomes as the trilinear factorisation of $\underline{\mathbf{D}}$, but in a significantly lower amount of time. In order to corroborate this statement, the following study will be conducted: first, several datasets together with their compressed counterparts will be processed by means of PARAFAC-ALS to attain their respective non-negative trilinear decomposition, which can be generically written in the form of a restricted Tucker3 model [50] as

$$\mathbf{D}_1 = \mathbf{A}\mathbf{T}_1(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}} + \mathbf{E} \quad (11)$$

$$\mathbf{D}_{\mathrm{r},1} = \mathbf{A}_{\mathrm{r}}\mathbf{T}_1\left(\mathbf{C}_{\mathrm{r}} \otimes \mathbf{B}_{\mathrm{r}}\right)^{\mathrm{T}} + \mathbf{E}_{\mathrm{r}} \quad (12)$$

with

- $\mathbf{D}_1$ and $\mathbf{D}_{\mathrm{r},1}$ being the $N \times JK$ and $N_{\mathrm{r}} \times J_{\mathrm{r}}K_{\mathrm{r}}$ data matrices yielded by the row-wise unfolding of $\underline{\mathbf{D}}$ and $\underline{\mathbf{D}}_{\mathrm{r}}$;
- $\mathbf{T}_1$ ($F \times F^2$) being the data matrix yielded by the row-wise unfolding of the PARAFAC-ALS core, say $\underline{\mathbf{T}}$ ($F \times F \times F$)—notice that $\underline{\mathbf{T}}$ is a unit superdiagonal core, that is, a three-dimensional array of zeros with superdiagonal elements equal to 1;
- $\mathbf{A}$ ($N \times F$), $\mathbf{B}$ ($J \times F$) and $\mathbf{C}$ ($K \times F$) being the full-size row, column and tube space PARAFAC-ALS loadings estimated for $\underline{\mathbf{D}}$;
- $\mathbf{A}_{\mathrm{r}}$ ($N_{\mathrm{r}} \times F$), $\mathbf{B}_{\mathrm{r}}$ ($J_{\mathrm{r}} \times F$) and $\mathbf{C}_{\mathrm{r}}$ ($K_{\mathrm{r}} \times F$) being the reduced-size row, column and tube space PARAFAC-ALS loadings estimated for $\underline{\mathbf{D}}_{\mathrm{r}}$;
- $\mathbf{E}$ ($N \times JK$) and $\mathbf{E}_{\mathrm{r}}$ ($N_{\mathrm{r}} \times J_{\mathrm{r}}K_{\mathrm{r}}$) encoding the residual unexplained variation of $\underline{\mathbf{D}}$ and $\underline{\mathbf{D}}_{\mathrm{r}}$; and
- $\otimes$ denoting the Kronecker product [51].

For comparing the factor profiles resulting from the two factorisations in Equations (11) and 12, though, once computed, the reduced-size loadings in $\mathbf{A}_{\mathrm{r}}$, $\mathbf{B}_{\mathrm{r}}$ and $\mathbf{C}_{\mathrm{r}}$ will need to be expanded so as to cover the entire row, column and tube space of the original observations. This operation can be performed as follows: let $\mathbf{D}_{1,\mathrm{ERs}}$ be the $N_{\mathrm{r}} \times JK$ matrix obtained from the row-wise unfolding of $\underline{\mathbf{D}}$ and carrying only its corresponding ERs. Regressing $\mathbf{D}_{1,\mathrm{ERs}}$ onto $\mathbf{A}_{\mathrm{r}}$, one can calculate an expanded mixed column-tube factor matrix $\widehat{\mathbf{B}\mathbf{C}_{\mathrm{r}}^{\mathrm{T}}}$ ($F \times JK$) as

$$\widehat{\mathbf{B}\mathbf{C}_{\mathrm{r}}^{\mathrm{T}}} = \mathbf{A}_{\mathrm{r}}^{+}\mathbf{D}_{1,\mathrm{ERs}} \quad (13)$$

where $^{+}$ denotes the pseudo-inverse operator. Every single row of $\widehat{\mathbf{B}\mathbf{C}_{\mathrm{r}}^{\mathrm{T}}}$ can then be refolded into a $J \times K$ matrix of rank 1 and subsequently subjected to a singular value decomposition (SVD [52]). Expanded column and tube space factors ($\widehat{\mathbf{B}_{\mathrm{r}}}$, $J \times F$, and $\widehat{\mathbf{C}_{\mathrm{r}}}$, $K \times F$) are finally retrieved gathering all the left and right singular vectors SVD returns. At this point, expanded row space factor profiles ($\widehat{\mathbf{A}_{\mathrm{r}}}$, $N \times F$) can be estimated in a least-squares sense as

$$\widehat{\mathbf{A}_{\mathrm{r}}} = \mathbf{D}_{1,\mathrm{ERs}}\left(\widehat{\mathbf{C}_{\mathrm{r}}} \odot \widehat{\mathbf{B}_{\mathrm{r}}}\right)^{+} \quad (14)$$

with $\odot$ connoting the Khatri-Rao product [53].

It is worth noticing here that given the inherent symmetry of the trilinear models PARAFAC-ALS provides, expanded PARAFAC-ALS loadings can also be derived through alternative column-wise or tube-wise unfolding-based approaches provided that Equations (13) and (14) are adapted and modified accordingly. Only the computational procedure relying on row-wise unfolding is detailed in this section since this procedure readily provides PARAFAC-ALS factor profiles normalised along all but the first data mode, which is a way to report them that is rather standard in the scientific literature.

3 | Datasets

The comparative study described in the previous section was carried out on two real-world and two simulated datasets.

3.1 | Time-Resolved Fluorescence Spectra of Dye Mixtures

Nine mixtures of three different fluorescent dyes, Alexa 647, Atto 655 and Atto 665, were prepared as detailed in [10] and their respective fluorescence decay curves recorded by using a PicoQuant Time-Correlated Single Photon Counting (TCSPC) system coupled to a FluoTime 200 spectrometer equipped with a picosecond laser diode emitting at 640 nm. These decay curves follow multiexponential functions that are the combination of monoexponential contributions with distinct rate constant, each related to one of the three fluorophores under study. Data acquisition was conducted at 8 distinct emission wavelengths (from 650 to 720 nm with a step of 10 nm) which

resulted in a three-way array of dimensions 9 samples × 2500 time points × 8 emission wavelength channels assumed to exhibit trilinearity among these three measurement modes. The multilinear resolution of such an array permits to retrieve the aforementioned monoexponential signal contributions together with the abundance/concentration profiles and the individual spectral fingerprint of the three chemical species underlying the investigated specimens.

3.2 | Sequential Illumination Fluorescence Image of Osteosarcoma-Affected Bone Tissue Cells

U2OS cells stained with two different fluorophores, $DiBAC_4(3)$ and Concavalin A-Alexa 488, were inspected under a Nikon Ti-E2 fluorescence microscope equipped with a 488 nm laser source and configured in sequential illumination mode [54, 55]. The conducted measurements yielded a four-dimensional image of dimensions $246 \times 238 \times 6 \times 45$ where pixels are underlain by 6 fluorescence decay signals recorded over 45 consecutive time instants across a series of photobleaching-photorecovery cycles. During these cycles, the laser power was modulated differently in an attempt to emphasise the distinctive dynamic behaviours of the various fluorescent species bound to the individual compartments of the system under study. Also such decays follow multiexponential functions which are the combination of a certain number of monoexponential contributions, each one related to a given fluorescent species located in a particular chemical environment. Before being finally processed by means of PARAFAC-ALS, the investigated image was preliminarily unfolded pixel-wise. This operation returned a three-dimensional data structure assumed to exhibit trilinearity among the pixel, the photobleaching-photorecovery cycle and the time mode.

3.3 | Simulated Multi-Way Images

In the light of the remarkable results obtained in the two real-world case-studies described before—whose comprehensive illustration is provided in the next sections—and in an attempt to validate the algorithmic procedure proposed in this article, two multi-way images of size $59 \times 59 \times 10 \times 60$ were generated for a two and a three component system, respectively. Every pixel of both images is underlain by a series of 10 time point decay signals recorded at 60 different wavelength channels and following multiexponential functions which are the combination of monoexponential contributions related to the aforementioned components. The data were simulated assuming trilinearity among the pixel, the time and the wavelength mode of the two four-dimensional images and accounting for heteroscedastic Gaussian noise. Therefore, before being finally processed by means of PARAFAC-ALS, they were preliminarily unfolded pixel-wise.

4 | Comparative Study

In all the scenarios outlined before, once full-size and reduced-size datasets were processed as described in Section 2.2, the quality and the adequacy of the final trilinear factorisation models were assessed in terms of residual lack-of-fit (LOF) and core consistency (CORCONDIA [56]). For full-size data, LOF and CORCONDIA are estimated as:

$$\mathrm{LOF} = \frac{\sum_{n=1}^{N} \sum_{jk=1}^{JK} e_{n,jk}^2}{\sum_{n=1}^{N} \sum_{jk=1}^{JK} d_{1\,n,jk}^{\;2}} \tag{15}$$

$$\mathrm{CORCONDIA} = 100\left(1 - \frac{\sum_{a=1}^{F} \sum_{b=1}^{F} \sum_{c=1}^{F} \left(g_{a,b,c} - t_{a,b,c}\right)}{F}\right) \tag{16}$$

with $e_{n,jk}$ and $d_{1\,n,jk}$ being the $n \times jk$ -th elements of $\mathbf{E}$ and $\mathbf{D}_1$, respectively, $g_{a,b,c}$ the $a \times b \times c$ -th element of the Tucker3 core, say $\underline{\mathbf{G}}$, computed via least squares from $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ and $t_{a,b,c}$ the $a \times b \times c$ -th element of the $F \times F \times F$ unit superdiagonal core, $\underline{\mathbf{T}}$. Equations (15) and (16) (appropriately adapted) were also exploited to calculate LOF and CORCONDIA values when reduced-size data were handled as well as after the expansion of the reduced-size PARAFAC-ALS factor profiles and the subsequent recalculation of the PARAFAC-ALS core and the PARAFAC-ALS residuals from such expanded factor profiles. Full-size and expanded PARAFAC-ALS factor profiles were also compared in terms of correlation by estimating the cosine of the angle between each pair of loadings corresponding to the same extracted component. In case one of the modes of the data under study encoded information regarding the pixels of an image, though, the PARAFAC-ALS loadings along this mode were contrasted in terms of Ruzicka's similarity [57], an extension of the Jaccard index [58–60] for real-valued vectors. The reduction rate yielded by the methodology proposed here and defined as

$$\mathrm{RR} = 100 - \frac{N_{\mathrm{r}} J_{\mathrm{r}} K_{\mathrm{r}}}{NJK} 100 \tag{17}$$

was also assessed together with the operational time required to factorise the full-size data and to derive, factorise and post-process the compressed data.

5 | Results and Discussion

5.1 | Time-Resolved Fluorescence Spectra of Dye Mixtures

Table 1 and Figures 1 and 2 illustrate the outcomes obtained for the time-resolved fluorescence spectra described in Section 3.1.

Here, F was set equal to 3 as the number of chemical compounds utilised for sample preparation and the trilinear model resulting from the application of PARAFAC-ALS to the compressed data yielded LOF and CORCONDIA values indicating that it can adequately explain all the sources of variability underlying them. Besides, the PARAFAC-ALS loadings profiles retrieved after the expansion of their corresponding reduced-size ones are in good agreement with those extracted from full-size measurements. The computational time required to individuate ERs, ECs and ETs, process them and, afterwards, *decompress* PARAFAC-ALS factors, though, was found to be about 800 times lower than when such full-size measurements were analysed. Moreover, the fact that 4 out of the 6 identified ERs relate to chemical mixtures constituted by only

TABLE 1 | Time-resolved fluorescence spectra of dye mixtures: dataset rank and dimensions before and after the EI-based compression, reduction rate yielded by the EI-based compression, processing time required for attaining the trilinear decomposition of the full-size and reduced-size data, LOF and CORCONDIA values resulting from the PARAFAC-ALS factorisation of the full-size and reduced-size data (before and after the expansion of the PARAFAC-ALS loadings profiles).

	F	Rows	Columns	Tubes	RR	Processing time (s)	LOF (%)	CORCONDIA	LOF (after loadings expansion, %)	CORCONDIA (after loadings expansion)
Full-size data	3	9	2500	8	—	881.33	0.03	99.51	—	—
Reduced-size data	3	6	24	6	99.52	1.12[a]	0.04	93.71	0.03	98.10

[a]This value accounts for (i) the time required for the HOSVD decomposition and all the convex hull-related calculations, (ii) the time required for the PARAFAC-ALS factorisation and (iii) the time required for the reconstruction of the expanded PARAFAC-ALS factor profiles.

2 of the 3 dyes concerned and for which the PARAFAC-ALS loadings profile of the missing component assumes values approximately equal to 0. This clearly evidences that the proposed EI-based reduction procedure is capable of dramatically decreasing the redundancy of the information encoded in the analysed datasets prior to their final decomposition. Emission wavelength channels characterised by a strong overlap among the three different PARAFAC-ALS factor profiles are also satisfactorily recognised as non-essential and, thus, excluded from the compression operations.

5.2 | Sequential Illumination Fluorescence Image of Osteosarcoma-Affected Bone Tissue Cells

The outcomes resulting from the analysis of the sequential illumination fluorescence image recorded as detailed in Section 3.2 are summarised in Table 2 and illustrated in Figures 3 and 4.

Also in this case, the EI-based compression procedure enabled a significant reduction of the data under study, which permitted to attain in a substantially reduced amount of time (namely, about 30-fold faster) a trilinear factor model capable of satisfactorily accounting for the main sources of variability underlying them and yielding expanded PARAFAC-ALS loadings profiles in good agreement with those extracted through the decomposition of the original measurements collected. Notice that F was here set equal to 3 based on preliminary knowledge about the number of components expected to be observed within the captured scene [55]. In addition, Figure 4 clearly shows that the ERs, ECs and ETs identified in this specific contingency correspond to rows, columns and tubes of the full-size tensor at hand in correspondence of which distinctive cell compartments can be observed, the overlap among PARAFAC-ALS factor profiles is minimal or their trend is not particularly correlated. This, once again, highlights how the algorithmic procedure outlined in Section 2.1 can considerably decrease the redundancy of the information encoded in the analysed datasets.

5.3 | Simulated Multi-Way Images

Given the remarkable performance that the EI-based data reduction approach proposed here guaranteed in the two real-world case-studies reported before and in order to assess its validity and reliability, the two- and three-component simulated imaging datasets described in Section 3.3 were finally processed. The outcomes obtained for the former are summarised in Table 3.

It should be noticed here that the reported indices actually refer to a series of 500 replications of the entire analysis procedure performed on as many images resulting from different random noise generation repetitions. As expected, in all cases, 2 ERs, 2 ECs and 2 ETs were identified by the aforementioned EI-based compression strategy: this stems from the fact that, having set F equal to 2, the recognition of ERs, ECs and ETs is always carried out within one-dimensional subspaces which lead to convex hulls supported only by 2 different normalised HOSVD scores points (see Figure S1a for an explanatory example). In spite of this substantial dimensionality reduction, all the two-factor trilinear models built on full-size and compressed datasets yielded satisfactory LOF and CORCONDIA values, which suggest that they can fully and adequately explain the meaningful sources of variability underlying the three-way arrays under study. Low LOF and high CORCONDIA estimates were also found after the expansion of the reduced-size PARAFAC-ALS loadings and the recalculation of the PARAFAC-ALS core tensors and residuals. For the sake of comparison, Figure S2 provides an illustration of the PARAFAC-ALS loadings returned by the decomposition of the original full-size image and those produced once expanded the PARAFAC-ALS loadings calculated from its reduced-size counterpart in one of the aforementioned replication rounds. It is evident how both approaches output strikingly matching factor profiles. The main advantage brought by the EI-based compression, though, relates to the dramatic decrease of the computational time required to derive reduced-size data, execute their decomposition and subsequently expand the reduced-size PARAFAC-ALS loadings: performing these three operations took approximately 10 times less (and only a single PARAFAC-ALS iteration when non-negativity was not explicitly constrained) than factorising full-size data. Last but not least, as shown in Figure S3 for the same simulation repetition considered before, the tensor rows, columns and tubes spotted as essential correspond to spatial locations (pixels), wavelength channels and time instants of minimal overlap between components. This implies a significant reduction of the redundancy of the captured information and, thus, a clear enhancement of its selectivity that does not necessarily

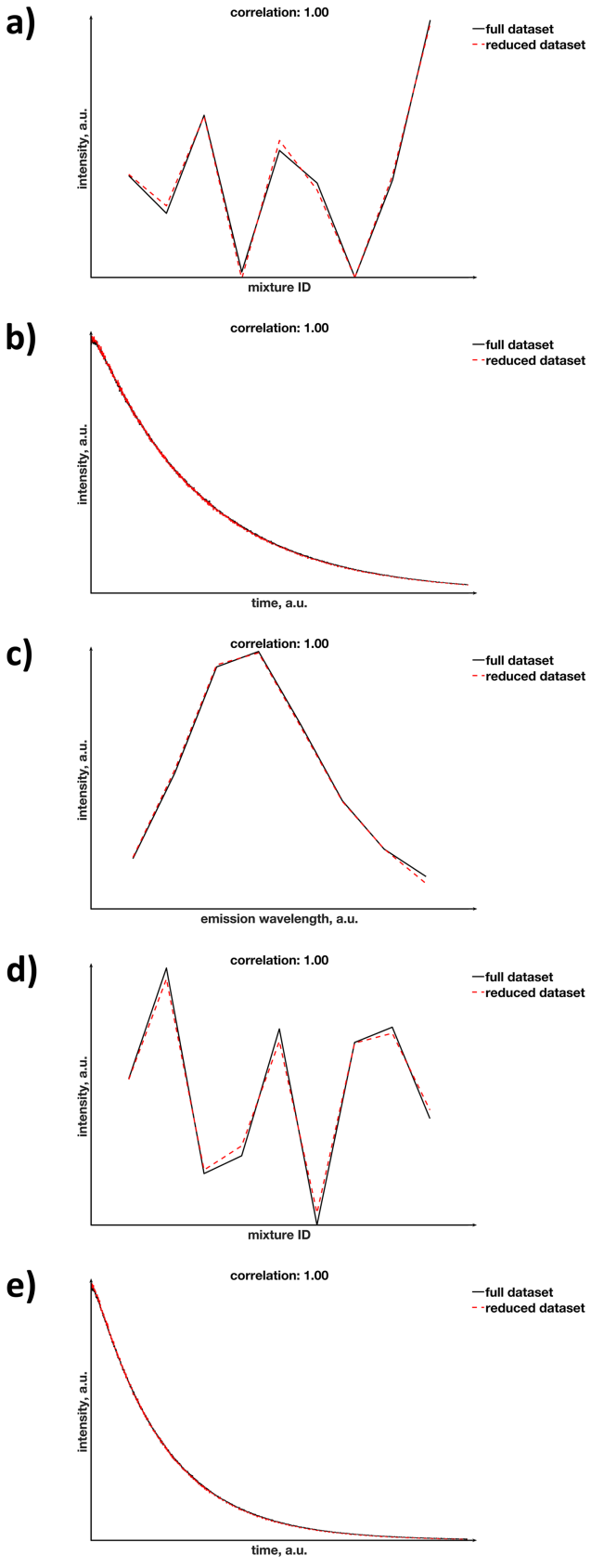

FIGURE 1 | Time-resolved fluorescence spectra of dye mixtures: comparison between full-size and expanded PARAFAC-ALS loadings profiles. Panels (a–c), (d–f) and (g–i) refer to the first, second, and third components resulting from the trilinear factorisation of the original and compressed data, respectively.

FIGURE 1 | (Continued)

translate into a more accurate trilinear factorisation (see also Section 1) but ensures all these components are appropriately represented even after the compression process and overall guarantees a faster convergence of PARAFAC-ALS. All these conclusions are also corroborated by the results attained when the three-component simulated data were concerned (see Table 4 and Figures S4 and S5).

Nonetheless, in this particular scenario, a larger amount of ERs, ECs and ETs was consistently extracted from the distinct multi-way images coped with: given their higher effective rank ($F = 3$), indeed, the normalised HOSVD scores points lie now on a plane (i.e., ERs, ECs and ETs are identified within

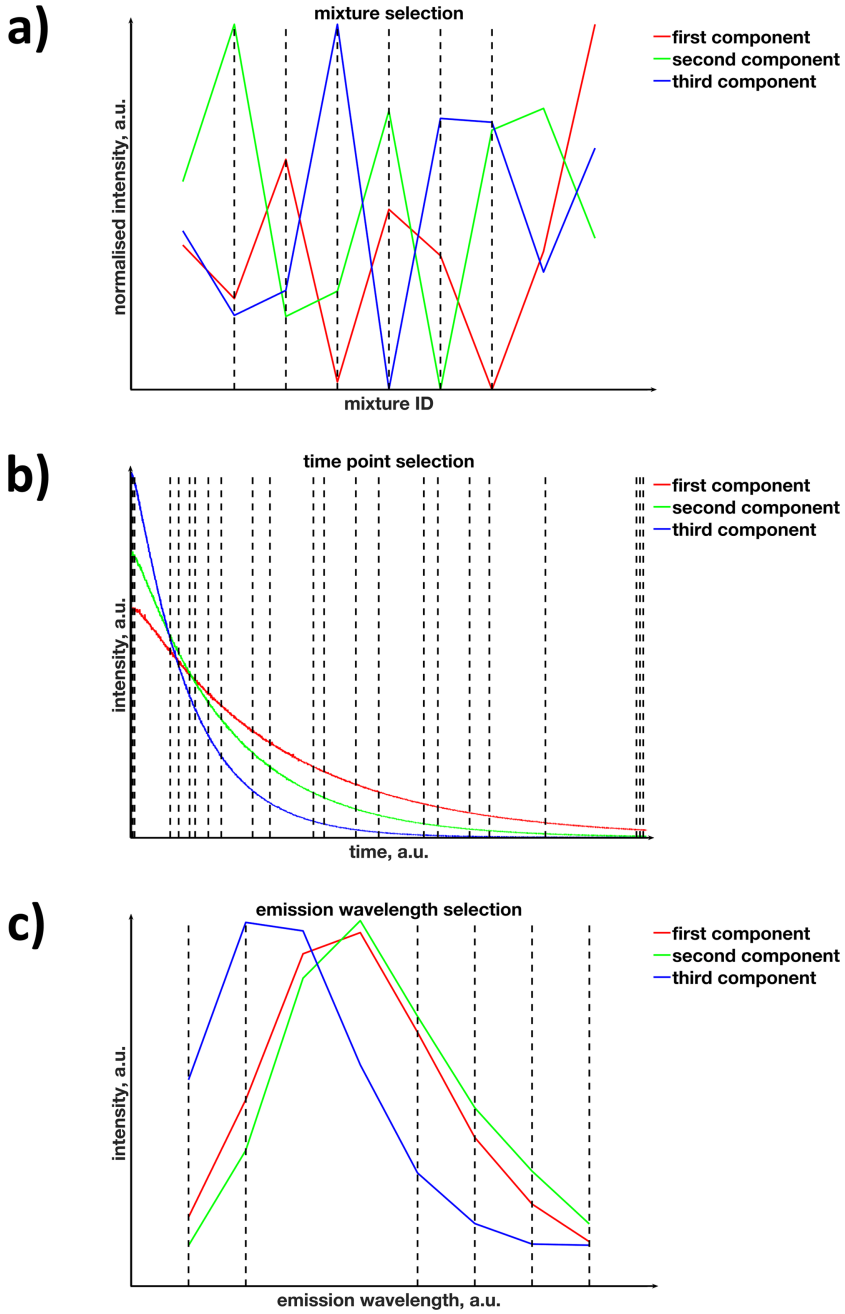

FIGURE 2 | Time-resolved fluorescence spectra of dye mixtures: first (red), second (green) and third (blue) PARAFAC-ALS component loadings along (a) the mixture, (b) the time and (c) the emission wavelength mode resulting from the trilinear factorisation of the full-size data at hand. Dashed lines highlight the mixture samples, time instants and emission wavelength channels identified as essential by the data reduction procedure outlined in Section 2.1.

TABLE 2 | Sequential illumination fluorescence image of osteosarcoma-affected bone tissue cells: dataset rank and dimensions before and after the EI-based compression, reduction rate yielded by the EI-base compression, processing time required for attaining the trilinear decomposition of the full-size and reduced-size data, LOF and CORCONDIA values resulting from the PARAFAC-ALS factorisation of the full-size and reduced-size data (before and after the expansion of the PARAFAC-ALS loadings profiles).

	F	Rows	Columns	Tubes	RR	Processing time (s)	LOF (%)	CORCONDIA	LOF (after loadings expansion, %)	CORCONDIA (after loadings expansion)
Full-size data	3	58,548	6	45	—	85.54	0.17	84.28	—	—
Reduced-size data	3	20	5	10	99.99	3.10[a]	0.10	97.74	0.18	84.98

[a]This value accounts for (i) the time required for the HOSVD decomposition and all the convex hull-related calculations, (ii) the time required for the PARAFAC-ALS factorisation and (iii) the time required for the reconstruction of the expanded PARAFAC-ALS factor profiles.

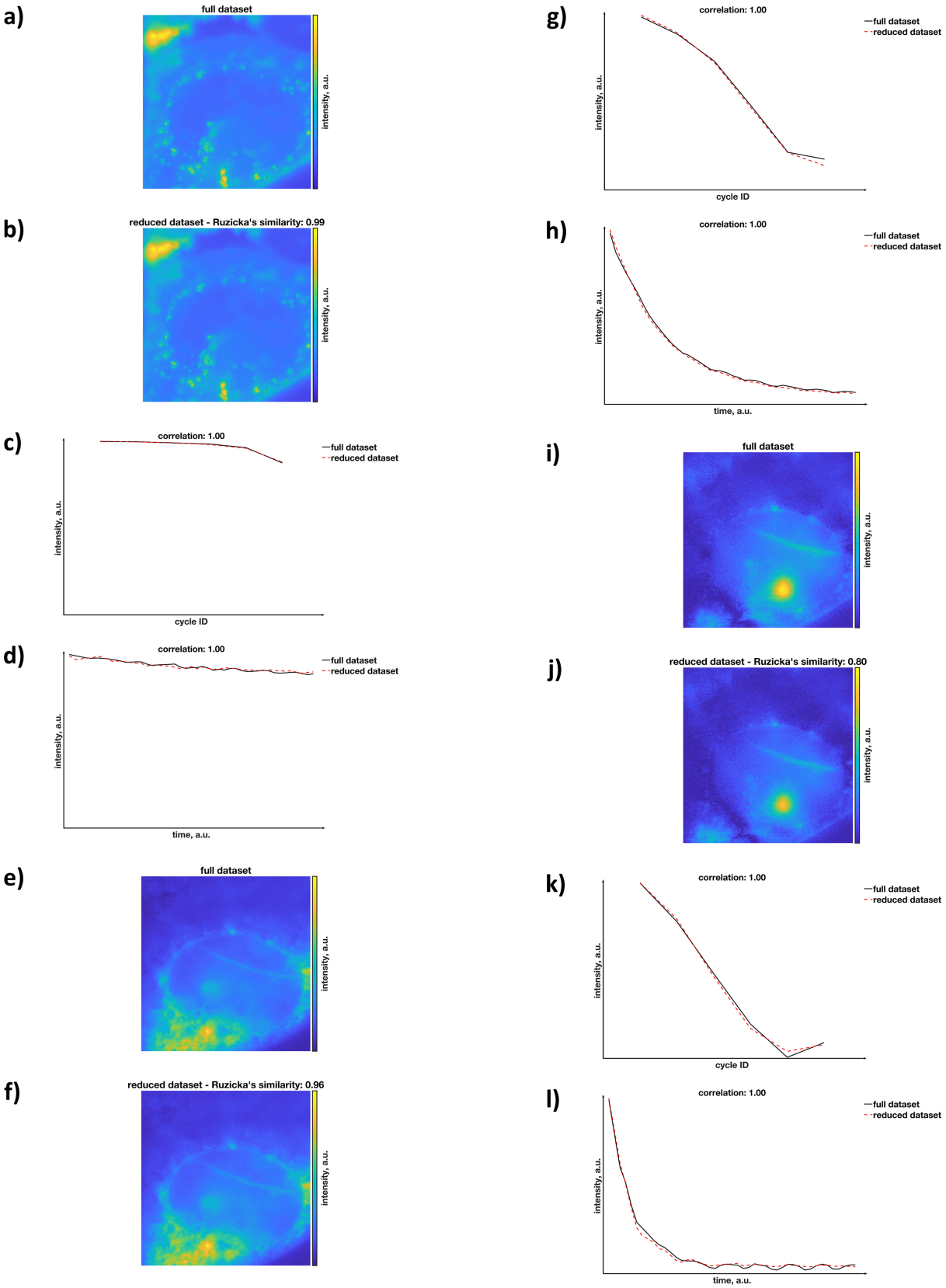

FIGURE 3 | Sequential illumination fluorescence image of osteosarcoma-affected bone tissue cells: comparison between full-size and expanded PARAFAC-ALS loadings profiles. Panels (a–d), (e–h) and (i–l) refer to the first, second, and third components resulting from the trilinear factorisation of the original and compressed data, respectively. PARAFAC-ALS loadings along the pixel model were refolded according to the size of the image under study.

FIGURE 3 | (Continued)

two-dimensional subspaces), but their distribution is made irregular by the presence of noise and the subsequent lack of strict selectivity among the investigated observations (see

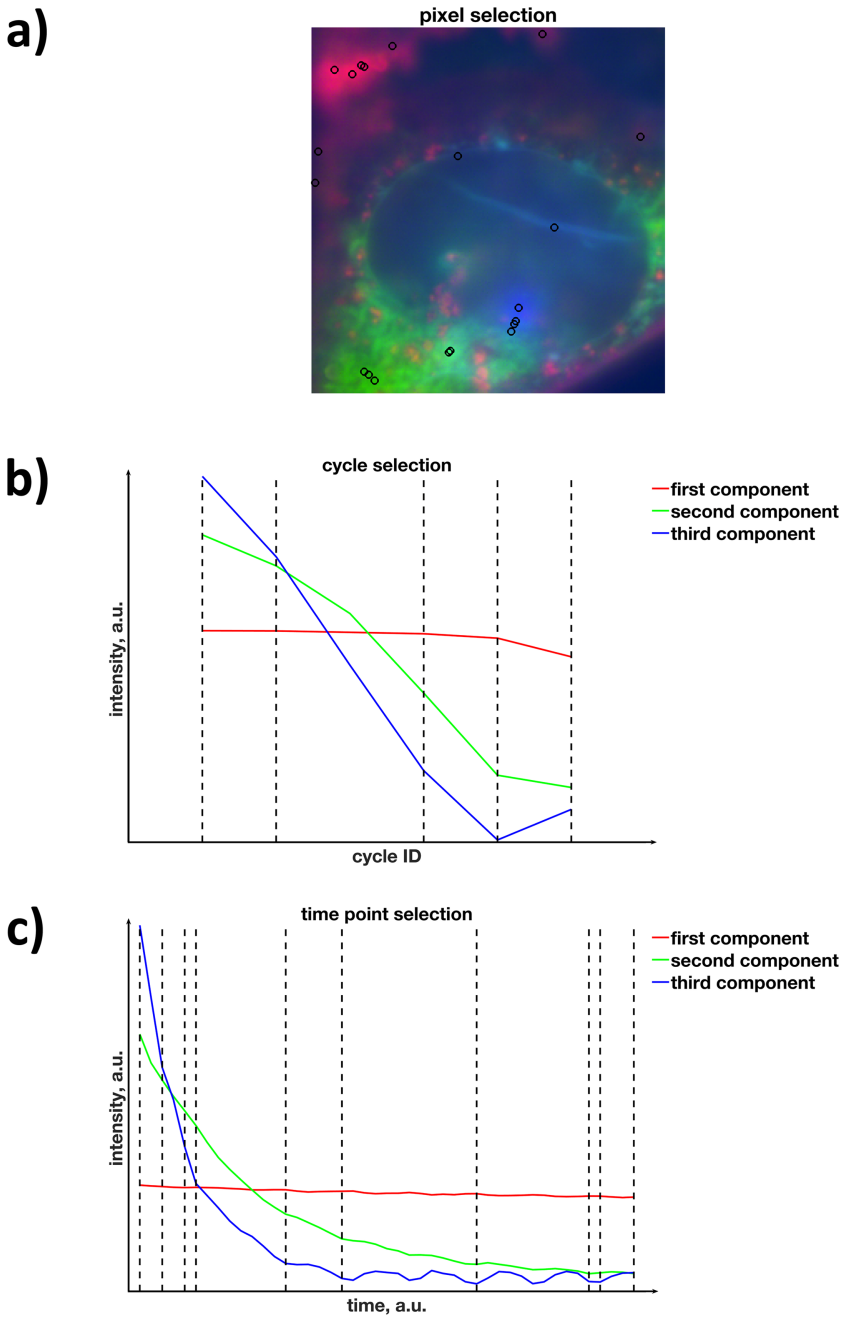

FIGURE 4 | Sequential illumination fluorescence image of osteosarcoma-affected bone tissue cells: first (red), second (green) and third (blue) PARAFAC-ALS component loadings along (a) the pixel, (b) the photobleaching-photorecovery cycle and (c) the time cycle mode resulting from the trilinear factorisation of the full-size data at hand. Black dots and dashed lines highlight the pixel locations, cycle repetitions and time instants identified as essential by the data reduction procedure outlined in Section 2.1. In (a), all the PARAFAC-ALS loadings profiles are simultaneously displayed (after refolding) in a single false-colour RGB image.

Figure S1b). This leads to convex hulls more complex-shaped and supported by a greater amount of vertices. Furthermore, it can also be highlighted that the whole ensemble of decay time instants was here systematically recognised as essential compared to when the sequential illumination fluorescence image was handled. This originates from the fact that the lower signal-to-noise ratio such an image features naturally induces the time-mode normalised HOSVD scores to be more widely scattered within their respective space and not aligned along a continuous non-linear trajectory like in this specific situation (see Figure S6). Ultimately, in an attempt to better point out the benefits that the algorithmic scheme sketched in Section 2.1 could bring in terms of processing time saving, additional tests were conducted gradually increasing the size of the time mode of this last dataset. The outcomes of these additional tests are illustrated in Figure S7.

TABLE 3 | Two-component simulated multi-way image: dataset rank and dimensions before and after the EI-based compression, reduction rate yielded by the EI-based compression, processing time required for attaining the trilinear decomposition of the full-size and reduced-size data, LOF and CORCONDIA values resulting from the PARAFAC-ALS factorisation of the full-size and reduced-size data (before and after the expansion of the PARAFAC-ALS loadings profiles).

	F	Rows	Columns	Tubes	RR	Processing time (s)	LOF (%)	CORCONDIA	LOF (after loadings expansion, %)	CORCONDIA (after loadings expansion)
Full-size data	2	3481	10	60	—	4.19 [2.98–5.55]	0.23 [0.22–0.24]	99.99 [99.98–100]	—	—
Reduced-size data	2	2 [2–2]	2 [2–2]	2 [2–2]	99.99 [99.99–99.99]	0.37 [0.34–0.42][a]	$2.6E^{-5}$ [$2.0E^{-6}$ to $2.0E^{-4}$]	99.91 [99.29–100]	0.35 [0.28–0.45]	99.82 [99.24–99.98]

Note: More specifically, the table provides (i) the median number and range (in brackets) of ERs, ECs and ETs identified across the 500 noise generation rounds, (ii) the mean reduction rate obtained across the 500 noise generation rounds, (iii) the mean processing time needed for achieving the PARAFAC-ALS factorisation of full-size and reduced-size data across the 500 noise generation rounds and (iv) mean LOF and CORCONDIA values computed across the 500 noise generation rounds. The intervals associated to reduction rate, processing time, LOF and CORCONDIA estimates are delimited by the 2.5th and 97.5th percentiles of their corresponding distributions across the 500 noise generation rounds.

[a]These values account for (i) the time required for the HOSVD decomposition and all the convex hull-related calculations, (ii) the time required for the PARAFAC-ALS factorisation and (iii) the time required for the reconstruction of the expanded PARAFAC-ALS factor profiles.

TABLE 4 | Three-component simulated multi-way image: dataset rank and dimensions before and after the EI-based compression, reduction rate yielded by the EI-based compression, processing time required for attaining the trilinear decomposition of the full-size and reduced-size data, LOF and CORCONDIA values resulting from the PARAFAC-ALS factorisation of the full-size and reduced-size data (before and after the expansion of the PARAFAC-ALS loadings profiles).

	F	Rows	Columns	Tubes	RR	Processing time (s)	LOF (%)	CORCONDIA	LOF (after loadings expansion, %)	CORCONDIA (after loadings expansion)
Full-size data	3	3481	10	60	—	12.90 [9.54–15.48]	0.24 [0.23–0.25]	99.99 [99.98–100]	—	—
Reduced-size data	3	18 [13–24]	10 [10–10]	12 [7–17]	99.89 [99.85–99.93]	0.60 [0.52–0.69][a]	0.20 [0.15–0.25]	99.11 [98.10–99.85]	0.25 [0.24–0.26]	99.74 [99.22–99.95]

Note: More specifically, the table provides (i) the median number and range (in brackets) of ERs, ECs and ETs identified across the 500 noise generation rounds, (ii) the mean reduction rate obtained across the 500 noise generation rounds, (iii) the mean processing time needed for achieving the PARAFAC-ALS factorisation of full-size and reduced-size data across the 500 noise generation rounds and (iv) mean LOF and CORCONDIA values computed across the 500 noise generation rounds. The intervals associated to reduction rate, processing time, LOF and CORCONDIA estimates are delimited by the 2.5th and 97.5th percentiles of their corresponding distributions across the 500 noise generation rounds.

[a]These values account for (i) the time required for the HOSVD decomposition and all the convex hull-related calculations, (ii) the time required for the PARAFAC-ALS factorisation and (iii) the time required for the reconstruction of the expanded PARAFAC-ALS factor profiles.

6 | Conclusions

In this article, a novel approach for the EI-based reduction of three-way datasets was proposed. Differently from existing methodologies devised for the same purpose and mainly relying on random sampling, such an approach exploiting the principle of HOSVD and the properties of convex geometry is capable of retaining uniquely the most meaningful information, encoded in the investigated measurements, that is needed for their sensible trilinear factorisation. The results obtained from the application of this innovative compression strategy in two real-world and two simulated case-studies permitted to highlight three key aspects:

- PARAFAC-ALS decompositions of reduced-size data can be up to 800 times faster than PARAFAC-ALS decompositions of full-size data;
- PARAFAC-ALS decompositions of full-size and reduced-size data yield PARAFAC-ALS loadings profiles that are indistinguishable from the perspective of self-modelling curve resolution—after the reduced-size ones are expanded as outlined in Section 2.2;
- PARAFAC-ALS model quality and adequacy is preserved once the aforementioned compression operation is carried out.

The benefits that users and operators could then get from it go substantially without saying. The only requirement to run the algorithmic procedure described in Section 2.1 is the prior estimation of the effective rank of the analysed data structures, F. Although tackling this task might not be trivial, it is worth mentioning that setting exactly this parameter may not be critical either: in fact, one could always think of slightly overestimating F in the compression stage—this would yield marginally larger datasets containing anyway the EI captured by the original recordings—and readjusting it only for the sake of the final trilinear factorisation depending also on the values quality metrics like LOF and CORCONDIA assume. It should additionally be noticed that, even if the main focus of this work was on PARAFAC-ALS, the designed data reduction scheme can be coupled to a wider range of three- and multi-way modelling techniques like Parallel Profiles with Linear Dependences-ALS (PARALIND-ALS [61]) or Tucker-ALS and all its variants [50]. This can be achieved by a different modulation of the compression rate along each individual data mode. Finally, it has to be stressed that the approach developed here builds upon recent advances in the domain of bilinear curve resolution [27–29] and constitutes an evidence of how the principle of convex hull estimation for the identification of essential rows and columns of two-way arrays can be successfully extended to higher order ones. This basically arises from the common mathematical properties shared by standard SVD and HOSVD.

Acknowledgements

The authors acknowledge Dario Cevoli and Prof. Peter Dedecker (KU Leuven, Belgium) for having kindly provided the sequential illumination fluorescence imaging data as well as Olivier Devos for the collection of the time-resolved fluorescence spectra. Nemat Omidikia and Azar Azizi acknowledge the University of Sistan and Baluchestan for financial support.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] For the sake of clarity, the local rank of a data matrix corresponds to the number of components/factors underlying specific portions of it, that is, specific windows of its rows and columns [30].

[2] It is worth noticing here that the columns of $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ correspond to the left singular vectors of the two-way arrays resulting from the row-wise, column-wise and tube-wise unfolding of $\underline{\mathbf{D}}$ [46].

References

1. R. A. Harshman, "Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis," in *UCLA Working Papers in Phonetics*, Vol. 16 (1970): 1–84.

2. J. D. Carroll and J. J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of "Eckart-Young" Decomposition," *Psychometrika* 35 (1970): 283–319.

3. R. Bro, "PARAFAC. Tutorial and Applications," *Chemometrics and Intelligent Laboratory* 38 (1997): 149–171.

4. M. H. Van Benthem, M. R. Keenan, R. W. Davis, et al., "Trilinear Analysis of Images Obtained With a Hyperspectral Imaging Confocal Microscope," *Journal of Chemometrics* 22 (2008): 491–499.

5. M. H. Van Benthem, T. W. Lane, R. W. Davis, P. D. Lane, and M. R. Keenan, "PARAFAC Modeling of Three-Way Hyperspectral Images: Endogenous Fluorophores as Health Biomarkers in Aquatic Species," *Chemometrics and Intelligent Laboratory* 106 (2011): 115–124.

6. G. L. Alexandrino, J. M. Amigo, M. R. Khorasani, J. Rantanen, A. V. Friderichsen, and R. J. Poppi, "Unveiling Multiple Solid-State Transitions in Pharmaceutical Solid Dosage Forms Using Multi-Series Hyperspectral Imaging and Different Curve Resolution Approaches," *Chemometrics and Intelligent Laboratory* 161 (2017): 136–146.

7. M. Ghaffari, A. L. Chateigner-Boutin, F. Guillon, M. F. Devaux, A. Abdollahi, and L. Duponchel, "Multi-Excitation Hyperspectral Autofluorescence Imaging for the Exploration of Biological Samples," *Analytica Chimica Acta* 1062 (2019): 47–59.

8. A. Gómez-Sánchez, M. Marro, M. Marsal, P. Loza-Alvarez, and A. de Juan, "3D and 4D Image Fusion: Coping With Differences in Spectroscopic Modes Among Hyperspectral Images," *Analytical Chemistry* 92 (2020): 9591–9602.

9. R. Rocha de Olivera and A. de Juan, "Design of Heterogeneity Indices for Blending Quality Assessment Based on Hyperspectral Images and Variographic Analysis," *Analytical Chemistry* 92 (2020): 15880–15889.

10. O. Devos, M. Ghaffari, R. Vitale, A. de Juan, M. Sliwa, and C. Ruckebusch, "Multivariate Curve Resolution Slicing of Multiexponential Time-Resolved Spectroscopy Fluorescence Data," *Analytical Chemistry* 93 (2021): 12504–12513.

11. M. Rajih, P. Comon, and R. A. Harshman, "Enhanced Line Search: A Novel Method to Accelerate PARAFAC," *SIAM Journal on Matrix Analysis and Applications* 30 (2008): 1128–1147.

12. K. Tian, L. Wu, S. Min, and R. Bro, "Geometric Search: A New Approach for Fitting PARAFAC2 Models on GC-MS Data," *Talanta* 185 (2018): 378–386.

13. H. Yu, D. Augustijn, and R. Bro, "Accelerating PARAFAC2 Algorithms for Non-Negative Complex Tensor Decomposition," *Chemometrics and Intelligent Laboratory* 214 (2021): 104312.

14. N. Omidikia, "The Effect of Multilinear Data Fusion on the Accuracy of Multivariate Curve Resolution Outputs," *Analytica Chimica Acta* 1227 (2022): 340325.

15. B. K. Alsberg and O. M. Kvalheim, "Compression of Three-Mode Data Arrays by B-Splines Prior to Three-Mode Principal Component Analysis," *Chemometrics and Intelligent Laboratory* 23 (1994): 29–38.

16. K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophical Magazine* 2 (1901): 559–572.

17. H. Hotelling, "Analysis of a Complex of Statistical Variables Into Principal Components," *Journal of Educational Psychology* 24 (1933): 417–441.

18. L. R. Tucker, *Problems in Measuring Change* (Madison, United States of America: The University of Wisconsin Press, 1963): 122–137.

19. L. R. Tucker, *Contributions to Mathematical Psychology* (New York, USA: Holt, Rinehart and Winston, Inc., 1964): 110–182.

20. L. R. Tucker, "Some Mathematical Notes on Three-Mode Factor Analysis," *Psychometrika* 31 (1966): 279–311.

21. P. M. Kroonenberg and J. de Leeuw, "Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms," *Psychometrika* 45 (1980): 69–97.

22. P. M. Kroonenberg, *Three-Mode Principal Component Analysis. Theory and Applications*, First (Leiden, The Netherlands: DSWO Press, 1983).

23. P. M. Kroonenberg, *Research Methods for Multimode Data Analysis* (New York, USA: Praeger Publishers, 1984), 64–103.

24. H. A. L. Kiers and R. A. Harshman, "Relating Two Proposed Methods for Speedup of Algorithms for Fitting Two- and Three-Way Principal Component and Related Multilinear Models," *Chemometrics and Intelligent Laboratory* 36 (1997): 31–40.

25. R. Bro and C. A. Andersson, "Improving the Speed of Multiway Algorithms: Part II: Compression," *Chemometrics and Intelligent Laboratory* 42 (1998): 105–113.

26. M. H. Van Benthem and M. R. Keenan, "Tucker1 Model Algorithms for Fast Solutions to Large PARAFAC Problems," *Journal of Chemometrics* 22 (2008): 345–354.

27. M. Ghaffari, N. Omidikia, and C. Ruckebusch, "Essential Spectral Pixels for Multivariate Curve Resolution of Chemical Images," *Analytical Chemistry* 91 (2019): 10943–10948.

28. C. Ruckebusch, R. Vitale, M. Ghaffari, S. Hugelier, and N. Omidikia, "Perspective on Essential Information in Multivariate Curve Resolution," *Trends in Analytical Chemistry* 132 (2020): 116044.

29. M. Ghaffari, N. Omidikia, and C. Ruckebusch, "Joint Selection of Essential Pixels and Essential Variables Across Hyperspectral Images," *Analytica Chimica Acta* 1141 (2021): 36–46.

30. R. Tauler, A. K. Smilde, and B. Kowalski, "Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution," *Journal of Chemometrics* 9 (1995): 31–58.

31. R. Zimmerleiter, R. Nikzad-Langerodi, C. Ruckebusch, et al., "QCL-Based Mid-Infrared Hyperspectral Imaging of Multilayer Polymer Oxygen Barrier-Films," *Polymer Testing* 98 (2021): 107190.

32. L. Coïc, P. Y. Sacré, A. Dispas, et al., "Pixel-Based Raman Hyperspectral Identification of Complex Pharmaceutical Formulations," *Analytica Chimica Acta* 1155 (2021): 338361.

33. S. Beyramysoltan, A. Abdollahi, and R. A. Musah, "Workflow for the Supervised Learning of Chemical Data: Efficient Data Reduction-Multivariate Curve Resolution (EDR-MCR)," *Analytical Chemistry* 93 (2021): 5020–5027.

34. L. Coïc, P. Y. Sacré, A. Dispas, et al., "Selection of Essential Spectra to Improve the Multivariate Curve Resolution of Minor Compounds in Complex Pharmaceutical Formulations," *Analytica Chimica Acta* 1198 (2022): 339532.

35. Q. Wu, C. Marina-Montes, J. O. Cáceres, J. Anzano, V. Motto-Ros, and L. Duponchel, "Interesting Features Finder (IFF): Another Way to Explore Spectroscopic Imaging Data Sets Giving Minor Compounds and Traces a Chance to Express Themselves," *Spectrochimica Acta Part B* 195 (2022): 106508.

36. R. Vitale and C. Ruckebusch, "On a *Black Hole* Effect in Bilinear Curve Resolution Based on Least Squares," *Journal of Chemometrics* 37 (2023): e3442.

37. M. Ahmad, R. Vitale, M. Cocchi, and C. Ruckebusch, "Weighted Multivariate Curve Resolution-Alternating Least Squares Based on Sample Relevance," *Journal of Chemometrics* 37 (2023): e3478.

38. S. K. Karimvand, J. M. Jafari, S. V. Zade, and A. Abdollahi, "Practical and Comparative Application of Efficient Data Reduction-Multivariate Curve Resolution," *Analytica Chimica Acta* 1243 (2023): 340824.

39. M. Ghaffari, M. C. J. Lukkien, N. Omidikia, et al., "Systematic Reduction of Hyperspectral Images for High-Throughput Plastic Characterization," *Scientific Reports* 13 (2023): 21591.

40. N. Omidikia, A. Abdollahi, and M. Kompany-Zareh, "On Uniqueness and Selectivity in Three-Component Parallel Factor Analysis," *Analytica Chimica Acta* 782 (2013): 12–20.

41. A. D. Palmer, J. Bunch, and I. B. Styles, "Randomized Approximation Methods for the Efficient Compression and Analysis of Hyperspectral Data," *Analytical Chemistry* 85 (2013): 5078–5086.

42. S. Kucheryavskiy, "Blessing of Randomness Against the Curse of Dimensionality," *Journal of Chemometrics* 32 (2018): e2966.

43. J. P. Cruz-Tirado, J. M. Amigo, D. Fernandes Barbin, and S. Kucheryavskiy, "Data Reduction by Randomization Subsampling for the Study of Large Hyperspectral Datasets," *Analytica Chimica Acta* 1209 (2022): 339793.

44. P. Drineas, R. Kannan, and M. W. Mahoney, "Fast Monte Carlo Algorithms for Matrices III: Computing a Compressed Approximate Matrix Decomposition," *SIAM Journal on Computing* 36 (2006): 184–206.

45. M. W. Mahoney and P. Drineas, "CUR Matrix Decomposition for Improved Data Analysis," *Proceedings of the National Academy of Sciences* 106 (2009): 697–702.

46. L. De Lathauwer, B. De Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM Journal on Matrix Analysis and Applications* 21 (2000): 1253–1278.

47. L. Omberg, G. H. Golub, and O. Alter, "A Tensor Higher-Order Singular Value Decomposition for Intergrative Analysis of DNA Microarray Data From Different Studies," *Proceedings of the National Academy of Sciences* 104 (2007): 18371–18376.

48. B. V. Grande and R. Manne, "Use of Convexity for Finding Pure Variables in Two-Way Data From Mixtures," *Chemometrics and Intelligent Laboratory* 50 (2000): 19–33.

49. R. Rajkó, "Studies on the Adaptability of Different Borgen Norms Applied in Self-Modeling Curve Resolution (SMCR) Method," *Journal of Chemometrics* 23 (2009): 265–274.

50. A. K. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*, First, (Chichester, United Kingdom: John Wiley & Sons, Ltd., 2004).

51. G. Zehfuss, "Ueber eine gewisse Determinante," *Zeitschrift für angewandte Mathematik und Physik* 3 (1858): 298–301.

52. C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika* 1 (1936): 211–218.

53. C. G. Khatri and C. R. Rao, "Solutions to Some Functional Equations and Their Applications to Characterization of Probability Distributions," *Sankhyā* 30 (1968): 167–180.

54. S. Hugelier, R. Van den Eynde, W. Vandenberg, and P. Dedecker, "Fluorophore Unmixing Based on Bleaching and Recovery Kinetics Using MCR-ALS," *Talanta* 226 (2021): 122117.

55. D. Cevoli, S. Hugelier, R. Van den Eynde, O. Devos, P. Dedecker, and C. Ruckebusch, "Multilinear Slicing for Curve Resolution of

Fluorescence Imaging With Sequential Illumination," *Talanta* 241 (2022): 123231.

56. R. Bro and H. A. L. Kiers, "A New Efficient Method for Determining the Number of Components in PARAFAC Models," *Journal of Chemometrics* 17 (2003): 274–286.

57. M. Růžička, "Anwendung mathematisch-statistischer Methoden in der Geobotanik (synthetische Bearbeitung von Aufnahmen)," *Biológia (Bratislava)* 13 (1958): 647–661.

58. G. K. Gilbert, "Finley's Tornado Predictions," *American Meteorological Journal* 1 (1884): 166–172.

59. P. Jaccard, "étude comparative de la distribution florale dans une portion des alpes et du jura," *Bulletin de la Société Vaudoise Des Sciences Naturelles* 37 (1901): 547–579.

60. T. T. Tanimoto, "An Elementary Mathematical Theory of Classification and Prediction," International Business Machines Corporation, New York, USA, (1958).

61. R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling Multi-Way Data With Linearly Dependent Loadings," *Journal of Chemometrics* 23 (2009): 324–340.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A

Here, the relationship between the HOSVD core $\underline{\mathbf{S}}$ and the singular values of the data matrices obtained via the row-wise, column-wise and tube-wise unfolding of $\underline{\mathbf{D}}$—say $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$—will be inferred.

Rearranging Equation (1), it results that $\underline{\mathbf{S}}$ can be expressed as:

$$\underline{\mathbf{S}} = \underline{\mathbf{D}} \times_1 \mathcal{U}^{\mathrm{T}} \times_2 \mathcal{V}^{\mathrm{T}} \times_3 \mathcal{W}^{\mathrm{T}} \tag{A1}$$

From this equivalence, it is easy to prove that:

$$\mathbf{S}_1 = \mathcal{U}^{\mathrm{T}} \mathbf{D}_1 \left(\mathcal{W}^{\mathrm{T}} \otimes \mathcal{V}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{A2}$$

$$\mathbf{S}_2 = \mathcal{V}^{\mathrm{T}} \mathbf{D}_2 \left(\mathcal{U}^{\mathrm{T}} \otimes \mathcal{W}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{A3}$$

$$\mathbf{S}_3 = \mathcal{W}^{\mathrm{T}} \mathbf{D}_3 \left(\mathcal{V}^{\mathrm{T}} \otimes \mathcal{U}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{A4}$$

with $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$ being the two-dimensional arrays yielded by the row-wise, column-wise and tube-wise unfolding of $\underline{\mathbf{S}}$. Given the SVD of $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$:

$$\mathbf{D}_1 = \mathcal{U} \boldsymbol{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}} \tag{A5}$$

$$\mathbf{D}_2 = \mathcal{V} \boldsymbol{\Sigma}_2 \mathbf{V}_2^{\mathrm{T}} \tag{A6}$$

$$\mathbf{D}_3 = \mathcal{W} \boldsymbol{\Sigma}_3 \mathbf{V}_3^{\mathrm{T}} \tag{A7}$$

where $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ carry the right singular vectors of $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$, respectively, then:

$$\mathbf{S}_1 = \mathcal{U}^{\mathrm{T}} \mathcal{U} \boldsymbol{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}} \left(\mathcal{W}^{\mathrm{T}} \otimes \mathcal{V}^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}} \left(\mathcal{W}^{\mathrm{T}} \otimes \mathcal{V}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{A8}$$

$$\mathbf{S}_2 = \mathcal{V}^{\mathrm{T}} \mathcal{V} \boldsymbol{\Sigma}_2 \mathbf{V}_2^{\mathrm{T}} \left(\mathcal{U}^{\mathrm{T}} \otimes \mathcal{W}^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{\Sigma}_2 \mathbf{V}_2^{\mathrm{T}} \left(\mathcal{U}^{\mathrm{T}} \otimes \mathcal{W}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{A9}$$

$$\mathbf{S}_3 = \mathcal{W}^{\mathrm{T}} \mathcal{W} \boldsymbol{\Sigma}_3 \mathbf{V}_3^{\mathrm{T}} \left(\mathcal{V}^{\mathrm{T}} \otimes \mathcal{U}^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{\Sigma}_3 \mathbf{V}_3^{\mathrm{T}} \left(\mathcal{V}^{\mathrm{T}} \otimes \mathcal{U}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{A10}$$