A survey of some tensor analysis techniques for biological systems

Farzane Yahyanejad
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607, USA
farzanehyahyanejad@gmail.com

Réka Albert
Department of Physics
Pennsylvania State University
University Park, PA 16802, USA
ralbert@phys.psu.edu

Bhaskar DasGupta
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607, USA
bdasgup@uic.edu

September 12, 2019

Abstract

Tensors are natural and powerful generalizations of vectors and matrices to higher dimensions and play a *fundamental* role in physics, mathematics and many other areas. Tensor analysis methods can be used to provide the foundations of systematic approaches to distinguish significant higher order correlations among the elements of a complex systems via finding ensembles of a small number of reduced systems that provide a concise and representative summary of these correlations. Since biological systems are complex and often involve multiple types of genomic relationships, tensor analysis methods can be utilized to elucidate these hidden complex relationships. There *is* a pressing need for this, as the interpretation of the results of high-throughput experiments has advanced at a much slower pace than the accumulation of data. In this review article we provide an overview of some tensor analysis methods for biological systems.

1 Introduction

The biological functioning and life of a cellular system is controlled by signaling and energy transfer interactions among its numerous constituents such as proteins, RNAs, DNAs, and other small molecules, and usually involve a cascade of biochemical reactions or other physical interactions among these constituents. An investigation of such interactions is usually done by selecting, implicitly or explicitly, one or more models to characterize the interactions (physical, chemical, or statistical dependencies) between components of the cellular environment. Naturally, the selection of the model depends on several factors such as the level of details desired, the characteristics of the particular interactions studied, and the overall goal of the investigation. Often, biologists describe the model by presenting the interaction data in the form of a diagram (*e.g.*, some type of graph), optionally along with some mathematical formulation of its dynamics. Most mathematical formulation of the dynamics typically assumes that each node in the diagram has an associated (discrete or continuous) "state" variable (representing, for example, concentration of the corresponding protein) that is a function of the time variable t, and describes how the value of this variable at a node ("state" of the node) depends on the state of the nodes interacting with it. Examples of some common models of the above type relevant for our article include *protein-protein interaction* networks represented

by undirected graphs *without* any explicit state variables for nodes and *signal transduction networks* represented by edge-labeled directed graphs *optionally* with state variables for nodes. In contrast, the more general (*ordinary* or *partial*) *differential equation* models omit the network diagram altogether and describe the behavior of the state variables via (ordinary or partial) differential equations. These representations may have inter-dependencies, *e.g.*, under certain technical assumptions one can construct a signal transduction network diagram from the differential equation model (see [17, Chap. 5] for details).

A major drawback of using graph-theoretic tools on a single network diagram lies in ignoring the time or ignoring the *higher-order* correlations of the interactions which may lead to *inaccurate* or *incomplete* analysis. For example, a network diagram only encodes pairwise correlations of node state variables, and thus cannot represent a joint k-way correlation among k state variables for any $k > 2$. If precise equations of time evolutions of state variables are given then we could of course completely ignore the network diagrams and work with the given equations, but then we lose the advantage of employing graph-theoretic tools and instead fall back on analysis techniques which are often *hard* to employ effectively because of difficulties of estimating precise equations and the non-trivial *non-linear* natures of these equations.

In this review article we provide an overview of some tensor analysis methods for biological systems. For this type of analysis, one usually models a given biological system as a k-dimensional matrix $\mathcal{X} \stackrel{\text{def}}{=} [x_{i_1,\ldots,i_k}]$ of size $I_1 \times \cdots \times I_k$ (formally an order k *tensor* $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{X}^{I_1 \times \cdots \times I_k}$, see Section 2) encoding higher-order correlation of a biological system with or without time evolution. Tensor analysis methods have already been successfully used in specific contexts of pathway reconstructions in cellular systems and microarray data integration from several sources [3, 18, 28, 54, 66]. Outside bioinformatics, tensor analysis methods have been very successfully applied to many other application areas such as *neuroscience* [23, 47–49], *psychology* [11, 23, 33, 50] and *chemometrics* [6]. Even though at first glance one is tempted to think that involving more dimensions as compared to vectors or matrices would further complicate the computational aspects of the relevant problems, this is *not* necessarily the case. There are *many* advantages of using higher mode tensors as opposed to matrices and vectors for analysis; see Section 2.2.1 for one such example.

2 Standard concepts and definitions related to tensor analysis

Tensors are natural and powerful generalizations of vectors and matrices to higher dimensions and play a *fundamental* role in physics, mathematics and other areas. In this section we briefly review some standard concepts and definitions associated with tensor analysis; see excellent survey articles or books such as [31, 33, 48] for further information.

2.1 Basic definitions and notations

A tensor $\mathcal{X}^{I_1 \times \cdots \times I_k}$ (or, simply $\mathcal{X}$) of *mode* (also called *order*) k is a k-dimensional array $\mathcal{X}$ of size $I_1 \times \cdots \times I_k$. Thus, a tensor of order 1 is a vector and a tensor of order 2 is a matrix; for $k > 2$ a tensor of order k is also known as a "higher-order" tensor. Following widely used conventions, we will denote tensors, matrices and column vectors by calligraphic uppercase (*e.g.*, $\mathcal{X}$), boldface uppercase (*e.g.*, $\mathbf{X}$) and boldface lowercase (*e.g.*, $\mathbf{x}$) letters, respectively. Individual elements will be denoted by the corresponding (non-bold) lowercase letter with appropriate indices, *e.g.*, $x_{i,j}$ for matrix $\mathbf{X}$. A sequence of vectors/matrices/tensors will be indicated by using parenthesized numbers as superscripts, *e.g.*, $\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots, \mathcal{X}^{(m)}$ will denote a sequence of m tensors. Table 1 succinctly summarizes some tensor operations and related notations. Many softwares such as MATLAB, Mathematica and Maple as well as packages in languages such as FORTRAN and C^{++} provide support for basic and advanced tensor operations. The following definitions are standard:

Simple tensor: A tensor $\mathcal{T}$ of order k is a *simple* tensor if and only if there are vectors $\mathbf{v^{(1)}}, \ldots, \mathbf{v^{(k)}}$ such that $\mathcal{T} = \mathbf{v^{(1)}} \circ \cdots \circ \mathbf{v^{(k)}}$. If the tensor $\mathcal{T}$ represents statistical correlations among k variables then $\mathcal{T}$

is simple exactly when the variables are mutually independent.

Symmetric tensor: A tensor $\mathcal{T}^{n\times\cdots\times n}$ of order k is symmetric if $\mathcal{T}_{i_1,\ldots,i_k} = \mathcal{T}_{\sigma_{i_1},\ldots,\sigma_{i_k}}$ for every permutation $\sigma \stackrel{\text{def}}{=} \{\sigma_{i_1},\ldots,\sigma_{i_k}\}$ of $\{i_1,\ldots,i_k\}$.

In the sequel, we will use $||$ to indicate a suitable tensor norm (*e.g.*, Frobenius norm, spectral norm or some other appropriate tensor norm).

Nomenclature	**Notation**	**Mathematical details**
sum of tensors	$\mathcal{Z} = \mathcal{X} + \mathcal{Y}$	$\forall\, i_1,\ldots,i_k : z_{i_1,\ldots,i_k} = x_{i_1,\ldots,i_k} + y_{i_1,\ldots,i_k}$
tensor-scalar product	$\mathcal{Y} = \alpha\,\mathcal{X}$	$\forall\, i_1,\ldots,i_k : y_{i_1,\ldots,i_k} = \alpha\, y_{i_1,\ldots,i_k}$
tensor inner product	$\langle \mathcal{X}, \mathcal{Y}\rangle$	$\langle \mathcal{X}, \mathcal{Y}\rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_k=1}^{I_k} x_{i_1,\ldots,i_k} y_{i_1,\ldots,i_k}$
Frobenius norm of a tensor	$\parallel \mathcal{X} \parallel_{\mathrm{F}}$	$\parallel \mathcal{X} \parallel_{\mathrm{F}} = \sqrt{\langle \mathcal{X}, \mathcal{X}\rangle}$
spectral norm of a tensor	$\parallel \mathcal{X}^{I_1\times\cdots\times I_k} \parallel_s$	$\sup_{\substack{\mathbf{y^{(1)}}\in\mathbb{R}^{I_1},\ldots,\mathbf{y^{(k)}}\in\mathbb{R}^{I_k} \\ \|\mathbf{y^{(1)}}\|=\cdots=\|\mathbf{y^{(k)}}\|=1}} \sum_{i_1=1}^{I_1} \cdots \sum_{i_k=1}^{I_k} x_{i_1,\ldots,i_k} y^{(1)}_{i_1} \ldots y^{(k)}_{i_k}$
vector outer product	$\mathcal{Y} = \mathbf{x^{(1)}} \circ \cdots \circ \mathbf{x^{(n)}}$	$\forall\, i_1,\ldots,i_n : y_{i_1,\ldots,i_n} = x^{(1)}_{i_1} \ldots x^{(n)}_{i_n}$
mode-r matricization of $\mathcal{X}^{I_1\times\cdots\times I_k}$	$\mathbf{X}^{I_r\times\left(\prod_{j\neq r} I_j\right)}_{(r)}$	$x_{(r)\,i_r,\,1+\sum_{\ell=1,\,\ell\neq r}^{k}(i_\ell-1)\left(\prod_{p=1,\,p\neq r}^{k} I_p\right)} = x_{i_1,\ldots,i_r,\ldots,i_k}$
mode-j product of $\mathcal{X}^{I_1\times\cdots\times I_k}$ and $\mathbf{Y}^{R\times I_j}$	$\mathcal{Z}^{I_1\times\cdots\times I_{j-1}\times R\times I_{j+1}\times\cdots\times I_k} = \mathcal{X}\times_j \mathbf{Y}$	$\mathbf{Z}_{(j)} = \mathbf{Y}\,\mathbf{X}_{(j)}$
κ-rank of matrix $\mathbf{X}$	$\kappa_{\mathbf{X}}$	largest j such that every set of j columns of $\mathbf{X}$ are linearly independent

Table 1: Some basic tensor operations and related notations.

2.2 Basic tensor decomposition methods and corresponding ranks

Philosophically, this step is similar to a type of principal component analysis for matrix data. In other words, we "factor" the input tensor into a combination of simpler tensors (*e.g.*, *rank one* tensors, tensor of *small column rank etc*.). For concreteness, we mention the following two factoring methods, but other factoring methods are also often considered based on specific applications and data types, such as the *multi-linear* SVD *factorizations* [38], *higher-order eigenvalue decompositions* [3, 54], and *Boolean tensor factorizations* [46].

2.2.1 CANDECOMP / PARAFAC (CP) decomposition [11, 23, 30]

Intuitively, this is a generalization to higher-order tensors of the standard SVD (*singular value decomposition*) of an $m \times n$ matrix $\mathbf{A}$ [21]:

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}} = \sum_{j=1}^{\min\{m,n\}} \sigma_j\, \mathbf{u_j} \circ \mathbf{v_j}$$

where σ_j's are the diagonal entries of the diagonal matrix Σ and vectors $\mathbf{u_j}$ and $\mathbf{v_j}$ are the j^{th} columns of the matrices $\mathbf{U}$ and $\mathbf{V}$, respectively. Likewise, in CP decomposition, one expresses the input tensor $\mathcal{X}$ as a sum (with minimum number of terms) of outer products of 1-rank tensors, *i.e.*, one finds $\lambda_r, \mathbf{a}^{(r,1)}, \ldots, \mathbf{a}^{(r,k)}$, for $k = 1, 2, \ldots, R_{\text{CP}}$, that solves the problem:

$$\textit{minimize } R_{\text{CP}} \textit{ subject to } \mathcal{X}^{I_1 \times \cdots \times I_k} = \sum_{r=1}^{R_{\text{CP}}} \lambda_r \, \mathbf{a}^{(r,1)} \circ \cdots \circ \mathbf{a}^{(r,k)} \tag{1}$$

where $\mathbf{a}^{(r,j)}$ is a vector of dimension I_j and λ_r's are scalars to bound the norms of $\mathbf{a}^{(r,j)}$'s (see Fig. 1 for a pictorial illustration). The matrix $\widehat{\mathbf{A}}^{(j)} = \left[\mathbf{a}^{(1,j)} \ldots \mathbf{a}^{(R_{\text{CP}},j)}\right]$ formed the taking $\mathbf{a}^{(1,j)} \ldots \mathbf{a}^{(R_{\text{CP}},j)}$ as columns is called the j^{th} *factor matrix* of this factoring. We call this *minimum* possible value of R_{CP} as the CP-rank of the tensor $\mathcal{X}$ and denote it by $\mathsf{rank}_{\text{CP}}(\mathcal{X})$.

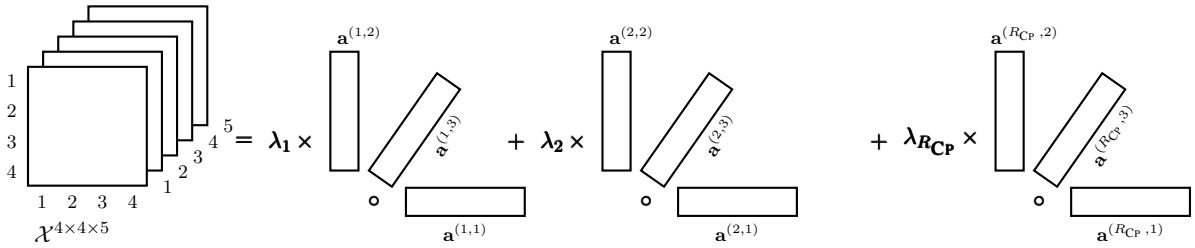


Figure 1: A pictorial representation of the CP decomposition in Equation (1) for $k = 3$.

A special case of CP decomposition of considerable interest to research communities is the *orthogonal tensor decomposition*. An orthogonal decomposition of a *symmetric* tensor $\mathcal{X}^{n\times\cdots\times n}$ of order k is a decomposition

$$\mathcal{X}^{n\times\cdots\times n} = \sum_{r=1}^{R_{\text{CP}}} \lambda_r \underbrace{\mathbf{a}^{(r)} \circ \cdots \circ \mathbf{a}^{(r)}}_{k \text{ times}} \tag{2}$$

such that the vectors $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(R_{\text{CP}})} \in \mathbb{R}^n$ form an orthonormal family of vectors. A tensor is called *orthogonally decomposable* if it has an orthogonal decomposition. Note that for an orthogonal decomposition $R_{\text{CP}} \leq n$.

The following point is worth mentioning about the CP decomposition and the CP-rank. Consider the matrix factoring $\mathbf{A} = \mathbf{X}\,\mathbf{Y}^{\text{T}}$ of a matrix $\mathbf{A}$ that plays a crucial role in many matrix analysis methods. Using the standard SVD of matrix $\mathbf{A}$, we can easily write $\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\text{T}} = \mathbf{X}\,\mathbf{Y}^{\text{T}}$ where $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}$ and $\mathbf{Y} = \mathbf{V}$. However, the factoring is *hardly* unique since we can also take $\mathbf{X} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{Z}$ and $\mathbf{Y} = \mathbf{V}\,\mathbf{Z}$ for *any* orthogonal matrix $\mathbf{Z}$. In contrast, the CP factoring of a higher mode tensor $\mathcal{X}$ *is* unique under less stringent conditions. For example, in [23] it is shown that there *is* any symmetric tensor of order 3 has a *unique* CP decomposition provided we only allows the case when all the vectors in the decomposition are mutually linearly independent, and such a decomposition can in fact be found in polynomial time. The reader is referred to [23, 35, 36] for further results in this direction.

2.2.2 TUCKER decomposition [41, 63–65]

This can be intuitively thought of as a generalization to higher-order tensors of a generic matrix factoring $\mathbf{A} = \mathbf{X}\,\mathbf{Y}^{\text{T}}$ of a matrix $\mathbf{A}$. In TUCKER factoring one expresses $\mathcal{X}$ as a *core* tensor $\mathcal{K}^{J_1\times\cdots\times J_k}$ of *minimal* size

transformed by a matrix $\mathbf{Y}^{(\ell)}$ of size $I_\ell \times J_\ell$ along *each* mode ℓ, *i.e.*, one finds $\mathcal{K}^{J_1 \times \cdots \times J_k}, \mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(k)}$, for arbitrary k that solves the minimization problem:

$$\textit{minimize } \ \mathsf{size}\big(\mathcal{K}^{J_1 \times \cdots \times J_k}\big) \ \ \textit{subject to } \ \mathcal{X} = \mathcal{K} \times_1 \mathbf{Y}^{(1)} \cdots \times_k \mathbf{Y}^{(k)} \tag{3}$$

It is sometimes more convenient to interpret the tensor equality $\mathcal{X} = \mathcal{K} \times_1 \mathbf{Y}^{(1)} \cdots \times_k \mathbf{Y}^{(k)}$ via the following equivalent element-wise equality:

$$x_{i_1,i_2,\ldots,i_n} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_n=1}^{J_n} k_{j_1,j_2,\ldots,j_n}\, y_{i_1,j_1}\, y_{i_2,j_2} \cdots y_{i_n,j_n}$$

In (3), $\mathsf{size}\big(\mathcal{K}^{J_1 \times \cdots \times J_k}\big)$ is a suitable measure of the complexity of the tensor $\mathcal{K}^{J_1 \times \cdots \times J_k}$ (*e.g.*, $\mathsf{size}\big(\mathcal{K}^{J_1 \times \cdots \times J_k}\big) = \big(\prod_{i=1}^{k} J_i\big)^{1/k}$). We call the *minimum* possible value of $\mathsf{size}\big(\mathcal{K}^{J_1 \times \cdots \times J_k}\big)$ as the TUCKER-rank of $\mathcal{X}$ and denote it by $\mathsf{rankTucker}(\mathcal{X})$.

2.2.3 Algorithmic and computational complexity aspects

A well-known algorithmic method to compute tensor decomposition while minimizing the corresponding rank is the *alternating least square* approach (see [11, 23] for CP factoring and [34] for TUCKER factoring), but unfortunately *no* known non-trivial *provable* accuracy guarantees are known for these heuristic approaches (*except* worst-case NP-hardness [24, 26] results which are mostly of theoretical interest only).

Very recently, there has been a a surge in interest in the algorithmic community in in applying the *sum-of-squares* (SOS) approach for special cases of CP decomposition of symmetric tensors to obtain provable guarantees with high probability [7, 8, 19, 27, 43, 56]. The SOS approach is a powerful mathematical technique that deals with determining the emptiness of a given semialgebraic set. Unfortunately, the mathematical details of full generalities of algorithmic applications of SOS appraoch for tensor decomposition is beyond the scope of this review paper, but we give some informal intuition about the approach; see [9, 39] for excellent surveys on the SOS approach and its applications. Consider the CP decomposition framework in Equation (1) for a symmetric "square" tensor (*i.e.*, $I_1 = \cdots = I_k = q$ and $\mathbf{a}^{(r,1)} = \cdots = \mathbf{a}^{(r,k)} = \mathbf{a}^{(r)}$). Using a binary search scheme similar to that used in transforming a linear-programming optimization problem to a problem of determining the feasibility of a system of linear inequalities (see [55, p. 172]), we assume that we know the value of R_{CP} up to any desired accuracy, and thus for some $p \approx R_{\text{CP}}$ we need to compute λ_r's and $\mathbf{a}^{(r)}$'s such that $\frac{1}{p}\mathcal{X} \approx \frac{1}{p}\sum_{r=1}^{p} \mathbf{b}^{(r)^{\circ k}} = \mathcal{M}(k)$, where $\mathbf{b}^{(r)^{\circ k}}$ denotes $(\lambda_r^{1/k}\mathbf{a}^{(r)}) \circ \cdots \circ (\lambda_r^{1/k}\mathbf{a}^{(r)})$ (k times). Note that $\mathcal{M}(k)$ can be thought of as the k^{th}-moment of all possible $\mathbf{b}^{(r)} \in \mathbb{R}^q$ over some unknown distribution $\mathcal{D} : \mathbb{R}^q \mapsto [0, 1]$, and under this same (unknown) distribution $\mathcal{M}(j) = \frac{1}{p}\sum_{r=1}^{p} \mathbf{b}^{(r)^{\circ j}}$ can also be thought of as the j^{th}-moment of all possible $\mathbf{b}^{(r)} \in \mathbb{R}^q$ for any j. Since finding $\mathcal{D}$ is in general NP-hard, the SOS approach pursues the following alternate route:

(I) (lifting moments higher) Select a suitable $\ell \geq k$ and *relax* the distribution $\mathcal{D}$ appropriately to another "almost distribution" $\widetilde{\mathcal{D}} : \mathbb{R}^q \mapsto \mathbb{R}$ such that ***(a)*** $\widetilde{\mathcal{D}}$ can be computed efficiently and ***(b)*** the "j^{th}-moment" under $\widetilde{\mathcal{D}}$ is very close to the j^{th}-moment under $\mathcal{D}$ for all $j = 1, \ldots, \ell$ (in SOS terminology $\widetilde{\mathcal{D}}$ is called a *degree-ℓ pseudo-distribution* [7–9]).

(II) (extracting factors from higher moments) Use a "postprocessing" step to "approximately" infer the $\mathbf{b}^{(r)}$'s from the pseudo-distribution $\widetilde{\mathcal{D}}$.

3 Representing biological systems as tensors

An order k *tensor* $\mathcal{X}^{I_1 \times \cdots \times I_k}$ can easily model higher-order correlation of a biological system of n components with or without time evolution in the following manner:

(tensors for static system) For a static model where time evolution is ignored (such as the ones modelled by *fixed* interaction maps), $I_1 = \cdots = I_k$ and $x_{i_1,\ldots,i_k}$ may denote the joint k-wise correlation value of the ${i_1}^{\text{th}}, \ldots, {i_k}^{\text{th}}$ components of the system. For $k = 2$, this corresponds to a fixed edge-weighted graph in which an undirected edge between two nodes represents a pairwise correlation with the edge weight representing a quantitative estimate of the correlation; for $k > 2$ it properly generalizes such graph-theoretic models.

(tensors for dynamic systems) For a (discrete) time-varying dynamical model, $I_1 = \cdots = I_{k-1}$ the last dimension I_k corresponds to discrete time steps and $x_{i_1,\ldots,i_n}$ denotes the joint $(k-1)$-wise correlation value of the ${i_1}^{\text{th}}, \ldots, {i_{k-1}}^{\text{th}}$ components of the system at time $t = i_n$. For $k = 2$, this corresponds to time-series data models generated by experimental methods such as those using DNA microarrays. For $k > 2$, such a model is popular in representing dynamic social networks of various types in the context of data mining (*e.g.*, see [62]).

Thus, we are led to the following natural questions

- ▷ How do we represent known real biological systems or models as order k tensors for some $k > 2$?
- ▷ How do we generate large number of simulated biological system tensors which are critical in providing statistical validity of tensor analysis methods?

Answers to these questions are discussed in the next few subsections.

3.1 Raw data sources for real biological systems

Interaction maps with node dynamics Curated repositories of published systems biology models include a large number of such systems that are *freely* accessible. For example, the BioModels Database [13] contains 1640 dynamic models (640 are manually curated and encoded in the Systems Biology Markup Language). Discrete dynamic models are available in the Cell Collective [25] and the GINsim model repository [12].

Time Series Data Published research works and large-scale repositories such as ArrayExpress [32] report a large amount of freely-accessible data, generated by various experimental methods (*e.g.*, using DNA microarrays or RNA-seq), in the form of a matrix $\mathbf{Y}^{q \times T}$, where $y_{i,t}$ is the value of the expression level of the i^{th} component (*e.g.*, gene) at the t^{th} time step. For example, Chou *et al.* [14] provide yeast cell cycle time-series gene expression data for a number of time points with relatively small time intervals.

3.2 Generating data for simulated biological systems

We can generate simulated biological systems that *faithfully* reproduce various types of real biological systems using a variety of approaches as discussed next.

Interaction maps with node dynamics To generate simulated tensors for a specific type of biological systems, we will start with the known interaction map (with node dynamics) of a system of the same type (for example, for plant signal transduction system, we may use the light and drought signal transduction system in plants from [42, 61]). For the case of order 3 dynamic tensors, we then use the method in Section 3.3 to generate a tensor $\mathcal{X}^{q \times q \times T}$. Using the mode 3 matricization of $\mathcal{X}$, we can view $\mathcal{X}^{q \times q \times T}$ as a sequence of $q \times q$ interaction/correlation maps, say $\mathcal{Y}_1^{q \times q}, \ldots, \mathcal{Y}_T^{q \times q}$. Independently for each map $\mathcal{Y}_j^{q \times q}$, we can use several known methods, such as the following, to generate a new simulated correlation map:

- ▷ We may generate random maps using the *Markov-chain algorithm* in [29] by repeatedly swapping randomly chosen *compatible* pairs of entries in $\mathcal{Y}_j^{q\times q}$. This approach is widely used in systems biology context, *e.g.*, see the textbook [17] and the references therein.

- ▷ If all the entries in $\mathcal{Y}_1^{q\times q}, \ldots, \mathcal{Y}_T^{q\times q}$ are from the set $\{-1, 0, 1\}$, we can treat each $\mathcal{Y}_j^{q\times q}$ as a graph (with activation and inhibition edge labels) and thus may generate random maps from the degree-distribution of nodes in $\mathcal{Y}_j^{q\times q}$ using the method pioneered by Newman and others in several publications [20, 40, 51–53] that preserves *in expectation* the degree distribution of each node, and label the edges independently randomly with appropriate probabilities as either activation or inhibition such that their percentages match those in $\mathcal{Y}_j^{q\times q}$ in expectation.

Simulated order k dynamic tensors for $k > 3$ can also be generated by a straightforward recursive generalization of the above procedures, *e.g.*, generate higher-order time-varying correlations using the matrix algebra in [44] and then use a *recursive* matricization.

Time-series data One can make use of already existing algorithmic implementations for generating time-series data, such as the software package in [45] that can generate gene regulatory networks with external perturbations from differential equation models.

Finally, as in many simulation methods that use random distribution generators, if necessary the *bias* of our random distributions can be corrected using standard statistical techniques used in statistical data mining [37]. For example, given sample values $x_1, \ldots, x_m$ with average μ and standard deviation σ, one such method is to first calculate the *standardized value* $s_i = (x_i - \mu)/\sigma$ of x_i, then calculate the *standardized range* $\alpha = \max_{1\leq i\leq n}\{s_i\} - \min_{1\leq i\leq n}\{s_i\}$, and finally replace each original x_i by s_i/α.

3.3 Biological systems to tensors

Interaction maps with node dynamics For such biological systems, we can start with an initial choice of states of nodes of the system as follows. If there is a *preferred* choice of initial non-steady state assignments, we can start with this choice (*e.g.*, if our goal is to study *apoptosis* in a disease network starting from expressions of specific genes then our initial choice of states will assign the expression levels of these gene nodes to a higher value and the rest of the nodes to a lower value). Otherwise, we may start with a suitable random choice of initial states of nodes. We then run the system up to a suitably large time step T. Our simulation outputs can be summarized in the form of a matrix $\mathbf{M}^{q\times T}$, where q is the number of nodes whose expression levels are measures over times $t = 1, 2, \ldots, T$ and $m_{i,t}$ is the value of the expression level of the i^{th} gene at the t^{th} time step. From this data, we can, for example, construct an order 3 tensor $\mathcal{X}^{q\times q\times T}$ representing the *second-order* time-evolving correlations among the nodes in the following manner:

- ▷ Let $\mathbf{M}^{(t)}$ be the sub-matrix obtained by taking the first t columns of $\mathbf{M}$. Ensure that the mean of the observed data for each gene in $\mathbf{M}^{(t)}$ is zero by subtracting $\sum_{j=1}^{t} m_{i,j}^{(t)}/t$ from each $m_{i,j}^{(t)}$ for $i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, t$.

- ▷ Compute $\mathbf{Z}^{(t)} = \mathbf{M}^{(t)}\mathbf{M}^{(t)^{\mathrm{T}}}$ and set $x_{i,j,t} = z_{i,j}^{(t)}$.

Higher-order time-varying correlations can also be constructed by generalizing the above approach; for the relevant matrix algebra, see, for example, [44].

Time series data These data types can be handled in the same manner as done for interaction maps with node dynamics, except that we do not need to run any model to generate the time series.

Time snapshots of interaction maps For input data consisting of explicit time snapshots of a given biological system, the corresponding tensor representation is *straightforward*.

4 System analysis via tensor decomposition

There are a few reasons why finding an "exact" tensor decomposition (*cf.* Equations (1)–(3)) could be unrealistic for real or simulated biological tensors. Firstly, as noted before, most tensor decompositions are NP-hard to compute in the worst case [24, 26]. Secondly, the input tensor data may be slightly noisy and therefore a computationally efficient inexact but *almost* accurate tensor decomposition may suffice. Finally, an exact decomposition may involve numbers (such as $\sqrt{2}$) that have *no* finite-precision representations. Thus, in practice, for tensor decompositions one usually uses an "suitably approximate" version of tensor decomposition. For example, for a (sufficiently small) real number $\varepsilon > 0$ and a suitable tensor norm $||$, Equations (1)–(3) can be modified to their approximate versions (1)′–(3)′ as follows:

$$\mathcal{X}^{I_1\times\cdots\times I_k} \approx_\varepsilon \mathcal{A}^{I_1\times\cdots\times I_k} \stackrel{\text{def}}{=} \Big|\Big|\mathcal{X}^{I_1\times\cdots\times I_k} - \sum_{r=1}^{R_{\text{CP}}(\varepsilon)} \lambda_r \underbrace{\mathbf{a}^{(r,1)}\circ\cdots\circ\mathbf{a}^{(r,k)}}_{=\mathcal{A}^{(r)^{I_1\times\cdots\times I_k}}}\Big|\Big| \leq \varepsilon \tag{1}'$$

$$\mathcal{X}^{n\times\cdots\times n} \approx_\varepsilon \mathcal{A}^{n\times\cdots\times n} \stackrel{\text{def}}{=} \Big|\Big|\mathcal{X}^{n\times\cdots\times n} - \sum_{r=1}^{R_{\text{CP}}(\varepsilon)} \lambda_r \underbrace{\mathbf{a}^{(r)}\circ\cdots\circ\mathbf{a}^{(r)}}_{=\mathcal{A}^{(r)^{n\times\cdots\times n}}}\Big|\Big| \leq \varepsilon \tag{2}'$$

$$\Big|\Big|\mathcal{X} - \mathcal{K}(\varepsilon)\times_1 \mathbf{Y}^{(1)}\cdots\times_k \mathbf{Y}^{(k)}\Big|\Big| \leq \varepsilon \tag{3}'$$

Informally, system analysis via tensor decomposition aims to address research questions on extracting an ensemble of a *small* number of *reduced* subsystems (*i.e.*, subsystems that have *less* complexity, such as *fewer* correlations, than the original one) out of a given system to provide a concise and representative summary of the important correlations between components of the system such that non-trivial system analysis methods can operate on these reduced subsystems. For concreteness, our remaining discussion in this section assumes the tensor decomposition as found in Equation (1)′ with $I_1 = \cdots = I_k = n$, but similar discussions hold for other tensor decompositions as well. Assume, without loss of generality, that we have re-scaled and re-indexed the values of λ_r's such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{R_{\text{CP}}(\varepsilon)} > 0$, and $||\mathbf{a}^{(r,j)}||_{\text{F}} = 1$ for all $r = 1, 2, \ldots, R_{\text{CP}}(\varepsilon)$ and $j = 1, 2, \ldots, k$. Let $\lambda'_j = \lambda_j/(\lambda_1 + \lambda_2 + \cdots + \lambda_{R_{\text{CP}}(\varepsilon)})$. Note that the value of $\mathcal{A}^{(r)}_{i_1,\ldots,i_k} \in [-1, 1]$ for any r and any $i_1, \ldots, i_k \in \{1, \ldots, n\}$ can be interpreted as the value of a k-way correlation between the k variables, say $x_{i_1}, \ldots, x_{i_k}$, corresponding to the k indices in the k dimensions, and each λ'_r can be thought of the significance probability of the corresponding r^{th} factor $\mathcal{A}^{(r)}$. Based on such interpretations, one can retrieve significant correlations from the factors in various ways. For example, one such method could be the following.

- ▶ First, select the significant factors in an appropriate way. Some possibilities could be as follows:
  - ▷ For a suitable threshold (real number) $0 < \Theta < 1$, select all factors $\mathcal{A}^{(j)}$ satisfying $\lambda_j \geq \Theta$.
  - ▷ Select factors probabilistically where the j^{th} factor $\mathcal{A}^{(j)}$ is selected with probability λ'_j. The randomized strategy may be more suitable when the λ'_j values are *not* sufficiently spread out (*e.g.*, their standard deviation is small).
  - ▷ In some applications that require strong statistical decoupling, it may be desirable to select factors such that vectors corresponding to the same variable in different factors are mutually linearly independent (or close to being mutually linearly independent). For such situations, strategies such as the following could be used. Recall that $\widehat{\mathbf{A}}^{(j)}$ denotes the j^{th} factor matrix of the tensor $\mathcal{A}$. For a suitable $\tau \in \left[\min_{j=1}^{k}\left\{\kappa_{\widehat{\mathbf{A}}^{(j)}}\right\} \leq \max_{j=1}^{k}\left\{\kappa_{\widehat{\mathbf{A}}^{(j)}}\right\}\right]$, select τ indices, say $\ell_1, \ldots, \ell_\tau$, such that for each j the vectors $\mathbf{a}^{(\ell_1,j)}, \ldots, \mathbf{a}^{(\ell_\tau,j)}$ form a *mutually orthogonal* family

of vectors. We can then select the τ factors $\mathcal{A}^{(\ell_1)}, \ldots, \mathcal{A}^{(\ell_\tau)}$. The selection can be further refined using one or more rules discussed before.

- Once a factor is selected, we can retrieve significant correlations encoded by it by using an appropriate threshold, *i.e.*, for a suitable threshold (real number) $\Phi \in [-1, 1]$, for a selected factor $\mathcal{A}^{(j)}$ consider the correlation among the variables $x_{i_1}, \ldots, x_{i_k}$ as statistically significant if $\mathcal{A}^{(r)}_{i_1,\ldots,i_k} \geq \Phi$. Other more complex application-specific strategies can also be designed.

One could interpret the value of the rank $R_{\text{CP}}(\varepsilon)$ itself as some measure of "complexity" of the input tensor $\mathcal{X}$. For other applications, it is possible to interpret the factors and the components of the factors in different ways (*i.e.*, not necessarily as correlations between variables); for example see [3, 4, 18, 54, 66].

5 Statistical and biological validations of tensor analysis methods

Validations of our tensor analysis can be classified into three categories:

Methodological validation: How do we estimate the "quality" of our tensor decomposition methods?

Statistical validation: How do we compute the "statistical significance" for our tensor analysis results?

Biological validation: Do our tensor decomposition methods recover reported correlations or pathways for known (published) biological systems?

For subsequent discussion purposes, assume that our tensor decomposition at the completion of final steps in Section 4 generated a collection of m significant tensor factors $\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(m)}$ and the combined single tensor $\mathcal{A}' = \sum_{r=1}^{m} \mathcal{A}^{(r)}$ for the input tensor $\mathcal{X}$.

5.1 Methodological validation

The parameter ε in the optimization framework for tensor decomposition (*cf.* Equations (1)–(3)) controls the initial accuracy of the decomposition and varying ε we can obtain decompositions of different accuracies. However, we need also to have control on the accuracy after the reduction steps in Section 4. One possible way to do this is via calculation of the relative error $\xi = ||\mathcal{X} - \mathcal{A}'|| \,/\, ||\mathcal{X}||$.

5.2 Statistical validation

Noise sensitivity and robustness of the decompositions in (1)**–**(3)

We can use the theoretical framework in [10, 22] inspired by the famous smoothed-analysis results in [59, 60]. We illustrate the framework for the rank measure for CP decomposition (Equation (1)); adoptions for other decompositions are very similar. We perturb each vector $\mathbf{a}^{(r,j)}$ suitably[1] to obtain a new vector $\widehat{\mathbf{a}}^{(r,j)}$, build the new tensor $\widehat{\mathcal{A}} = \sum_{r=1}^{R_{\text{CP}}(\varepsilon)} \lambda_r \, \widehat{\mathbf{a}}^{(r,1)} \circ \cdots \circ \widehat{\mathbf{a}}^{(r,k)}$, use the same CP decomposition algorithm on $\widehat{\mathcal{A}}$ to compute the new rank $\widehat{R_{\text{CP}}(\varepsilon)}$, and finally measure the sensitivity by the relative error $(\,|R_{\text{CP}}(\varepsilon) - \widehat{R_{\text{CP}}(\varepsilon)}|\,)/R_{\text{CP}}(\varepsilon)$.

[1] The authors in [10, 22] add independent Gaussian noise to each coordinate of $\mathbf{a}^{(r,j)}$, but other distributions may be used.

***p*-value calculation**

There are many ways to do this; we illustrate one such approach. We can use the method used in [1, 2] to calculate the p-values for the rank or individual significant correlations that we found at the completion of final steps in Section 4. We illustrate the method for the rank. Suppose that we wish to calculate the p-value of a particular evaluation of the rank $R(\mathcal{X})$ of a biological system (tensor) $\mathcal{X}$. We will generate a large number q of simulated systems $\mathcal{X}^{(1)}, \ldots, \mathcal{X}^{(q)}$ of the same type as $\mathcal{X}$ using the Markov-chain algorithm of Section 3.2, compute the corresponding ranks $R(\mathcal{X}^{(1)}, \ldots, R(\mathcal{X}^{(q)})$ of these simulated systems, and then use an appropriate statistical test, such as a (unpaired) one-sample student's t-test, to determine the probability that $R(\mathcal{X})$ can be generated by a distribution that fits the data points $R(\mathcal{X}^{(1)}, \ldots, R(\mathcal{X}^{(q)})$.

5.3 Biological validation

Here we determine how close our tensor analysis is in preserving important properties of a *known* system. These validations are *conceptually straightforward* and can check a variety of properties of a known biological system in its corresponding reduced system. For example:

- We can check the known presence or absence of a significant correlation of a known (published) system in most significant tensors factors produced after Section 4. Based on the number of true positives, false positives, true negatives and false negatives, we may compute the four standard metrics *true positive rate*, *false positive rate*, *accuracy rate* and *precision* to assess the validity of our methods. For fine tuning a specific parameter of our method, such as the ε parameter in Equation (1), we may use the ROC (*receiver operating characteristic*) plot over the ranges of ε.
- For dynamical tensors, we may check if our significant tensors produce a system that preserve known significant dynamical properties (such as *limit cycles*, *attractors*, *monotonicity*, *controllability* and *observability*) of a known (published) system.

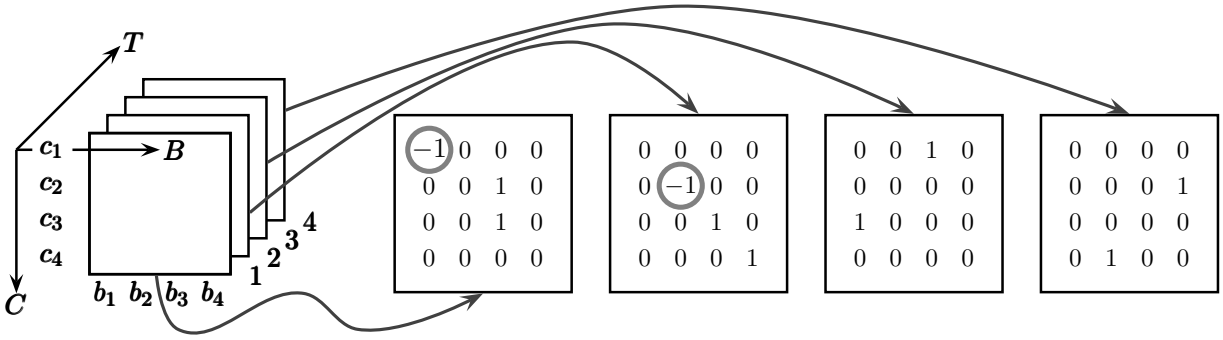
As a concrete illustration, let $\mathcal{X}$ be a order 3 tensor generated from the Boolean dynamic model for the *guard cell ABA signaling* as obtained in [42] using the method described in Section 3.3, and suppose that our CP decomposition together with the optimizations in Section 4 produces q most significant tensors $\mathcal{A}^{(1)} = \mathbf{a}^{(1,1)} \circ \mathbf{a}^{(1,2)} \circ \mathbf{a}^{(1,3)}, \ldots, \mathcal{A}^{(q)} = \mathbf{a}^{(q,1)} \circ \mathbf{a}^{(q,2)} \circ \mathbf{a}^{(q,3)}$ for $\mathcal{X}$. The double and triple knockout experiments in [42] suggest that the states of CaIM (Ca^{2+} influx through the plasma membrane) and AnionEM (anion efflux at the plasma membrane) are jointly correlated to the state of Closure (ABA signalling). Assuming this to be the ground truth, we may check if this correlation also exists in at least one of $\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(q)}$.

6 Specific illustrations of tensor analysis

In this section we provide two specific illustrations of tensor analysis for biological systems. The first one is an artificial toy example. The second illustration summarizes tensor analysis results for a specific research result.

6.1 A toy example of tensor analysis insights of an artificial biological system

Consider the following mode 3 tensor $\mathcal{X} \stackrel{\text{def}}{=} \mathcal{X}^{\mathbf{t}\times\mathbf{b}\times\mathbf{c}}$ that describes time-evolving *cross-correlations* between two sets of four components $\mathbf{b} = (b_1, b_2, b_3, b_4)$ and $\mathbf{c} = (c_1, c_2, c_3, c_4)$ of a biological system over discrete time $\mathbf{t} = (1, 2, 3, 4)$ (modified from an example in [15, 16]):



Consider the following CP decomposition with appropriate normalization of the inter-modal factors (*c.f.* Section 2.2.1):

$$\begin{array}{cccc} \lambda_1 & \mathbf{t}^{(1)} & \mathbf{b}^{(1)} & \mathbf{c}^{(1)} \\ 2\left[\begin{pmatrix}-1\\0\\0\\0\end{pmatrix}\circ\begin{pmatrix}-1\\0\\0\\0\end{pmatrix}\circ\begin{pmatrix}-1\\0\\0\\0\end{pmatrix}\right] & + & \frac{1}{\sqrt{2}}\begin{pmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\\0\end{pmatrix}\circ\begin{pmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\\0\end{pmatrix}\circ\begin{pmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\\0\end{pmatrix} & + \frac{1}{\sqrt{2}}\begin{pmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\\0\end{pmatrix}\circ\begin{pmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\\0\end{pmatrix}\circ\begin{pmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\\0\end{pmatrix} \\ \lambda_2 \quad \mathbf{t}^{(2)} \quad \mathbf{b}^{(2)} \quad \mathbf{c}^{(2)} & & \lambda_3 \quad \mathbf{t}^{(3)} \quad \mathbf{b}^{(3)} \quad \mathbf{c}^{(3)} & \end{array}$$

$$\begin{array}{cccc} & & & \\ +2\left[\begin{pmatrix}0\\-1\\0\\0\end{pmatrix}\circ\begin{pmatrix}0\\-1\\0\\0\end{pmatrix}\circ\begin{pmatrix}0\\-1\\0\\0\end{pmatrix}\right] & + & \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1/\sqrt{2}\\0\\1/\sqrt{2}\end{pmatrix}\circ\begin{pmatrix}0\\1/\sqrt{2}\\0\\1/\sqrt{2}\end{pmatrix}\circ\begin{pmatrix}0\\1/\sqrt{2}\\0\\1/\sqrt{2}\end{pmatrix} & + \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1/\sqrt{2}\\0\\-1/\sqrt{2}\end{pmatrix}\circ\begin{pmatrix}0\\1/\sqrt{2}\\0\\-1/\sqrt{2}\end{pmatrix}\circ\begin{pmatrix}0\\1/\sqrt{2}\\0\\-1/\sqrt{2}\end{pmatrix} \\ \lambda_4 \quad \mathbf{t}^{(4)} \quad \mathbf{b}^{(4)} \quad \mathbf{c}^{(4)} & & \lambda_5 \quad \mathbf{t}^{(5)} \quad \mathbf{b}^{(5)} \quad \mathbf{c}^{(5)} & \lambda_6 \quad \mathbf{t}^{(6)} \quad \mathbf{b}^{(6)} \quad \mathbf{c}^{(6)} \end{array}$$

Assuming that this tensor decomposition has been statistically validated, a *simplest* predictive algorithm will identify $\begin{pmatrix}-1\\0\\0\\0\end{pmatrix}\circ\begin{pmatrix}-1\\0\\0\\0\end{pmatrix}\circ\begin{pmatrix}-1\\0\\0\\0\end{pmatrix}$ and $\begin{pmatrix}0\\-1\\0\\0\end{pmatrix}\circ\begin{pmatrix}0\\-1\\0\\0\end{pmatrix}\circ\begin{pmatrix}0\\-1\\0\\0\end{pmatrix}$ (shown by dotted boxes) as the two most significant factors, and they can be traced back to the correlations between b_1 and c_1 at $t = 1$ and between b_2 and c_2 at $t = 2$ (circled in gray color).

6.2 A review of specific tensor analysis in two research papers

In this section, we briefly review the tensor analysis methods and the biological conclusions drawn therefrom in the two research papers [3, 54]. The application of tensor decomposition methods in these two papers to yeast (*S. cerevisiae*) time course expression data illustrates the biological insights that can be gained from this type of analysis. All the measurements were made in *S. cerevisiae* cell cultures under the influence of the pheromone α-factor (to synchronize their cell cycles).

Alter and Golub [3] use matrix eigenvalue decomposition to decompose a matrix of pairwise gene correlations into rank-1 sub-matrices. They then construct a tensor that integrates information about gene expression and the binding of select transcription factors to the promoter regions to each gene. By applying tensor higher order eigenvalue decomposition, they identify decorrelated rank-1 sub-networks that can be associated with *independent* biological pathways. They apply this methodology to a time-course of mRNA expression data for more than 4000 *S. cerevisiae* genes, integrated with binding data on 12 cell cycle related transcription factors and 12 developmental transcription factors. The analysis uncovers *three* significant sub-networks that capture 40%, 15%, and 9%, respectively, of the expression correlation among genes. The first subnetwork is associated with the α-factor signal transduction pathway, and expresses the correlations among genes that are up-regulated (or down-regulated, respectively) in response to pheromone. The second and third sub-networks are associated with the two known pathways of antipodal (opposite) cell cycle expression oscillations. The coupling between the first sub-network and the second sub-network expresses the exit from pheromone-induced cell cycle arrest and entry into cell cycle progression.

Omberg *et al.* [54] use higher-order singular value decomposition to decompose a data tensor of genes versus experimental settings into rank-1 sub-tensors. They apply this methodology to mRNA expression data for more than 4000 *S. cerevisiae* genes, measured over 13 time points, for three conditions, namely oxidative stress due to exposure to hydrogen peroxide or menadione, respectively, and oxidative-stress-free control. They find that the significant sub-tensors represent independent biological programs. The first and most significant sub-tensor, which captures 70% of the expression information, represents the steady state of mRNA expression in response to hydrogen peroxide, menadione, or α-factor, averaged over time and conditions. Three sub-tensors that follow in significance (explaining 1% to 6% of the information) represent change in expression in response to oxidative stress. The three following sub-tensors, each explaining around 1% of the expression information, represent pheromone responses and pheromone-induced oxidative stress responses. The three following sub-tensors (explaining 0.6% to 0.9% of the information) reflect the differences in the responses to hydrogen peroxide and to menadione.

7 Conclusion

In this article we have reviewed some basic aspects of powerful tensor analysis methods that provide the foundations of *systematic* approaches to determine significant higher order correlations among elements of biological systems by finding ensembles of small number of reduced systems that provide a concise and representative summary of these correlations.

Admittedly, a short review article such as this one can cover *only* some aspects of tensor analysis of biological systems, leaving other aspects in the references. For example, following are some of the topics that are not covered in this article but may be of significance to some researchers:

Learning models via tensor decompositions: References such as [4, 5] discuss algorithms for learning hidden Markov models or other types of models using tensor decompositions.

Eigenvalues and eigenvectors of tensors: Eigenvalues and eigenvectors of matrices (*i.e.*, tensors of order 2) play a crucial role in spectral analysis of network algorithms and processes, and in principal component analysis for matrix data. One can extend the these definitions to higher-order tensors (*e.g.*, see [57]) but the full potential of these generalizations to higher-order tensor analysis is still not clear.

Although tensor analysis methods have been used in specific contexts of computational biology before, their usage in bioinformatics is *not* as widespread as matrix-based or linear algebraic methods. In our opinion, there is a *pressing* need for more research on applying tensor analysis methods for biological systems, as interpretations of results of high-throughput experiments have advanced at a much slower pace than the data accumulation. The state of the art in gene expression data analysis is still to focus on a *small* group of *key* genes (*e.g.*, those that are most highly correlated, or most differentially expressed when comparing two contexts) and *discard* the rest of the information. This is partly because of the computational complexity of all types of follow-up analyses, and partly because of the noise and uncertainty affecting all biological measurements. Tensor-based analysis provides a *principled* way of using all available information to achieve a clearer understanding. In addition, by identifying reduced systems, a *smaller* set of *most-supported* correlations naturally emerges that are optimal for all follow-up analyses *without* making potentially limiting assumptions. The tensor analysis framework does *not* have the limitations and specific requirements of methods, such as the bi-clustering approach [58], that are used to determine clusters of genes correlated over a set of conditions but not in conditions outside of this set. The tensor framework is also naturally suited to incorporate and study the effect of *additional* variables (*e.g.*, variable *environmental* influences), thus allowing *integrated* studies of genetic and environmental factors. It is our hope that this article will catalyze and motivate further research in the fascinating inter-disciplinary interplay between biology and tensor analysis methods.

Acknowledgements

B.D. and F.Y. was supported by NSF grant IIS-1814931. R.A. was supported by NSF grant IIS-1814405.

References

[1] R. Albert, B. DasGupta, A. Gitter, G. Gürsoy, R. Hegde, P. Pal, G. S. Sivanathan and E. D. Sontag. *A New Computationally Efficient Measure of Topological Redundancy of Biological and Social Networks*, Physical Review E, 84(3), 036117, 2011.

[2] R. Albert, B. DasGupta and N. Mobasheri. *Topological implications of negative curvature for biological and social networks*, Physical Review E, 89(3), 032811, 2014.

[3] O. Alter and G. H. Golub. *Reconstructing the pathways of a cellular system from genome-scale signals by using matrix and tensor computations*, PNAS, 102(49), 17559-17564, 2005.

[4] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade and M. Telgarsky. *Tensor Decompositions for Learning Latent Variable Models*, Journal of Machine Learning Research, 15(1), 2773-2832, 2014.

[5] A. Anandkumar, D. Hsu, and S. M. Kakade. *A method of moments for mixture models and hidden markov models*, Conference on Learning Theory, 33.1-33.34, 2012.

[6] C. J. Appellof and E. R. Davidson. *Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents*, Analytical Chemistry, 53, 2053-2056, 1981.

[7] B. Barak, J. A. Kelner and D. Steurer. *Dictionary learning and tensor decomposition via the sum-of-squares method*, 47^{th} ACM Symposium on Theory of Computing, 143-151, 2015.

[8] B. Barak and A. Moitra. *Noisy tensor completion via the sum-of-squares hierarchy*, Proceedings of Machine Learning Research, 49, 417-445, 2016.

[9] B. Barak and D. Steurer. *Sum-of-squares proofs and the quest toward optimal algorithms*, proceedings of International Congress of Mathematicians, 2014.

[10] A. Bhaskara, M. Charikar, A. Moitra and A. Vijayaraghavan. *Smoothed Analysis of Tensor Decompositions*, 46^{th} Annual ACM Symposium on Theory of Computing, 594-603, 2014.

[11] 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, 283-319, 1970.

[12] C. Chaouiya, A. Naldi and D. Thieffry. *Logical Modelling of Gene Regulatory Networks with GINsim*, Methods in molecular biology, 804, 463-79, 2012.

[13] V. Chelliah, N. Juty, I. Ajmera, R. Ali, M. Dumousseau, M. Glont, M. Hucka, G. Jalowicki, S. Keating, V. Knight-Schrijver, A. Lloret-Villas, K. Nath Natarajan, J.-B. Pettit, N. Rodriguez, M. Schubert, S. M. Wimalaratne, Y. Zhao, H. Hermjakob,N. Le Novère and C. Laibe. *BioModels: ten-year anniversary*, Nucleic Acids Research, 43, D542-D548, 2015.

[14] R. J. Chou, M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. J. Lockhart and R. W. Davis. *A genome-wide transcriptional analysis of the mitotic cell cycle*, Molecular Cell, 2, 65-73, 1998.

[15] P. Comon. *Canonical Tensor Decompositions*, Research report ISRN I3S/RR-2004-17-FR, 2004.

[16] P. Comon and B. Mourrain. *Decomposition of quantics in sums of powers of linear forms*, Signal Processing, 53(2-3), 93-107, 1996.

[17] B. DasGupta and J. Liang. *Models and Algorithms for Biomolecules and Molecular Networks,* John Wiley and Sons, Inc., 2016.

[18] M. Dyrby, D. Baunsgaard, R. Bro and S. B. Engelsen. *Multiway chemometric analysis of the metabolic response to toxins monitored by nmr*, Chemom Intell Lab Systems, 76, 79-89, 2005.

[19] R. Ge and T. Ma. *Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms*, 18^{th} International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, 829-849, 2015.

[20] M. Girvan and M. E. J. Newman. *Community structure in social and biological networks*, Proceedings of the National Academy of Sciences, 99, 7821-7826, 2002.

[21] G. H. Golub and C. F. Van Loan. *Matrix Computations*, 3^{rd} edition, The Johns Hopkins University Press, 1996.

[22] N. Goyal, S. Vempala and Y. Xiao. *Fourier PCA and robust tensor decomposition*, 46^{th} Annual ACM Symposium on Theory of Computing, 584-593, 2014.

[23] R. A. Harshman. *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis*, UCLA Working Papers in Phonetics, 16, 1-84, 1970.

[24] J. Håstad. *Tensor rank is NP-complete*, Journal of Algorithms, 11, 644-654, 1990.

[25] T. Helikar, B. Kowal, S. McClenathan, M. Bruckner, T. Rowley, A. Madrahimov, B. Wicks, M. Shrestha, K. Limbu and J. A. Rogers. *The Cell Collective: Toward an open and collaborative approach to systems biology*, BMC Systems Biology, 6, 96, 2012.

[26] C. J. Hillar and L.-H. Lim. *Most tensor problems are* NP*-hard*, Journal of the ACM, 60 (6), 45:1–45:38, 2013.

[27] S. B. Hopkins, J. Shi and D. Steurer. *Tensor principal component analysis via sum-of-square proofs*, Proceedings of Machine Learning Research, 40, 956-1006, 2015.

[28] V. Hore, A. Viñuela, A. Buil, J. Knight, M. I McCarthy, K. Small and J. Marchini. *Tensor decompositions for multiple tissue gene expresion experiments*, Nature Genetics, 48, 1094-1100, 2016.

[29] R. Kannan, P. Tetali and S. Vempala. *Markov-chain algorithms for generating bipartite graphs and tournaments*, Random Structures and Algorithms, 14, 293-308, 1999.

[30] H. A. L. Kiers. *Towards a standardized notation and terminology in multiway analysis*, Journal of Chemometrics, 14, 105-122, 2000.

[31] T. G. Kolda and B. W. Bader. *Tensor decompositions and applications*, SIAM Review, 51(3), 455-500, 2009.

[32] N. Kolesnikov, E. Hastings, M. Keays, O. Melnichuk, Y. A. Tang, E. Williams, M. Dylag, N. Kurbatova, M. Brandizi, T. Burdett, H. Megy, E. Pilicheva, G. Rustici, A. Tikhonov, H. Parkinson, R. Petryszak, U. Sarkans and A. Brazma. *ArrayExpress update–simplifying data submissions*, Nucleic Acids Research, 43, D1113-D1116, 2015.

[33] P. M. Kroonenberg. *Applied Multiway Data Analysis*, John Wiley & Sons, 2008.

[34] P. M. Kroonenberg and J. De Leeuw. *Principal component analysis of three-mode data by means of alternating least squares algorithms*, Psychometrika, 45, 69-97, 1980.

[35] J. B. Kruskal. *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Applications, 18, 95-138, 1977.

[36] J. B. Kruskal. *Rank, decomposition, and uniqueness for 3-way and N-way arrays*, in Multiway Data Analysis, R. Coppi and S. Bolasco (Eds.), 7-18, North-Holland, Amsterdam, 1989.

[37] R. J. Larsen and M. L. Marx. *An Introduction to Mathematical Statistics and Its Applications*, Third Edition, 2000.

[38] L. D. Lathauwer, B. D. Moor and J. Vandewalle. *A Multilinear Singular Value Decomposition*, SIAM Journal of Matrix Analysis and Applications, 21(4), 1253-1278.

[39] M. Laurent. *Sums of Squares, Moment Matrices and Optimization Over Polynomials*, in M. Putinar and S. Sullivant (eds), Emerging Applications of Algebraic Geometry, The IMA Volumes in Mathematics and its Applications, 149, 157-270, 2009.

[40] E. A. Leicht and M. E. J. Newman. *Community Structure in Directed Networks*, Physical Review Letters, 100, 118703, 2008.

[41] J. Levin. *Three-Mode Factor Analysis*, Ph.D. thesis, University of Illinois, Urbana, 1963.

[42] S. Li, S. M. Assmann and R. Albert. *Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling*, PLoS Biology, 4(10), e312, 2006.

[43] T. Ma, J. Shi and D. Steurer. *Polynomial-Time Tensor Decompositions with Sum-of-Squares*, 57^{th} Annual IEEE Symposium on Foundations of Computer Science, 438-446, 2016.

[44] E. Meijer. *Matrix algebra for higher order moments*, Linear Algebra and its Applications, 410, 112-134, 2005.

[45] P. Mendes. *Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3*, Trends in Biochemical Science, 22, 361-363, 1997.

[46] P. Miettinen. *Boolean Tensor Factorizations*, 11^{th} IEEE International Conference on Data Mining, 447-456, 2011.

[47] F. Miwakeichi, E. Martnez-Montes, P. A. Valds-Sosa, N. Nishiyama, H. Mizuharam and Y. Yamaguchi. *Decomposing eeg data into space time frequency components using parallel factor analysis*, Neuroimage, 22, 1035-1045, 2004.

[48] M. Mørup. *Applications of tensor (multiway array) factorizations and decompositions in data mining*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 24-40, John Wiley & Sons, Inc., 2011.

[49] M. Mørup, L. Hansen and S. M. Arnfred. *Erpwavelab a toolbox for multi-channel analysis of time-frequency transformed event related potentials*, Journal of Neuroscience Methods, 161, 361-368, 2007.

[50] T. Murakami and P. M. Kroonenberg. *Three-mode models and individual differences in semantic differential data*, Multivariate Behavioral Research, 38, 247-283, 2003.

[51] M. E. J. Newman. *Detecting community structure in networks*, European Physics Journal B, 38, 321-330, 2004.

[52] M. E. J. Newman. *The structure and function of complex networks*, SIAM Review, 45, 167-256, 2003.

[53] M. E. J. Newman and M. Girvan. *Finding and evaluating community structure in networks*, Physical Review E, 69, 026113, 2004.

[54] L. Omberg, G. H. Golub and O. Alter. *A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies*, PNAS, 104(47), 18371-18376, 2007.

[55] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*, Dover Publications, 1988.

[56] A. Potechin and D. Steurer. *Exact tensor completion with sum-of-squares*, Proceedings of Machine Learning Research, 65, 1-54, 2017.

[57] L. Qi. *Eigenvalues and invariants of tensors*, Journal of Mathematical Analysis and Applications, 325, 1363-1377, 2007.

[58] D. J. Reiss, N. S. Baliga and R. Bonneau. *Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks*, BMC Bioinformatics, 7, 1-22, 2006.

[59] D. A. Spielman and S. H. Teng. *Smoothed Analysis of Algorithms: Why the Simplex Algorithm usually takes Polynomial Time*, Journal of the ACM, 385-463, 2004.

[60] D. A. Spielman and S. H. Teng. *Smoothed Analysis: An Attempt to Explain the Behavior of Algorithms in Practice*, Communications of the ACM, 76-84, 2009.

[61] Z. Sun, X. Jin, R. Albert and S. M. Assmann. *Multi-level Modeling of Light-Induced Stomatal Opening Offers New Insights into Its Regulation by Drought*, PLOS Computational Biology, 10: e1003930, 2014.

[62] J. Sun, D. Tao and C. Faloutsos. *Beyond Streams and Graphs: Dynamic Tensor Analysis*, 12^{th} ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 374-383, 2006.

[63] L. R. Tucker. *Implications of factor analysis of three-way matrices for measurement of change*, in Problems in Measuring Change, C. W. Harris (ed.), University of Wisconsin Press, 1963, 122-137, 1963.

[64] L. R. Tucker. *The extension of factor analysis to three-dimensional matrices*, in Contributions to Mathematical Psychology, H. Gulliksen and N. Frederiksen (eds.), Holt, Rinehardt & Winston, New York, 110-127, 1964.

[65] L. R. Tucker. *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31, 279-311, 1966.

[66] B. Yener, E. Acar, P. Aguis, K. Bennett, S. L. Vandenberg and G. E Plopper. *Multiway modeling and analysis in stem cell systems biology*, BMC Systems Biology, 2:63, 2008.