Equations for Multilinear Generalization of the General Linear Model :

Factor/component and cluster analysis examine "internal relationships"-- relations within a single dataset -- while the GLM and its various special cases (which include most standard statistical methods) examine "external relationships" -- relations between two datasets (one of which might be a logical or 'design' matrix). Many *internal* analysis methods have been generalized to higher-order multilinearity and are used to analyze three-(or higher)-way data arrays. Some of these generalized methods have been attracting interest because they possess significantly stronger properties in certain applications. As a result, they are now being incorporated into a growing number of such applications in a widening range of disciplines.[*] My talk develops similar higher-order generalizations of *external* analysis. In particular, it offers multilinear generalizations of the GLM that give it enhanced capabilities similar to those that are proving useful in internal analysis.

Generalization at Level 1: Canonical weights become multilinear (for data that are higher-order arrays)
At Level 1, the information-source objects (data and/or design matrices) and their associated canonical weights are extended to a higher number of 'ways'. This can be done on both sides of the relation or just on one side. Even when restricted to one side, this is sometimes enough to give the solution stronger properties. Thus, in many of the equations given here, only the left side of the canonical relation is explicitly given. The other side may, or may not, have a similar form. (Of course, both sides must evaluate to canonical objects of the same shape.)

Generalization at Level 2: Canonical variates become multilinear (the tensor products of optimal 'canonical factors')
Canonical variates are the objects created to have maximum correlation across sides of the relation. For level two, they are extended from linear to multilinear functions of the data, i.e., from weighted linear combinations of data columns to outer or Kronecker products of several weighted linear combinations of data fibers. Each variate still represents a pattern (of covariation of **X** elements) that is extracted on a given side to most resemble a similar pattern extracted on the other side. However, since they are Kronecker products, they are 'vectors' with multilinear structure (repeating proportional sub-patterns) and so they identify patterns generated by sources that influence variation in (or interact with and hence are modulated by) several 'ways' or modes at once. This requires ${}^{1}\mathbf{X}$ and ${}^{2}\mathbf{X}$ to have more than one matching 'way'.

Generalization at Level 3: Canonical tensors become 'hybridized' (i.e., incorporate patterns from multiple data sources in one tensor)
At level 3, the canonical factors for a given canonical variate (tensor) need not all be formed from fibers in the same data array. The variates can be tensor products of patterns extracted from fibers in several data sources, or even sums of such patterns. Level 3 is generally beyond the scope of the talk but a few examples are provided.

This addendum provides the equations to accompany the concepts discussed in my talk. In these equations, $\otimes$ denotes the Kronecker product, and $\odot$ denotes the Khatri-Rao (or KRB =Khatri-Rao-Bro) *column-wise* Kronecker product[†]. The symbol δ represents the (generalized) Kronecker delta[‡], here used mostly to eliminate cross-factor products in PARACCON models, where a factor in Mode A only interacts with 'itself' in Modes B and C. This also requires that R=S (and =T,U, etc.).

Although these generalizations can incorporate, and be applied to, higher-way canonical/data objects of all orders, the examples given here are kept simple, and usually increase the order of (multi)linearity by only one step—from two way to three-way data and from one-way to two-way canonical variates.

Algorithms have been developed that have successfully performed some of these generalized canonical analyses. An example is shown during the talk. However, it is important to note that important research questions concerning these generalizations are still under active study. The foremost of these is the question of how best to define optimality for different classes of problems, particularly when the canonical factors have unique axis positions in their respective canonical spaces.

---

[*] to see some of these, perform a Google search for "parafac" or "Tucker T3"

[†] The KR(B) product is defined as $\mathbf{A} \odot \mathbf{B} = \left[\mathbf{a}_1 \otimes \mathbf{b}_1 \;\middle|\; \mathbf{a}_2 \otimes \mathbf{b}_2 \;\middle|\; \cdots \;\middle|\; \mathbf{a}_R \otimes \mathbf{b}_R\right]$ where **A** and **B** are matrices with the same number of columns.

[‡] δ has zero elements except when all indices are equal, in which case the elements are 1. Delta tensors are isotropic, so their indices can be raised or lowered freely.

Addendum to a talk presented to the York Univ. Dept. of Mathematics and Statistics on Nov.24,2006, by R. A. Harshman, U.W.O. Dept. of Psych.

Some relevant existing models, written using three basic notations

Scalar notation | *Matrix notation* | *Tensor notation*

1. FA/PCA (Factor Analysis / Principal Components Analysis) a bilinear model

$$x_{ij} \approx \sum_{r=1}^{R} a_{ir}\, b_{jr} \qquad \mathbf{X} \approx \mathbf{A}\mathbf{B}' \qquad X^{IJ} \approx A^{IR} B^{JS} \delta_{RS}$$

2. Parafac (PARAllel FACtor analysis) simply FA/PCA with added level(s) of multilinearity [trilinear case shown]

$$x_{ijk} \approx \sum_{r=1}^{R} a_{ir}\, b_{jr}\, c_{kr} \qquad \mathbf{X}_k \approx \mathbf{A}\,\mathbf{D}_k\,\mathbf{B}' \qquad X^{IJK} \approx A^{IR} B^{JS} C^{KT} \delta_{RST}$$

where $\mathbf{D}_k = diag([c_{k1}\ c_{k2} \cdots c_{kR}])$

3. Tucker's T3 (Tucker-3Mode, or Multi--Mode Factor Analysis), another "higher-way PCA" that includes factor 'interactions'

$$x_{ijk} \approx \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T} a_{ir}\, b_{js}\, c_{kt}\, g_{rst} \qquad \mathbf{X}_k \approx \mathbf{A}\left(\sum_{t=1}^{T} c_{kt}\mathbf{G}_t\right)\mathbf{B}' \qquad X^{IJK} \approx A^{IR} B^{JS} C^{KT} G_{RST}$$

4. Canonical Correlation & GLM (standard linear model)

(for 1 canonical variate)

$$\sum_{j=1}^{J} {}^{1}x_{ij}\, {}^{1}w_{j} = {}^{1}y_{i} \approx {}^{2}y_{i} = \sum_{j'=1}^{J'} {}^{2}x_{ij'}\, {}^{2}w_{j'} \qquad {}^{1}\mathbf{X}\,{}^{1}\mathbf{w} = {}^{1}\mathbf{y} \approx {}^{2}\mathbf{y} = {}^{2}\mathbf{X}\,{}^{2}\mathbf{w} \qquad {}^{1}X^{\mathrm{IJ}}\,{}^{1}W_{\mathrm{J}} = {}^{1}Y^{\mathrm{I}} \approx {}^{2}Y^{\mathrm{I}} = {}^{2}X^{\mathrm{IJ'}}\,{}^{2}W_{\mathrm{J'}}$$

(for R canonical variates)

$$\sum_{j=1}^{J} {}^{1}x_{ij}\, {}^{1}w_{jr} = {}^{1}y_{ir} \approx {}^{2}y_{ir} = \sum_{j'=1}^{J'} {}^{2}x_{ij'}\, {}^{2}w_{j'r} \qquad \left({}^{1}\mathbf{X}\right)\left({}^{1}\mathbf{W}\right) = \left({}^{1}\mathbf{Y}\right) \approx \left({}^{2}\mathbf{Y}\right) = \left({}^{2}\mathbf{X}\right)\left({}^{2}\mathbf{W}\right) \qquad {}^{1}X^{\mathrm{IJ}}\,{}^{1}W_{\mathrm{J}}^{\mathrm{R}} = {}^{1}Y^{\mathrm{IR}} \approx {}^{2}Y^{\mathrm{IR}} = {}^{2}X^{\mathrm{IJ'}}\,{}^{2}W_{\mathrm{J'}}^{\mathrm{R}}$$

(In general, $J \neq J'$ although in particular cases they may be equal.)

Level 1 Generalization of GLM: Multilinear canonical weights and three-way data sources

TUCCON at L1 (Tensor/Tucker Unrestricted Canonical CorrelatiON)

Raw Product form -- no linear recombinations of basis vectors; produces *S*x*T* canonical variates (flattened to 1 index in matrix version)

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\left({}^{1}x_{ijk}\right)\left({}^{1}w_{js}\;{}^{1}w_{kt}\right)=\left({}^{1}y_{ist}\right)\approx\cdots \qquad \underset{I\ \mathrm{x}\ JK}{\left(\left[{}^{1}\mathbf{X}_1 \mid {}^{1}\mathbf{X}_2 \mid \cdots \mid {}^{1}\mathbf{X}_K\right]\right)}\underset{JK\ \mathrm{x}\ ST}{\left(\left({}^{1\mathrm{C}}\mathbf{W}\otimes{}^{1\mathrm{B}}\mathbf{W}\right)\right)}=\underset{I\,\mathrm{x}\,ST}{\left({}^{1}\mathbf{Y}\right)}\approx\cdots \qquad \left({}^{1}X^{IJK}\right)\left({}^{1}W_J^{S}\;{}^{1}W_K^{T}\right)=\left({}^{1}Y^{IST}\right)\approx\cdots$$

General weighted form -- incorporates G (an $S\times T\times R$ array, unfolded to $ST\times R$ for matrix version) and reindexes components

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{s=1}^{S}\sum_{t=1}^{T}\left({}^{1}x_{ijk}\right)\left({}^{1}w_{js}\;{}^{1}w_{kt}\;{}^{1}g_{str}\right)=\left({}^{1}y_{ir}\right)\approx\cdots \qquad \underset{I\ \mathrm{x}\ JK}{\left(\left[{}^{1}\mathbf{X}_1 \mid {}^{1}\mathbf{X}_2 \mid \cdots \mid {}^{1}\mathbf{X}_K\right]\right)}\left(\underset{JK\ \mathrm{x}\ ST}{\left({}^{1\mathrm{C}}\mathbf{W}\otimes{}^{1\mathrm{B}}\mathbf{W}\right)}\underset{ST\,\mathrm{x}\,R}{{}^{1}\mathbf{G}}\right)=\underset{I\,\mathrm{x}\,R}{\left({}^{1}\mathbf{Y}\right)}\approx\cdots \qquad \left({}^{1}X^{IJK}\right)\left({}^{1}W_J^{S}\;{}^{1}W_K^{T}G_{ST}^{R}\right)=\left({}^{1}Y^{IR}\right)\approx\cdots$$

PARACCON at L1 (PARallel Canonical CorrelatiON) incorporates only 'fully distinct' outer-products (this is accomplished by using the Khatri-Rao product in matrix notation and the Kronecker delta in tensor notation).

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\left({}^{1}x_{ijk}\right)\left({}^{1}w_{jr}\;{}^{1}w_{kr}\right)=\left({}^{1}y_{ir}\right)\approx\cdots \qquad \underset{I\ \mathrm{x}\ JK}{\left(\left[{}^{1}\mathbf{X}_1 \mid {}^{1}\mathbf{X}_2 \mid \cdots \mid {}^{1}\mathbf{X}_K\right]\right)}\underset{JK\ \mathrm{x}\ R}{\left(\left({}^{1\mathrm{C}}\mathbf{W}\odot{}^{1\mathrm{B}}\mathbf{W}\right)\right)}=\underset{I\,\mathrm{x}\,R}{\left({}^{1}\mathbf{Y}\right)}\approx\cdots \qquad \left({}^{1}X^{IJK}\right)\left({}^{1}W_J^{S}\;{}^{1}W_K^{T}\;\delta_{ST}^{R}\right)=\left({}^{1}Y^{IR}\right)\approx\cdots$$

Level 2 generalization: Multilinear canonical variates (canonical tensors)

TUCCON at Level 2

$$\sum_{r=1}^{R} {}^{1}a_{is}\,{}^{1}b_{jt} = {}^{1}y_{(ij)(st)} \approx {}^{2}y_{(ij)(st)} = \sum_{r=1}^{R} {}^{2}a_{is}\,{}^{2}b_{jt}$$

$${}^{1}\mathbf{A} \otimes {}^{1}\mathbf{B} = {}^{1}\mathbf{Y} \approx {}^{2}\mathbf{Y} = {}^{2}\mathbf{A} \otimes {}^{2}\mathbf{B}$$

$${}^{1}A^{IR}\;{}^{1}B^{JS} = {}^{1}Y^{(IJ)(RS)} \approx {}^{2}Y^{(IJ)(RS)} = {}^{2}A^{IR}\;{}^{2}B^{JS}$$

where

$${}^{1}a_{ir} = {}^{1}a_{i(st)} = \sum_{j=1}^{J}\sum_{k=1}^{K} {}^{1}x_{ijk}\;{}^{1}w_{js}\;{}^{1}w_{kt}$$

where

$$\underset{I\,x\,ST}{{}^{1}\mathbf{A}} = \left(\underset{I\quad x\quad JK}{\left[\,{}^{1}\mathbf{X}_1 \mid {}^{1}\mathbf{X}_2 \mid \cdots \mid {}^{1}\mathbf{X}_K\,\right]}\right)\left(\underset{JK\quad x\quad ST}{\left({}^{1C}\mathbf{W} \otimes {}^{1B}\mathbf{W}\right)}\right)$$

where

$${}^{1}A^{IR} = {}^{1}X^{IJK}\;{}^{1}W_J^{S}\;{}^{1}W_K^{T}\;\delta_{ST}^{R}$$

and

$${}^{1}b_{js} = {}^{1}b_{s(jt)} = \sum_{i=1}^{I}\sum_{k=1}^{K} {}^{1}x_{ijk}\;{}^{1}w_{is}\;{}^{1}w_{kt}$$

and

$$\underset{J\,x(RT)}{{}^{1}\mathbf{B}} = \left(\underset{J\quad x\quad IK}{\left[\,{}^{1}\mathbf{X}'_1 \mid {}^{1}\mathbf{X}'_2 \mid \cdots \mid {}^{1}\mathbf{X}'_K\,\right]}\right)\left(\underset{IK\quad x\quad RT}{\left({}^{1C}\mathbf{W} \otimes {}^{1A}\mathbf{W}\right)}\right)$$

and

$${}^{1}B^{IS} = {}^{1}X^{IJK}\;{}^{1}W_I^{R}\;{}^{1}W_K^{T}\;\delta_{RT}^{S}$$

There are other versions of Level 2 TUCCON that differ in the amount to which the subscripts of the products are combined into fewer indices, and how this is done. The goal here is to give the general idea, so not all these alternatives and their rationales are covered here.

PARACCON Level 2 [an example with three-way data and two-way canonical variates])

$$\sum_{r=1}^{R} {}^{1}a_{ir}\,{}^{1}b_{jr} = {}^{1}y_{(ij)r} \approx {}^{2}y_{(ij)r} = \sum_{r=1}^{R} {}^{2}a_{ir}\,{}^{2}b_{jr}$$

$${}^{1}\mathbf{A} \odot {}^{1}\mathbf{B} = {}^{1}\mathbf{Y} \approx {}^{2}\mathbf{Y} = {}^{2}\mathbf{A} \odot {}^{2}\mathbf{B}$$

$${}^{1}A^{IR}\;{}^{1}B^{JS}\delta_{RS}^{U} = {}^{1}Y^{(IJ)U} \approx {}^{2}Y^{(IJ)U} = {}^{2}A^{IR}\;{}^{2}B^{JS}\delta_{RS}^{U}$$

where

$${}^{1}a_{ir} = \sum_{j=1}^{J}\sum_{k=1}^{K} {}^{1}x_{ijk}\;{}^{1}w_{jr}\;{}^{1}w_{kr}$$

where

$$\underset{I\,x\,R}{{}^{1}\mathbf{A}} = \left(\underset{I\quad x\quad JK}{\left[\,{}^{1}\mathbf{X}_1 \mid {}^{1}\mathbf{X}_2 \mid \cdots \mid {}^{1}\mathbf{X}_K\,\right]}\right)\left(\underset{JK\quad x\quad R}{\left({}^{1C}\mathbf{W} \odot {}^{1B}\mathbf{W}\right)}\right)$$

where

$${}^{1}A^{IR} = {}^{1}X^{IJK}\;{}^{1}W_J^{S}\;{}^{1}W_K^{T}\;\delta_{ST}^{R}$$

and

$${}^{1}b_{jr} = \sum_{i=1}^{I}\sum_{k=1}^{K} {}^{1}x_{ijk}\;{}^{1}w_{ir}\;{}^{1}w_{kr}$$

and

$$\underset{J\,x(RT)}{{}^{1}\mathbf{B}} = \left(\underset{J\quad x\quad IK}{\left[\,{}^{1}\mathbf{X}'_1 \mid {}^{1}\mathbf{X}'_2 \mid \cdots \mid {}^{1}\mathbf{X}'_K\,\right]}\right)\left(\underset{IK\quad x\quad S}{\left({}^{1C}\mathbf{W} \odot {}^{1A}\mathbf{W}\right)}\right)$$

and

$${}^{1}B^{JS} = {}^{1}X^{IJK}\;{}^{1}W_I^{R}\;{}^{1}W_K^{T}\;\delta_{RT}^{S}$$

Addendum to a talk presented to the York Univ. Dept. of Mathematics and Statistics on Nov.24,2006, by R. A. Harshman, U.W.O. Dept. of Psych.

Notes on Level 2

It is apparent by comparison of Level 1 and 2 equations that the Level 2 canonical *factors* have the same internal structure as the Level 1 canonical *variates* (e.g., ${}^{1}A^{IR}$ and ${}^{1}Y^{IR}$ are written the same way). However, the numerical values of their elements will differ, because their weights are chosen so that they optimize correlations of different things. At Level 1, the c-variates themselves (e.g., columns of ${}^{1}Y_R^I$ and ${}^{2}Y_R^I$) are what must be optimally correlated, so the weights are selected to accomplish this. At Level 2, the rank-1 patterns (e.g., matrices ${}^{1}Y_r^{IJ}$, ${}^{2}Y_r^{IJ}$), when vectorized, are what must be optimally correlated, and this is accomplished by finding optimal weights for generating the c-factors (e.g., columns of ${}^{1}A^{IR}$, ${}^{1}B^{JS}$, ${}^{2}A^{IR}$,and ${}^{2}B^{JS}$) so their *AB products* are optimally correlated.

These models can, of course, also be written explicitly in terms of the original source objects and weights. In the case of PARACCON this is straightforward:
In scalar product notation, it is

$$\left(\sum_{j=1}^{J}\sum_{k=1}^{K} {}^{1}x_{ijk}\ {}^{1}w_{jr}\ {}^{1}w_{kr}\right)\left(\sum_{i=1}^{I}\sum_{k=1}^{K} {}^{1}x_{ijk}\ {}^{1}w_{ir}\ {}^{1}w_{kr}\right) = {}^{1}y_{ijr} \approx {}^{2}y_{ijr} = \left(\sum_{j=1}^{J}\sum_{k=1}^{K} {}^{2}x_{ijk}\ {}^{2}w_{jr}\ {}^{2}w_{kr}\right)\left(\sum_{i=1}^{I}\sum_{k=1}^{K} {}^{2}x_{ijk}\ {}^{2}w_{ir}\ {}^{2}w_{kr}\right),$$

in matrix notation it is

$$\left(\left[\,{}^{1}\mathbf{X}_1 \mid {}^{1}\mathbf{X}_2 \mid \cdots \mid {}^{1}\mathbf{X}_K\right] \left({}^{1C}\mathbf{W} \odot {}^{1B}\mathbf{W}\right)\right) \odot \left(\left[\,{}^{1}\mathbf{X}'_1 \mid {}^{1}\mathbf{X}'_2 \mid \cdots \mid {}^{1}\mathbf{X}'_J\right] \left({}^{1C}\mathbf{W} \odot {}^{1A}\mathbf{W}\right)\right) = {}^{1}\mathbf{Y} \approx \cdots$$

and in tensor notation:

$$\left({}^{1}X^{IJK}\ {}^{1}W_J^S\ {}^{1}W_K^T\ \delta_{ST}^R \right) \otimes \left({}^{1}X^{IJK}\ {}^{1}W_I^R\ {}^{1}W_K^T\ \delta_{RT}^S \right) \delta_{RS}^U = {}^{1}Y^{IJU} \approx {}^{2}Y^{IJU} = \left({}^{2}X^{IJK}\ {}^{2}W_J^S\ {}^{2}W_K^T\ \delta_{ST}^R \right) \otimes \left({}^{2}X^{IJK}\ {}^{2}W_I^R\ {}^{2}W_K^T\ \delta_{RT}^S \right) \delta_{RS}^U \quad .$$

However, in the case of TUCCON at Level 2, there are several choices to make concerning how to combine or vectorize the subscripts generated (there can be up to four modes of the components generated unless some combination and vectorization is done). There are also intermediate models that define the canonical factors as PARACCON would, but then use them to create canonical tensors using unrestricted tensor (Kronecker) products. For example, here are some alternative TUCCON models:

$$\left({}^{1}X^{IJK}\ {}^{1}W_J^S\ {}^{1}W_K^T\ \delta_{ST}^R \right) \otimes \left({}^{1}X^{IJK}\ {}^{1}W_I^R\ {}^{1}W_K^T\ \delta_{RT}^S \right) = {}^{1}Y^{IJ,RS} \approx {}^{2}Y^{IJ,RS} = \left({}^{2}X^{IJK}\ {}^{2}W_J^S\ {}^{2}W_K^T\ \delta_{ST}^R \right) \otimes \left({}^{2}X^{IJK}\ {}^{2}W_I^R\ {}^{2}W_K^T\ \delta_{RT}^S \right)$$

or, instead, as

$$\left({}^{1}X^{IJK}\ {}^{1}W_J^S\ {}^{1}W_K^{T'} \right) \otimes \left({}^{1}X^{IJK}\ {}^{1}W_I^R\ {}^{1}W_K^{T''} \right) \delta_{T'T''}^T = {}^{1}Y^{IJ,RST} \approx {}^{2}Y^{IJU} = \left({}^{2}X^{IJK}\ {}^{2}W_J^S\ {}^{2}W_K^{T'} \right) \otimes \left({}^{2}X^{IJK}\ {}^{2}W_I^R\ {}^{2}W_K^{T''} \right) \delta_{T'T''}^T \ .$$

These issues would take us too far beyond the introductory scope of this addendum, and so will not be discussed here.

Level 3 generalization: Hybrid (multi-source) canonical variates

In all these models, canonical factors **A** , **B**, etc. are defined in terms of their roles in the canonical variates, regardless of whether they are weighted composites of fibers from Mode A, B or C etc. of the data object from which they were formed. The data objects are simply regarded as sources of linear or multilinear *patterns* which could reside in any mode or combination of modes.

1. **Example 1**: A simple example would be the use of two different three-way arrays (e.g., ${}^{1}\overset{*}{\mathbf{X}}$ and ${}^{1}\overline{\mathbf{X}}$) to get the Mode A and Mode B canonical factors for the left side of the relation:

$$ {}^{1}\overset{*}{\mathbf{A}} \odot {}^{1}\overline{\mathbf{B}} = {}^{1}\overset{*-}{\mathbf{Y}} \approx \cdots $$

or, written explicitly,

$$ \left(\left[{}^{1}\overset{*}{\mathbf{X}}_1 \,\middle|\, {}^{1}\overset{*}{\mathbf{X}}_2 \,\middle|\, \cdots \,\middle|\, {}^{1}\overset{*}{\mathbf{X}}_K \right] \left({}^{1C}\mathbf{W} \odot {}^{1B}\mathbf{W} \right) \right) \odot \left(\left[{}^{1}\overline{\mathbf{X}}_1 \,\middle|\, {}^{1}\overline{\mathbf{X}}_2 \,\middle|\, \cdots \,\middle|\, {}^{1}\overline{\mathbf{X}}_J \right] \left({}^{1C}\mathbf{W} \odot {}^{1B}\mathbf{W} \right) \right) = {}^{1}\mathbf{Y} \approx \cdots $$

2. **Example 2:** Another example would be to use two different *two-way* arrays:

$$ \left({}^{1}\dot{\mathbf{X}} \;\; {}^{1}\dot{\mathbf{W}} \right) \odot \left({}^{1}\tilde{\mathbf{X}} \;\; {}^{1}\tilde{\mathbf{W}} \right) = {}^{1}\dot{\mathbf{A}} \odot {}^{1}\tilde{\mathbf{B}} = {}^{2}\overset{\bullet\sim}{\mathbf{Y}} \approx \cdots $$

3. **Example 3:** Finally, to give an idea of the wide range of pattern-interaction exploration strategies possible, here is a case (in tensor notation) in which the Mode A canonical factors are obtained by contracting three modes of a four way array while the Mode B canonical factors are obtained by combining patterns from a three-way array and a two way array:

$$ {}^{1}A^{IR} \; {}^{1}B^{JS} \; \delta^{U}_{RS} = {}^{1}Y^{IJU} \approx \cdots $$

where

$$ {}^{1}A^{IR} = {}^{1}X^{IJKL} \; {}^{1}W^{S}_{J} \; {}^{1}W^{T}_{K} \; {}^{1}W^{U}_{L} \; \delta^{R}_{STU} $$

$$ {}^{1}B^{JS} = {}^{1B.1}X^{IJK} \; {}^{1B.1}W^{R}_{I} \; {}^{1B.1}W^{T}_{K} \; \delta^{S}_{RT} + {}^{1B.2}X^{I'J} \; {}^{1B.2}W^{R}_{I'} \; \delta^{S}_{R} \quad . $$

Level 3 was only very briefly mentioned in the talk and further details would take us beyond the scope of this addendum.