Coupled Matrix and Tensor Factorizations: Models, Algorithms & Computational Issues

Evrim Acar

University of Copenhagen

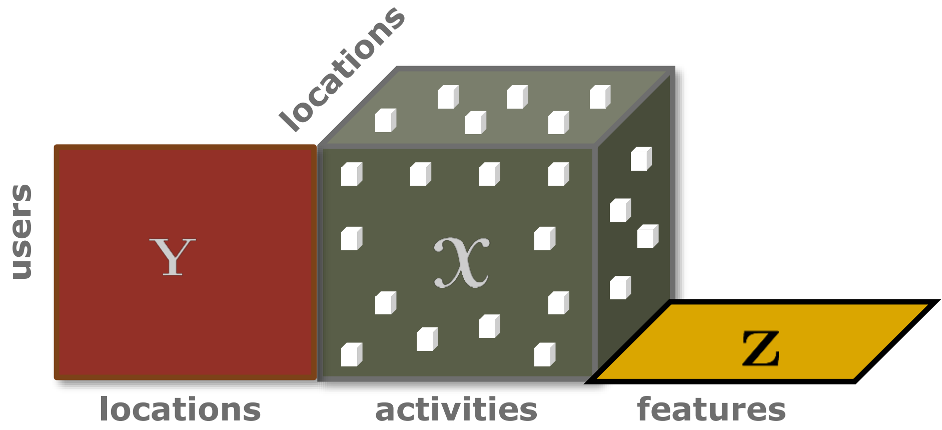
SIAM Conference on Parallel Processing for Scientific Computing

April 14, 2016

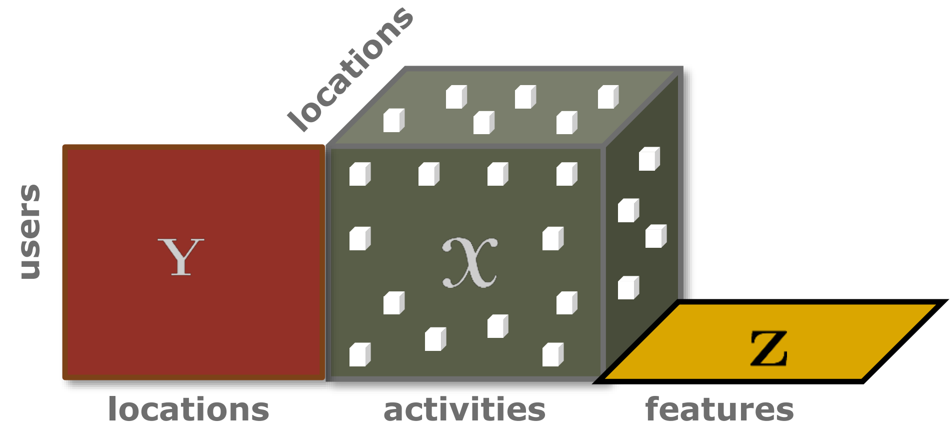
Data Fusion Applications of Interest:

Recommender Systems
Goal: To recommend users activities that they may be interested in doing at various locations

→ **Missing Data Estimation**



Data Fusion Applications of Interest:



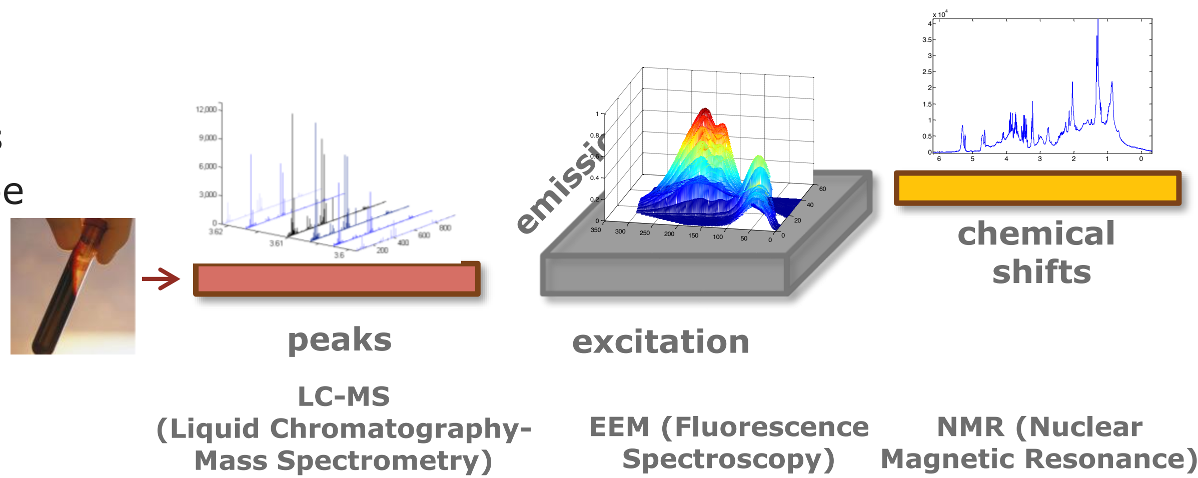
Recommender Systems

Goal: To recommend users activities that they may be interested in doing at various locations

→ **Missing Data Estimation**

Metabolomics

Goal: To capture underlying patterns uniquely so that these patterns can be used for biomarker discovery.

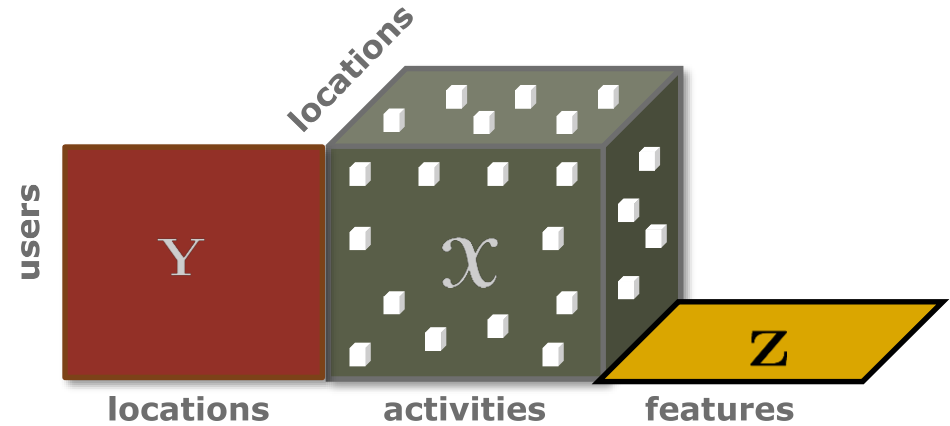


Data Fusion Applications of Interest:

Recommender Systems

Goal: To recommend users activities that they may be interested in doing at various locations

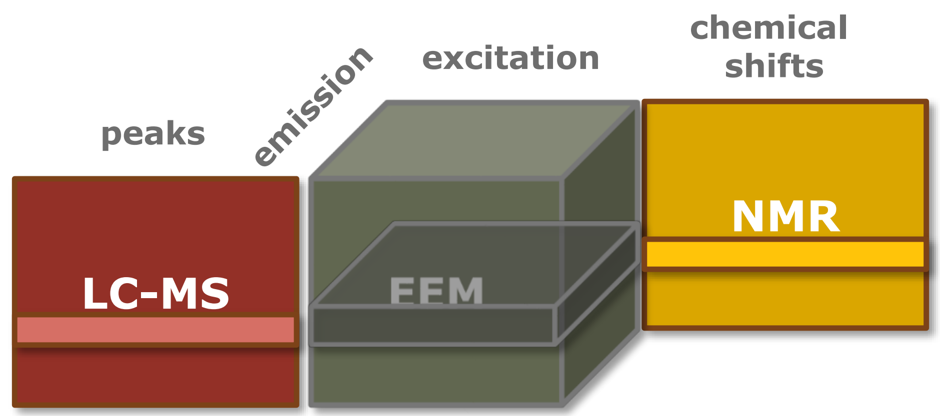
→ **Missing Data Estimation**



Metabolomics

Goal: To capture underlying patterns uniquely so that these patterns can be used for biomarker discovery.

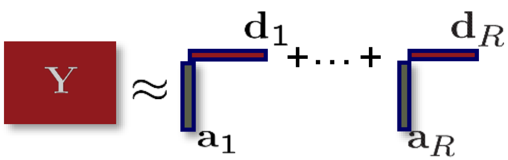
→ **Capturing underlying structures accurately and uniquely**



Coupled Matrix and Tensor Factorizations (CMTF)

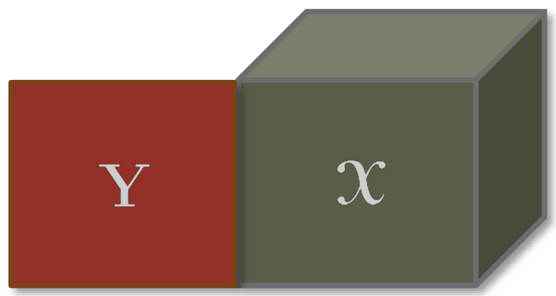
Joint analysis of heterogeneous data from multiple sources can be formulated as a coupled matrix and tensor factorization problem. In CMTF, higher-order tensors and matrices are simultaneously factorized by fitting a CP model to higher-order tensors and factorizing matrices in a coupled manner.

Matrix Factorization:



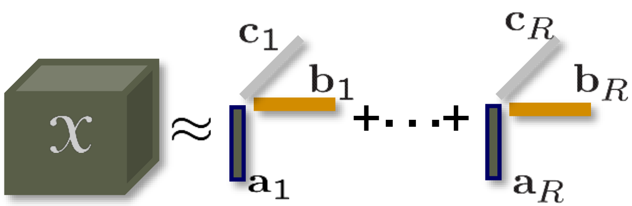
$$\mathbf{Y} \approx \sum_{r=1}^{R} \mathbf{a}_r \mathbf{d}_r^\top$$
$$\approx \mathbf{A}\mathbf{D}^\top$$

$$\mathbf{A} \in \mathbb{R}^{I \times R} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix}$$
$$\mathbf{D} \in \mathbb{R}^{M \times R} = \begin{bmatrix} \mathbf{d}_1 & \cdots & \mathbf{d}_R \end{bmatrix}$$



$$\mathbf{Y} \approx \mathbf{A}\mathbf{D}^\top \quad \mathcal{X} \approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

Tensor Factorization:
CANDECOMP/PARAFAC(CP)
[Harshman, 1970; Carroll & Chang, 1970]



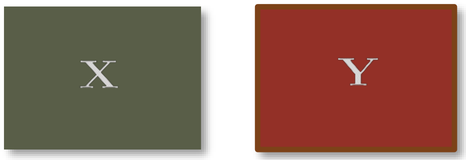
$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$
$$\approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\mathbf{A} \in \mathbb{R}^{I \times R} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix}$$
$$\mathbf{B} \in \mathbb{R}^{J \times R} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix}$$
$$\mathbf{C} \in \mathbb{R}^{K \times R} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

The problem can be formulated as:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\mathbf{D}^\top \|^2$$

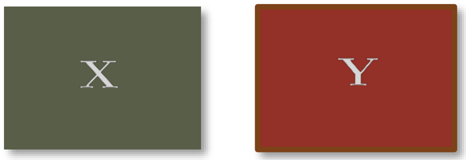
Data Fusion based on Coupled Factorizations



$$\min_{\mathbf{U},\mathbf{V},\mathbf{W}} \left\| \mathbf{X} - \mathbf{U}\mathbf{V}^{\mathsf{T}} \right\|^2 + \left\| \mathbf{Y} - \mathbf{U}\mathbf{W}^{\mathsf{T}} \right\|^2$$

- **Psychometrics:** Simultaneous factorization of Gramian matrices **[Levin, 1966];** Simultaneous component analysis **[Kiers and Ten Berge, 1994]**
- **Chemometrics:** Principal Component Analysis of multiple matrices **[Westerhuis et al., 1998];** Augmented Multivariate Curve Resolution **[De Juan and Tauler, 2000]**
- **Bioinformatics:** Comparing genome-scale expression data from multiple organisms **[Alter et al., 2003]**; Clustering microarray data **[Badea, 2007]**; Simultaneous component analysis with rotation to common and distinct components **[Van Deun et al., 2012];** Decomposing multiple matrices into terms explaining joint and individual variation **[Lock et al., 2013]**
- **Signal Processing:** Joint diagonalization of multiple matrices **[Yeredor, 2002; Ziehe et al., 2004]**; Audio source separation **[Yoo et al., 2010]**; Joint independent component analysis **[Calhoun et al., 2006]**
- **Data Mining:** Collective matrix factorization **[Singh and Gordon, 2008]**; Clustering multi-type relational data **[Long et al., 2006]**; Social recommendation **[Ma et al., 2008]**; Coupled matrix factorization with sparse factors **[Van Deun et al., 2011; Acar et al., 2012]**; Nonnegative shared subspace learning **[Gupta et al., 2010]**; Bayesian interbattery factor analysis **[Klami et al., 2013]**

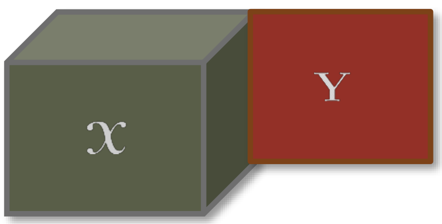
Data Fusion based on Coupled Factorizations



$$\min_{\mathbf{U},\mathbf{V},\mathbf{W}} \| \mathbf{X} - \mathbf{U}\mathbf{V}^{\mathsf{T}} \|^2 + \| \mathbf{Y} - \mathbf{U}\mathbf{W}^{\mathsf{T}} \|^2$$

psychometrics, chemometrics, bioinformatics, signal processing, data mining, …

Cannot handle joint analysis of matrices and higher-order tensors!



$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\mathbf{D}^{\mathsf{T}} \|^2$$

- **Psychometrics:** Linked-mode PARAFAC **[Harshman and Lundy, 1984]**
- **Chemometrics:** Multi-way Multi-block component models **[Smilde et al., 2000]**
- **Bioinformatics:** Coupled analysis of in vitro and histology tissue samples **[Acar et al., 2012]**
- **Signal Processing:** Joint analysis of a covariance matrix and a cumulant tensor **[De Lathauwer and Vandewalle, 2004; Comon, 2004]**; Generalized Coupled Tensor Factorizations **[Yilmaz et al., 2011]**; Structured Data Fusion **[Sorber et al., 2015]**
- **Data Mining:** Multi-way Clustering **[Banerjee et al., 2007]**; Community detection **[Lin et al., 2009]**; Missing value estimation **[Zheng et al., 2010]**; Link prediction **[Ermis et al., 2012]**; Scalable CMTF approaches (sampling–based **[Papalexakis et al., 2014]**, distributed stochastic gradient running on MapReduce **[Beutel et al., 2014]**, distributed ALS running on MapReduce **[Jeon et al., 2016]**)

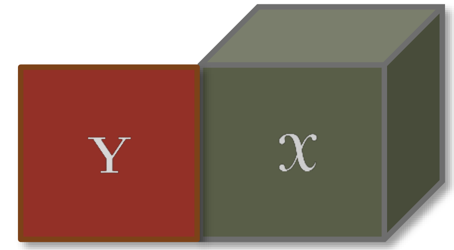
Assume that all components are shared!

All-at-once Optimization for CMTF

[Acar, Dunlavy and Kolda, *KDD Workshop MLG*, 2011]

CMTF-OPT is a gradient–based optimization approach for joint factorization of coupled matrices and higher-order tensors.

$$\min_{\mathrm{A,B,C,D}} \| \mathcal{X} - [\![\mathrm{A,B,C}]\!] \|^2 + \| \mathrm{Y} - \mathrm{AD}^\mathsf{T} \|^2$$



$$\mathrm{Y} \approx \mathrm{AD}^\mathsf{T} \quad \mathcal{X} \approx [\![\mathrm{A,B,C}]\!]$$

Step1: Define the objective function

$$f(\mathrm{A,B,C,D}) = \frac{1}{2} \| \mathcal{X} - [\![\mathrm{A,B,C}]\!] \|^2 + \frac{1}{2} \| \mathrm{Y} - \mathrm{AD}^\mathsf{T} \|^2$$

Step2: Compute the gradient

$$\frac{\partial f}{\partial \mathrm{A}} = -\mathrm{X}_{(1)}(\mathrm{C} \odot \mathrm{B}) + \mathrm{A}(\mathrm{C}^\mathsf{T}\mathrm{C} * \mathrm{B}^\mathsf{T}\mathrm{B}) - \mathrm{YD} + \mathrm{AD}^\mathsf{T}\mathrm{D}$$

$$\frac{\partial f}{\partial \mathrm{B}} = -\mathrm{X}_{(2)}(\mathrm{C} \odot \mathrm{A}) + \mathrm{B}(\mathrm{C}^\mathsf{T}\mathrm{C} * \mathrm{A}^\mathsf{T}\mathrm{A})$$

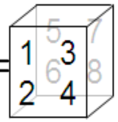
$$\frac{\partial f}{\partial \mathrm{C}} = -\mathrm{X}_{(3)}(\mathrm{B} \odot \mathrm{A}) + \mathrm{C}(\mathrm{B}^\mathsf{T}\mathrm{B} * \mathrm{A}^\mathsf{T}\mathrm{A})$$

$$\frac{\partial f}{\partial \mathrm{D}} = -\mathrm{Y}^\mathsf{T}\mathrm{A} + \mathrm{DA}^\mathsf{T}\mathrm{A}$$

Vectorize and concatenate the partials

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial \mathrm{a}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathrm{a}_R} \\ \frac{\partial f}{\partial \mathrm{b}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathrm{b}_R} \\ \frac{\partial f}{\partial \mathrm{c}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathrm{c}_R} \\ \frac{\partial f}{\partial \mathrm{d}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathrm{d}_R} \end{bmatrix}$$

Matricization

$\mathcal{X} =$ 

$$\mathrm{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

Khatri-Rao Product

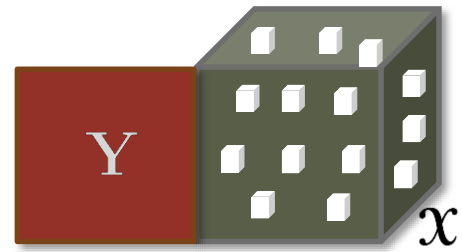
$$\mathrm{A} \odot \mathrm{B} = \begin{bmatrix} \mathrm{a}_1 \otimes \mathrm{b}_1 & \cdots & \mathrm{a}_R \otimes \mathrm{b}_R \end{bmatrix}$$

Step3: Pick a first-order optimization method

e.g., Nonlinear Conjugate Gradient (NCG) and Limited-memory BFGS (L-BFGS) from **Poblano Toolbox** [Dunlavy, Kolda and Acar, 2010]

CMTF easily extends to incomplete data sets

We fit the model only to the known data entries and ignore the missing entries (in higher-order tensors and/or matrices)



$$\mathbf{Y} \approx \mathbf{A}\mathbf{D}^\mathsf{T} \quad \mathcal{X} \approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \| \mathcal{W} * (\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]) \|^2 + \left\| \mathbf{Y} - \mathbf{A}\mathbf{D}^\mathsf{T} \right\|^2$$

$$w_{ijk} = \begin{cases} 1 & \text{if } x_{ijk} \text{ is known,} \\ 0 & \text{if } x_{ijk} \text{ is missing.} \end{cases}$$

Our objective:

$$f_{\mathbf{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = \frac{1}{2} \| \mathcal{W} * (\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]) \|^2 + \frac{1}{2} \left\| \mathbf{Y} - \mathbf{A}\mathbf{D}^\mathsf{T} \right\|^2$$

Gradient: Let $\mathcal{Z} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$

$$\frac{\partial f_{\mathbf{W}}}{\partial \mathbf{A}} = (\mathbf{W}_{(1)} * \mathbf{Z}_{(1)} - \mathbf{W}_{(1)} * \mathbf{X}_{(1)})(\mathbf{C} \odot \mathbf{B}) - \mathbf{Y}\mathbf{D} + \mathbf{A}\mathbf{D}^\mathsf{T}\mathbf{D}$$

$$\frac{\partial f_{\mathbf{W}}}{\partial \mathbf{B}} = (\mathbf{W}_{(2)} * \mathbf{Z}_{(2)} - \mathbf{W}_{(2)} * \mathbf{X}_{(2)})(\mathbf{C} \odot \mathbf{A})$$

$$\frac{\partial f_{\mathbf{W}}}{\partial \mathbf{C}} = (\mathbf{W}_{(3)} * \mathbf{Z}_{(3)} - \mathbf{W}_{(3)} * \mathbf{X}_{(3)})(\mathbf{B} \odot \mathbf{A})$$

$$\frac{\partial f_{\mathbf{W}}}{\partial \mathbf{D}} = -\mathbf{Y}^\mathsf{T}\mathbf{A} + \mathbf{D}\mathbf{A}^\mathsf{T}\mathbf{A}$$

Vectorize and concatenate the partials

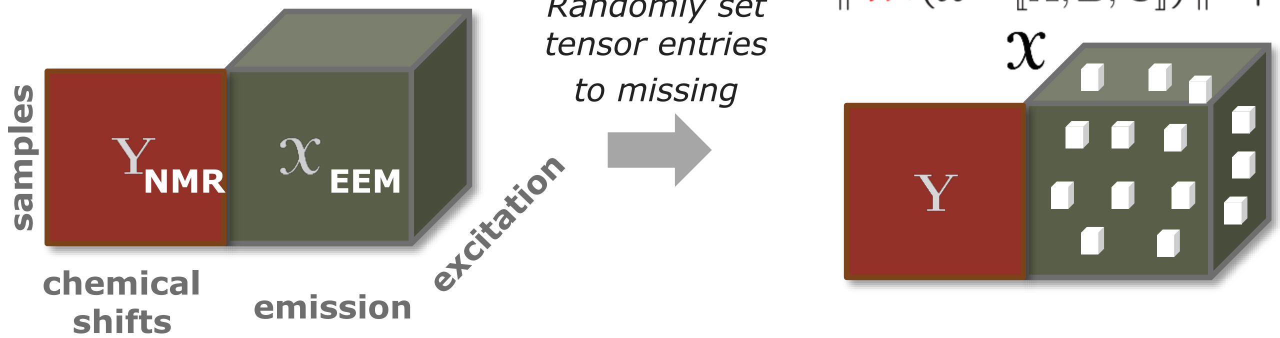
$$\nabla f_{\mathbf{W}} = \begin{bmatrix} \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{a}_1} \\ \vdots \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{a}_R} \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{b}_1} \\ \vdots \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{b}_R} \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{c}_1} \\ \vdots \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{c}_R} \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{d}_1} \\ \vdots \\ \frac{\partial f_{\mathbf{W}}}{\partial \mathbf{d}_R} \end{bmatrix} \quad \cdots$$

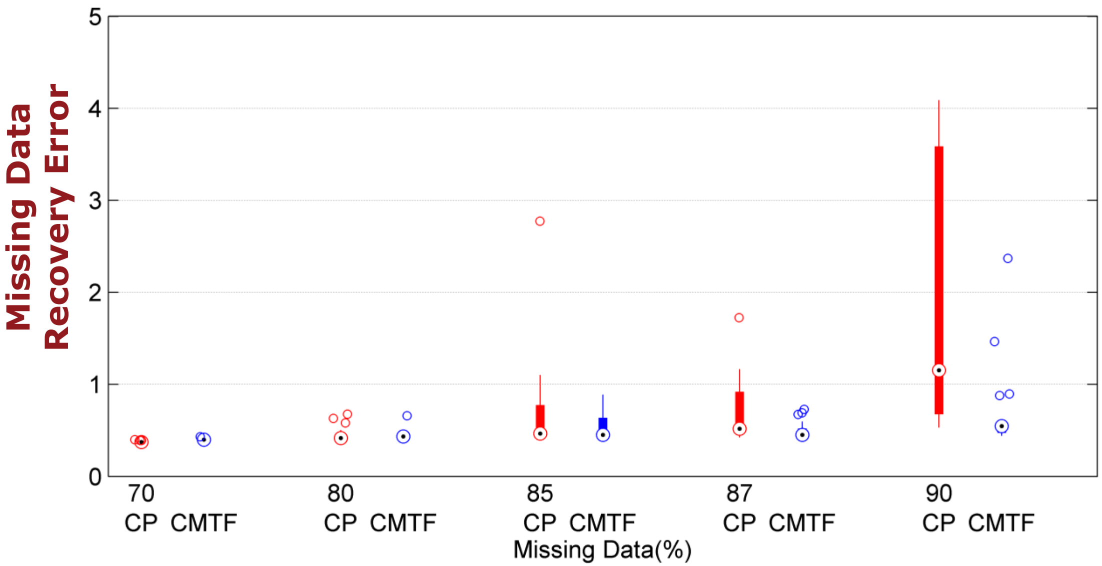
CMTF can improve missing data estimation performance!

[Acar et al., *Chemometrics and Intelligent Lab. Systems*, 2013]

Metabolomics: We have plasma samples measured using different analytical techniques, i.e., NMR and Fluorescence Spectroscopy.

$$\| \mathcal{W} * (\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]) \|^2 + \| \mathbf{Y} - \mathbf{A}\mathbf{D}^{\mathsf{T}} \|^2$$

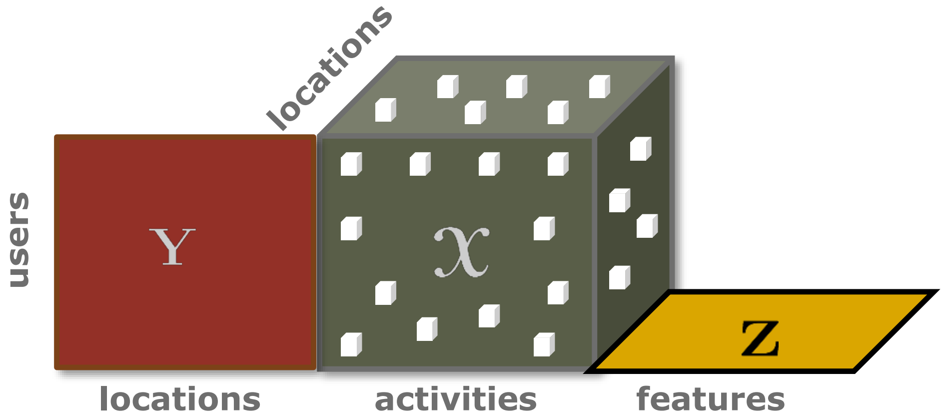




$$\text{Missing Data Recovery Error} = \frac{\| (1 - \mathcal{W}) * (\mathcal{X} - \hat{\mathcal{X}}) \|}{\| (1 - \mathcal{W}) * \mathcal{X} \|}$$

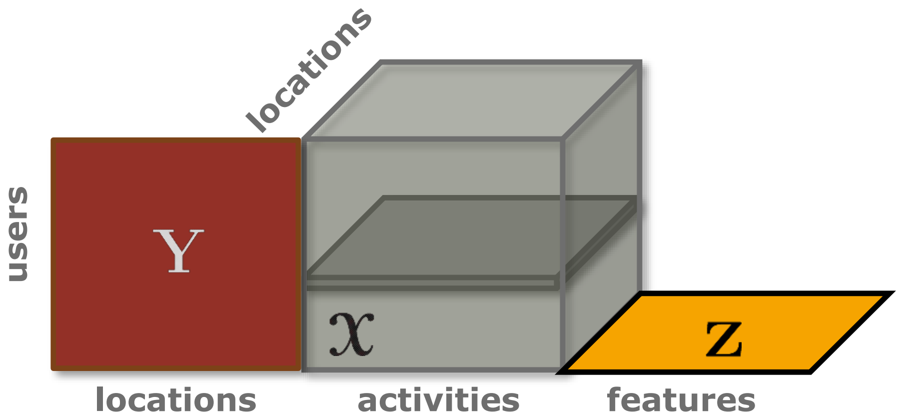
Missing data recovery error is lower using the coupled approach at high amounts of missing data.

Coupling can handle structured missing data!



$$x_{ijk} = \begin{cases} 1 & \text{if user i performs activity j at location k,} \\ 0 & \text{otherwise.} \end{cases}$$

Coupling can handle structured missing data!



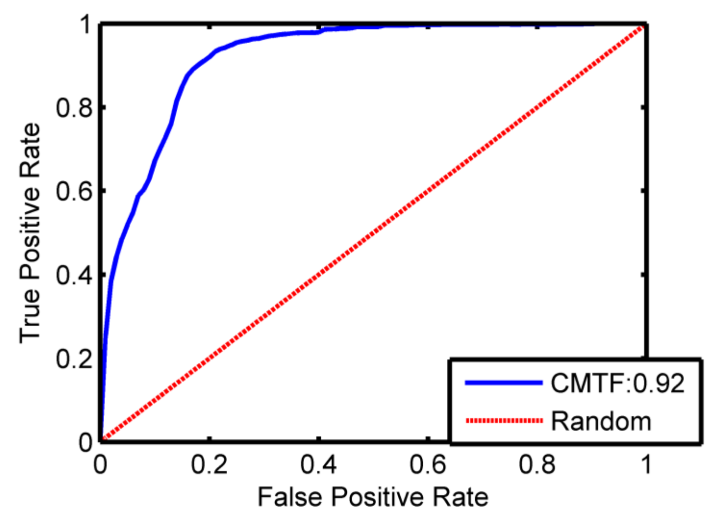
We face with the cold-start problem when a new user starts using an application, e.g., location-activity recommender system. This will correspond to a missing slice for the new user.

For the missing slice i (for i=1,2,...,*I*):

Original values	Estimated values using CMTF
$\text{vec}(\mathbf{X}_i)$	$\text{vec}(\hat{\mathbf{X}}_i)$

$$x_{ijk} = \begin{cases} 1 & \text{if user i performs activity j at location k,} \\ 0 & \text{otherwise.} \end{cases}$$

Average ROC curve for *I*=146 users



We cannot use low-rank approximation of a tensor to fill in the missing slice. However, we can make use of additional sources of information through the coupled model.

$$\mathbf{Y} \approx \mathbf{A}\mathbf{D}^{\mathsf{T}}$$

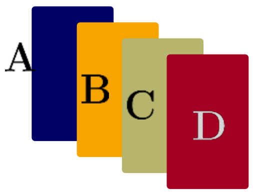
$$\mathcal{X} \approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\mathbf{Z} \approx \mathbf{C}\mathbf{E}^{\mathsf{T}}$$

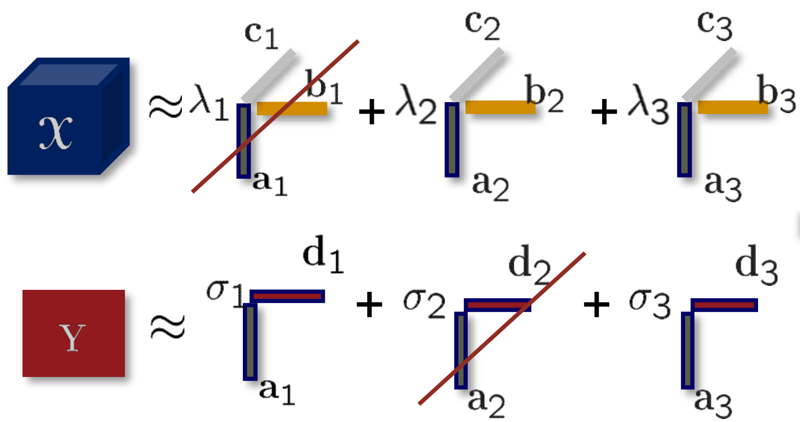
CMTF fails to identify shared/unshared factors!

In real applications, coupled data sets often have both ***shared*** and ***unshared*** factors. However, CMTF formulation focuses on modeling only the ***shared*** factors and fails to identify shared/unshared factors.

Generate factor matrices

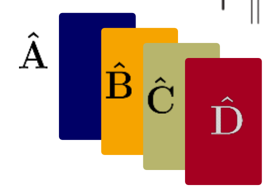


Construct coupled data sets



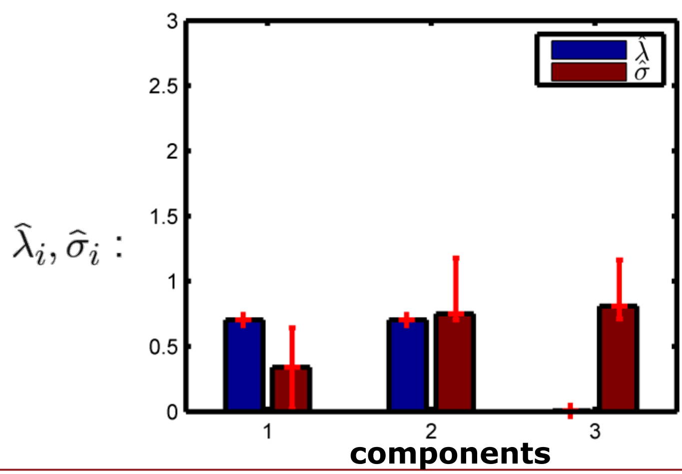
Minimize *f* using CMTF-OPT

$$f(\hat{\mathrm{A}}, \hat{\mathrm{B}}, \hat{\mathrm{C}}, \hat{\mathrm{D}}) = \left\| \mathcal{X} - [\![\hat{\mathrm{A}}, \hat{\mathrm{B}}, \hat{\mathrm{C}}]\!] \right\|^2 + \left\| \mathrm{Y} - \hat{\mathrm{A}}\hat{\mathrm{D}}^{\mathsf{T}} \right\|^2$$



Example: One shared and one unshared component in each data set:

$$[\lambda_1 \ \lambda_2 \ \lambda_3] = [0 \ 1 \ 1]$$
$$[\sigma_1 \ \sigma_2 \ \sigma_3] = [1 \ 0 \ 1]$$

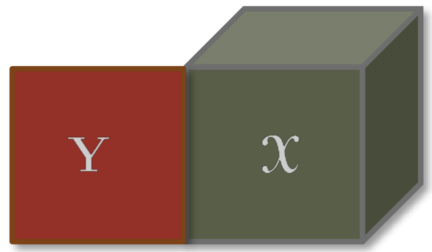


Fails to identify shared and unshared components!

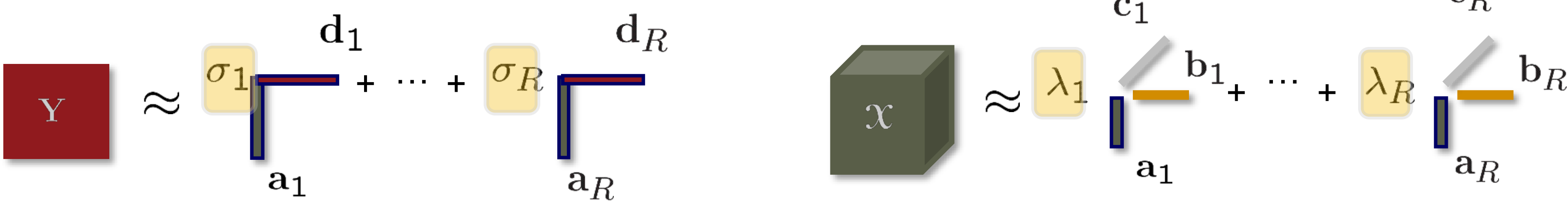
ACMTF: Structure-Revealing CMTF [Acar et al., *BMC Bioinformatics*, 2014]

We reformulate the coupled matrix and tensor factorization problem by having factor matrices with unit norm columns and explicitly representing the weights of rank-one components in the formulation. Through modeling constraints/penalties, we let the model identify shared/unshared components.

$$\mathbf{Y} \approx \mathbf{A}\boldsymbol{\Sigma}\mathbf{D}^{\mathsf{T}}$$



$$\mathcal{X} \approx [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$



l1-norm

$$\|\mathbf{x}\|_1 = \sum |x_i|$$

Structure-revealing model:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\boldsymbol{\Sigma},\lambda} \|\mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\|^2 + \|\mathbf{Y} - \mathbf{A}\boldsymbol{\Sigma}\mathbf{D}^{\mathsf{T}}\|^2 + \beta\|\lambda\|_1 + \beta\|\sigma\|_1$$

$$\text{s.t.} \; \|\mathbf{a}_r\|_2 = \|\mathbf{b}_r\|_2 = \|\mathbf{c}_r\|_2 = \|\mathbf{d}_r\|_2 = 1, \text{for } r = 1, \ldots, R$$

Original CMTF

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \|\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\|^2 + \|\mathbf{Y} - \mathbf{A}\mathbf{D}^{\mathsf{T}}\|^2$$

ACMTF: Unconstrained Optimization

Optimization Problem:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\Sigma,\lambda} \| \mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^\top \|^2 + \beta \| \lambda \|_1 + \beta \| \sigma \|_1$$

$$\text{s.t. } \boxed{\| \mathbf{a}_r \|_2 = \| \mathbf{b}_r \|_2 = \| \mathbf{c}_r \|_2 = \| \mathbf{d}_r \|_2 = 1,} \text{ for } r = 1, ..., R$$

Define the objective function:

Add as quadratic penalty terms

$$f(\lambda, \Sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = \| \mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^\top \|^2 + \boxed{\beta \| \lambda \|_1 + \beta \| \sigma \|_1} + ...$$

Smooth Approximation:

Replace sparsity penalties with differentiable approximations

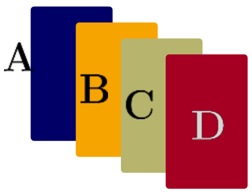
$$f(\lambda, \Sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = \| \mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^\top \|^2 + \boxed{\beta \sum_{r=1}^{R} \sqrt{\lambda_r^2 + \epsilon} + \beta \sum_{r=1}^{R} \sqrt{\sigma_r^2 + \epsilon}} + ...$$

Compute the gradient and pick a first order optimization method

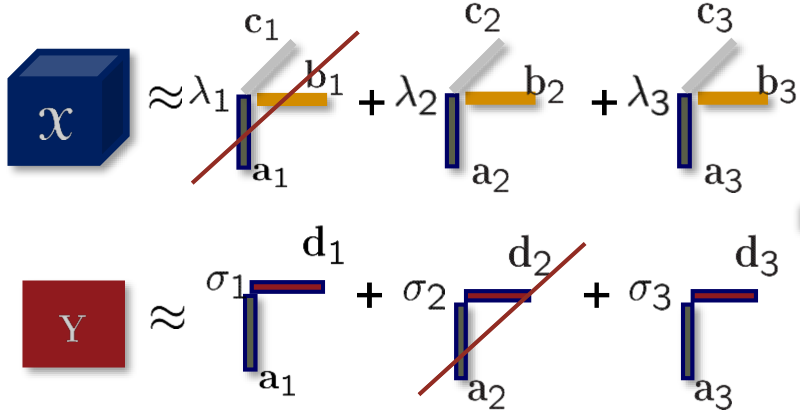
Nonlinear Conjugate Gradient from **Poblano Toolbox** **[Dunlavy, Kolda and Acar, 2010]**

Sparsity penalties enable us to capture the true structure!

Generate factor matrices



Construct coupled data sets



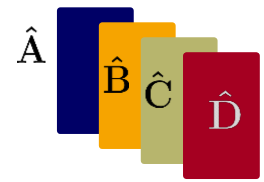
One shared and one unshared component in each data set:

$$[\lambda_1\ \lambda_2\ \lambda_3] = [0\ 1\ 1]$$
$$[\sigma_1\ \sigma_2\ \sigma_3] = [1\ 0\ 1]$$

Minimize

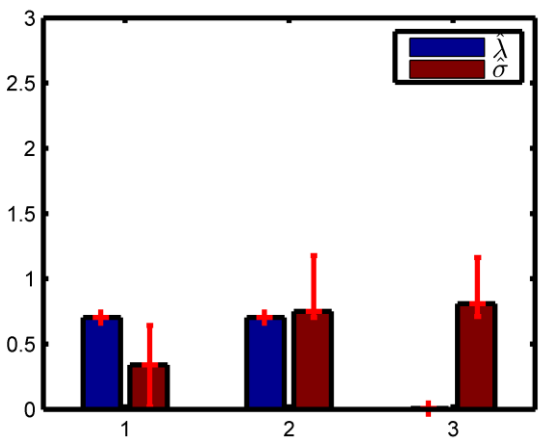
CMTF: $f(\hat{A}, \hat{B}, \hat{C}, \hat{D})$

ACMTF: $f(\hat{\lambda}, \hat{\Sigma}, \hat{A}, \hat{B}, \hat{C}, \hat{D})$



CMTF

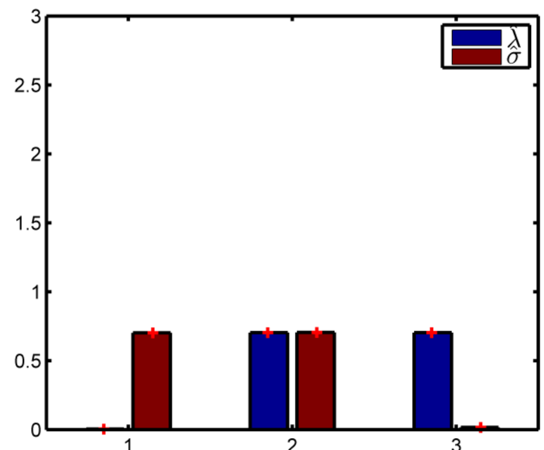
$\hat{\lambda}_i, \hat{\sigma}_i$:



Fails to identify shared/unshared components!

ACMTF

$\hat{\lambda}_i, \hat{\sigma}_i$:



True structure captured!

ACMTF: Constrained Optimization

[Acar, Nilsson, and Saunders, *EUSIPCO*, 2014]

In order to have a flexible modeling framework, we use a general-purpose optimization solver SNOPT (Sparse Nonlinear OPTimizer) [Gill, Murray and Saunders, 2005].

SNOPT is designed for large constrained optimization problems with smooth nonlinear functions in the objective and constraints.

SNOPT uses a sequential quadratic programming (SQP) algorithm to minimize an augmented Lagrangian.

SNOPT

$$\min_{x \in \mathbb{R}^n} \quad \phi(x)$$

$$\texttt{s.t.} \quad l \leq \begin{pmatrix} x \\ Ax \\ c(x) \end{pmatrix} \leq u$$

where *c(x)* indicates nonlinear functions, and A is a sparse matrix.

Structure-revealing CMTF model:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\Sigma,\lambda} \| \mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^\mathsf{T} \|^2$$

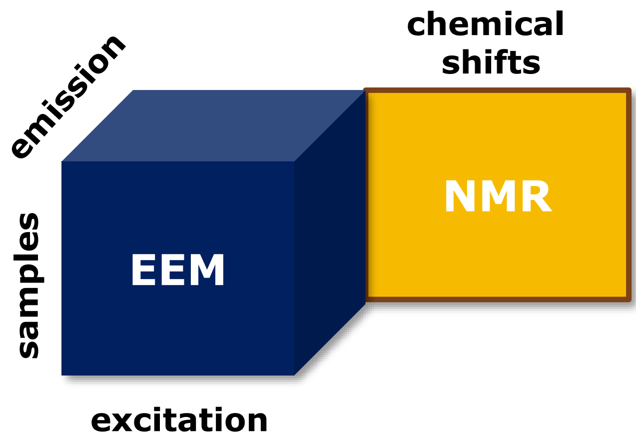
$$\texttt{s.t.} \quad \| \mathbf{a}_r \|_2 = \| \mathbf{b}_r \|_2 = \| \mathbf{c}_r \|_2 = \| \mathbf{d}_r \|_2 = 1$$

$$\sum_{r=1}^{R} \lambda_r \leq \beta, \ \sum_{r=1}^{R} \sigma_r \leq \beta$$

$$\sigma_r, \lambda_r \geq 0, \text{for } r = 1, \ldots, R.$$

Additional constraints can easily be incorporated!

In many data fusion problems, we may need the following constraints to capture the underlying structures accurately.



Nonnegativity Constraints:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\Sigma,\lambda} \| \mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^\mathsf{T} \|^2$$
$$\text{s.t.} \quad \| \mathbf{a}_r \|_2 = \| \mathbf{b}_r \|_2 = \| \mathbf{c}_r \|_2 = \| \mathbf{d}_r \|_2 = 1$$
$$\sum_{r=1}^{R} \lambda_r \leq \beta, \sum_{r=1}^{R} \sigma_r \leq \beta$$
$$\sigma_r, \lambda_r \geq 0, b_{jr}, c_{kr}, d_{mr} \geq 0$$
$$\text{for } r = 1:R, j = 1:J, k = 1:K, m = 1:M.$$

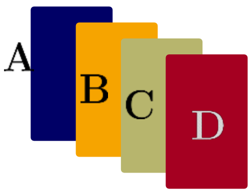
Angular Constraints: When coupled data sets are overfactored, one shared factor may be represented by two closely-correlated factors. In that case, the structure–revealing model will fail to identify shared factors accurately.

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\Sigma,\lambda} \| \mathcal{X} - [\![\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^\mathsf{T} \|^2$$
$$\text{s.t.} \quad \| \mathbf{a}_r \|_2 = \| \mathbf{b}_r \|_2 = \| \mathbf{c}_r \|_2 = \| \mathbf{d}_r \|_2 = 1$$
$$|\mathbf{a}_r^\mathsf{T}\mathbf{a}_p| \leq \theta, |\mathbf{b}_r^\mathsf{T}\mathbf{b}_p| \leq \theta, |\mathbf{c}_r^\mathsf{T}\mathbf{c}_p| \leq \theta, |\mathbf{d}_r^\mathsf{T}\mathbf{d}_p| \leq \theta$$
$$\sum_{r=1}^{R} \lambda_r \leq \beta, \sum_{r=1}^{R} \sigma_r \leq \beta$$
$$\sigma_r, \lambda_r \geq 0 \text{ for } r, p \in \{1:R\}, r \neq p.$$

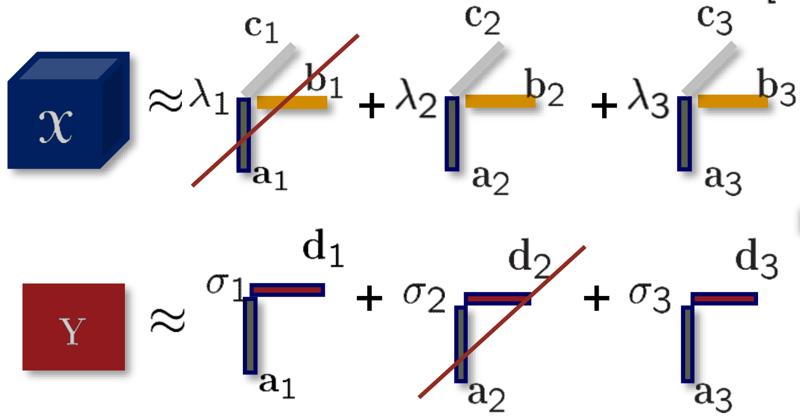
Angular constraints have a promising performance in the case of overfactoring!

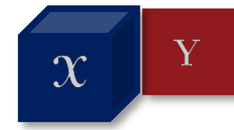
$$[\lambda_1\ \lambda_2\ \lambda_3] = [0\ 1\ 1]$$
$$[\sigma_1\ \sigma_2\ \sigma_3] = [1\ 0\ 1]$$

Generate factor matrices



Construct coupled data sets





Solve for A, B, C, D, λ, and Σ using constrained optimization

Overfactoring (R=4):

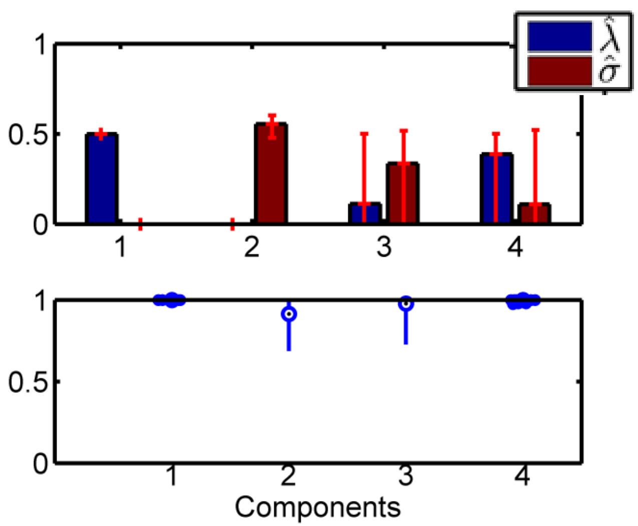
Angular Constraints Inactive

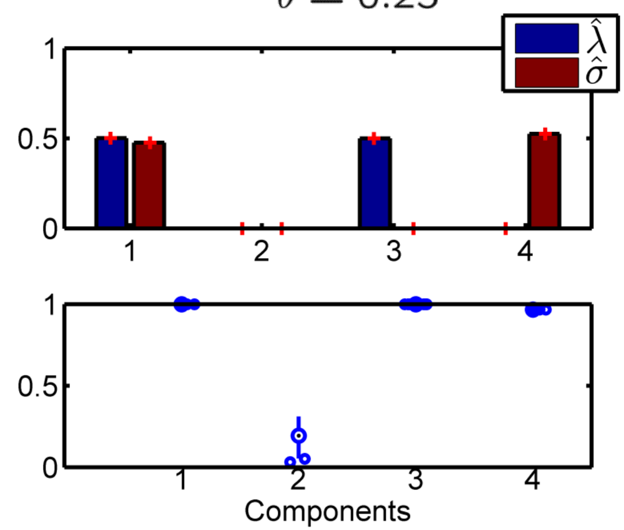
Angular Constraints Active

$\theta = 0.25$

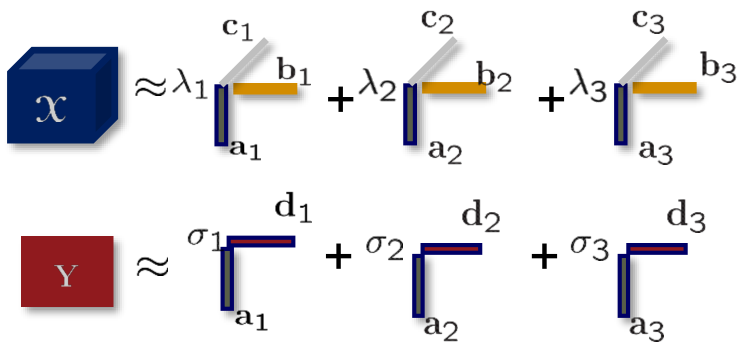
Match score:

$$\frac{\hat{\mathbf{a}}_r^T \mathbf{a}_r}{\|\hat{\mathbf{a}}_r\|\,\|\mathbf{a}_r\|}$$

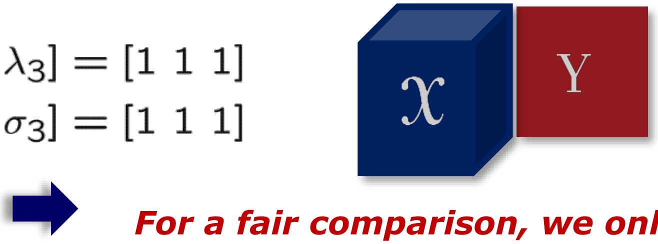




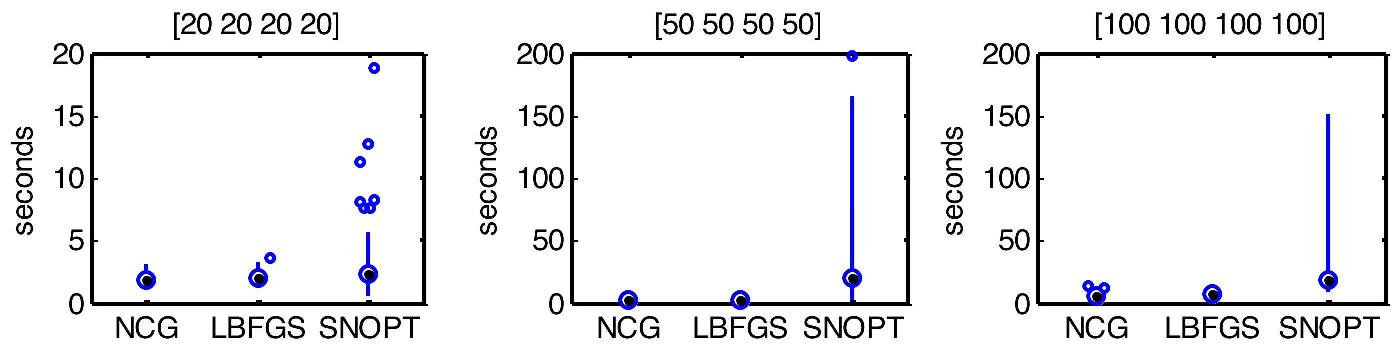
Performance Comparison: Unconstrained Optimization vs. SNOPT

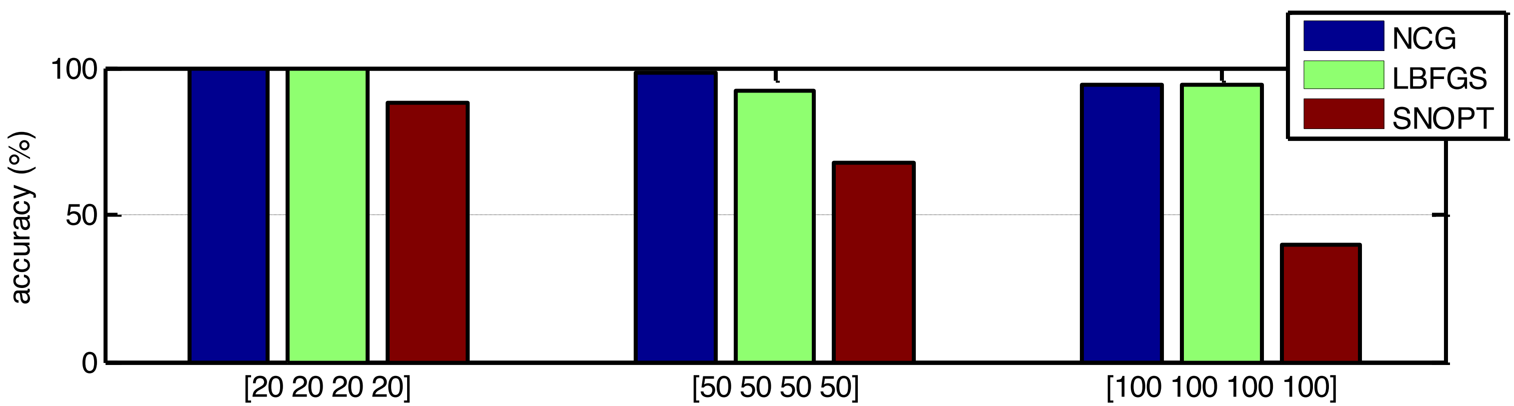


$[\lambda_1\ \lambda_2\ \lambda_3] = [1\ 1\ 1]$
$[\sigma_1\ \sigma_2\ \sigma_3] = [1\ 1\ 1]$



For a fair comparison, we only have norm constraints (treated as quadratic penalties in the unconstrained version).





Application: Joint analysis of LC-MS and NMR measurements

[Acar et al., *BMC Bioinformatics*, 2014]

Goal: To identify shared/unshared factors in each data set

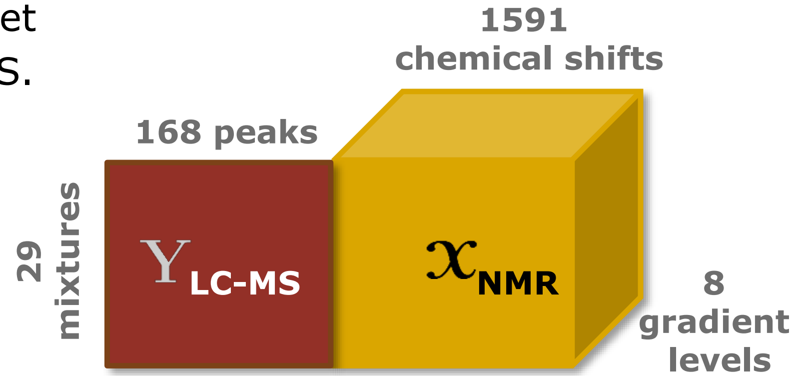
Data: 29 mixtures measured using DOSY-NMR and LC-MS.

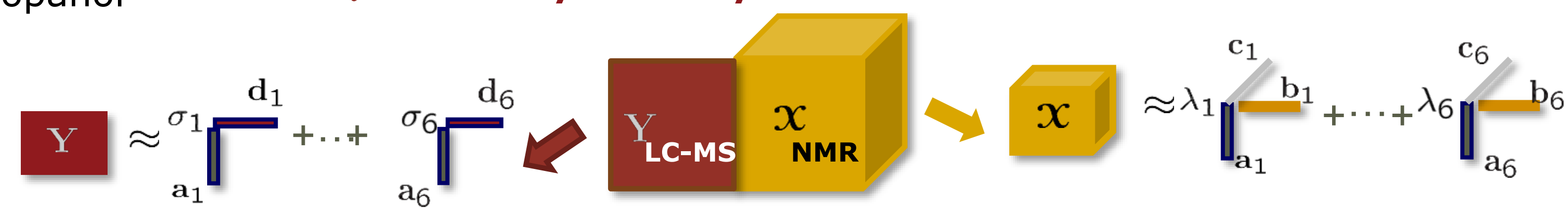
Mixtures are prepared using the following five chemicals:

- Val-Try-Val
- Trp – Gly
- Phe
- Maltoheptaose
- Propanol

Visible by LC-MS and NMR (Val-Try-Val, Trp – Gly, Phe, Maltoheptaose)

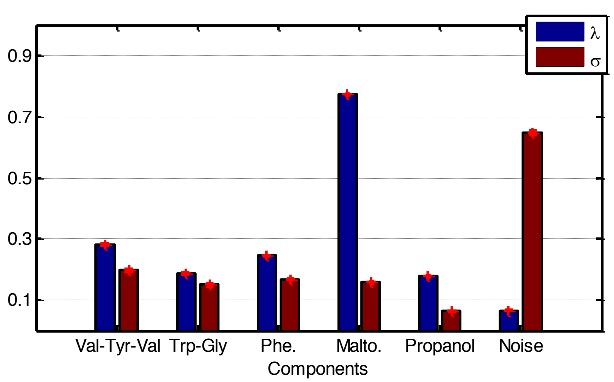
Visible by NMR only (Propanol)

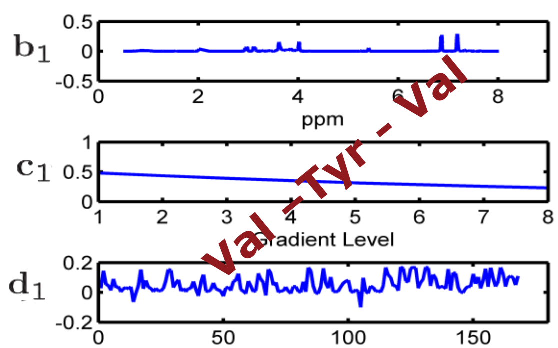


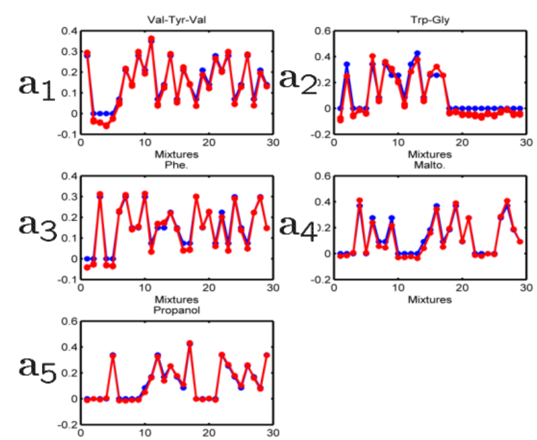


$$\min_{A,B,C,D,\Sigma,\lambda} \left\| \mathcal{X} - [\![\lambda; A, B, C]\!] \right\|^2 + \left\| Y - A\Sigma D^T \right\|^2 + \beta\|\lambda\|_1 + \beta\|\sigma\|_1$$

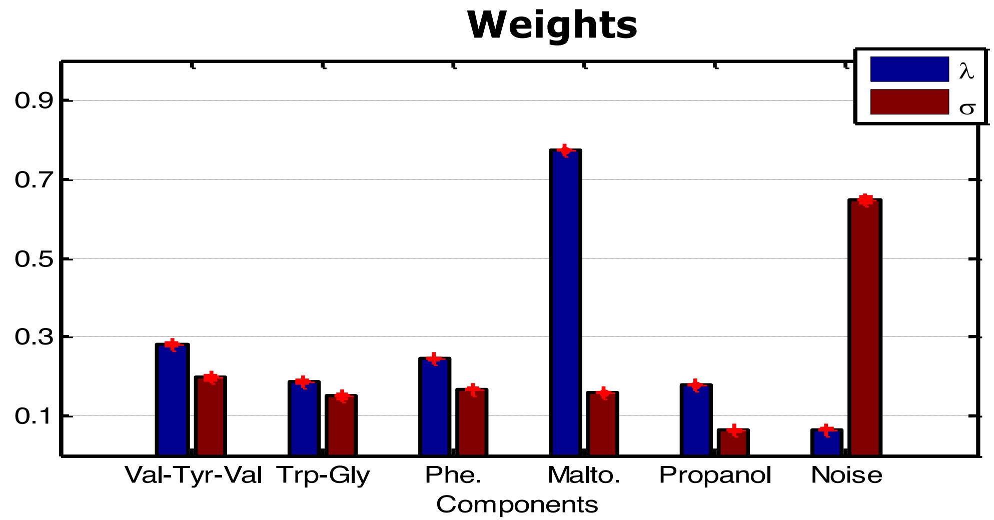
$$\text{s.t.} \; \|a_r\|_2 = \|b_r\|_2 = \|c_r\|_2 = \|d_r\|_2 = 1, \text{for } r = 1, ..., R$$



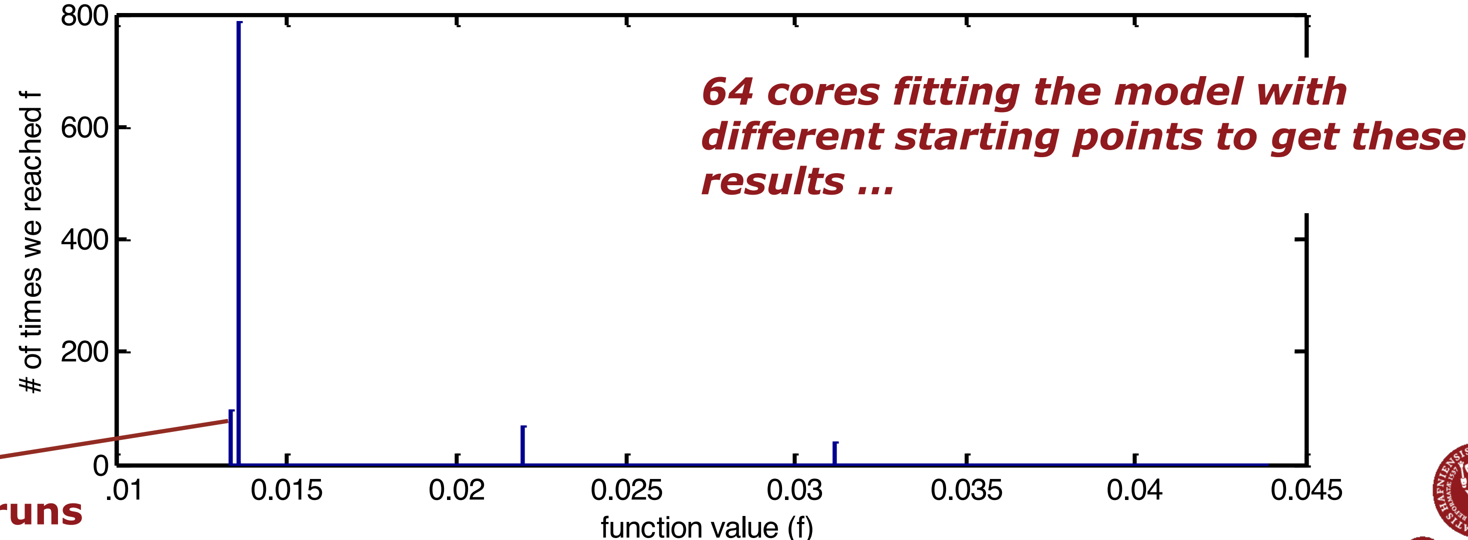




Need better ways for dealing with the initialization problem!



- The minimum function value: $f(\hat{\lambda}, \hat{\Sigma}, \hat{A}, \hat{B}, \hat{C}, \hat{D}) = 0.0134$
- Out of 1000 runs with random initializations, we get the minimum function value 98 times.

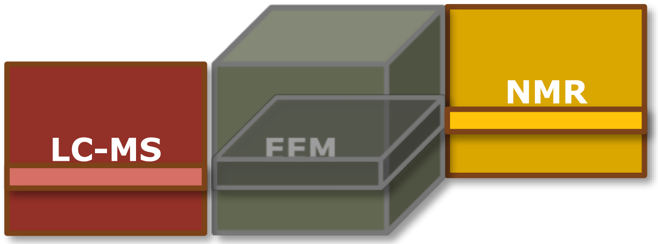


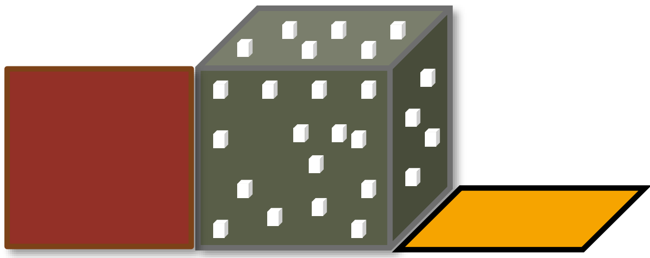
These are the runs we interpret!

Summary

- **Goal:** Joint analysis of heterogeneous data sets
- **Our Approach:** Coupled matrix and tensor factorizations
  - Original formulation assuming all components are shared
  - Reformulation of the original model to identify shared/unshared factors accurately

  >> All can handle missing data!
- **Algorithmic Approach:** All-at-once optimization
  - Unconstrained optimization
  - Constrained optimization
- **Applications:**
  - Chemometrics/Metabolomics
  - Social network analysis





- **Open issues:**
  - **More flexible structure-revealing data fusion models, e.g., constraints** [Acar, Nilsson, and Saunders, *EUSIPCO*, 2014], **flexible couplings** [Farias et al., *LVA/ICA*, 2015]**...**
  - **More robust and/or computationally efficient approaches**

Thank you!

CMTF:

E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once Optimization for Coupled Matrix and Tensor Factorizations. *KDD Workshop on Mining and Learning with Graphs,* 2011 (arXiv:1105.3422)

ACMTF: Structure-revealing data fusion model

E. Acar, R. Bro, and A. K. Smilde, Data Fusion in Metabolomics using Coupled Matrix and Tensor Factorizations, *Proceedings of the IEEE*, 103: 1602-1620, 2015.

E. Acar, E. E. Papalexakis, G. Gurdeniz, M. A. Rasmussen, A. J. Lawaetz, M. Nilsson, and R. Bro, Structure Revealing Data Fusion, *BMC Bioinformatics*, 15: 239, 2014.

E. Acar, M. Nilsson, and M. Saunders, A Flexible Modeling Framework for Coupled Matrix and Tensor Factorizations, *EUSIPCO,* pp. 111-115, 2014.

E. Acar, A. J. Lawaetz, M. A. Rasmussen, and R. Bro, Structure Revealing Data Fusion Model with Applications in Metabolomics, *IEEE EMBC,* pp. 6023 - 6026, 2013.

Missing Data Estimation: E. Acar, M. A. Rasmussen, F. Savorani, T. Næs, and R. Bro. Understanding Data Fusion within the Framework of Coupled Matrix and Tensor Factorizations, *Chemometrics and Intelligent Laboratory Systems,* 129: 53-63, 2013.

Link Prediction: B. Ermis, E. Acar, and A. T. Cemgil. Link Prediction in Heterogeneous Data via Generalized Coupled Tensor Factorization, *Data Mining and Knowledge Discovery*, 29(1): 203-236, 2015.

Danish Agency for Science Technology and Innovation
Ministry of Science, Innovation and Higher Education

Sapere Aude

Evrim Acar
http://www.models.life.ku.dk/~acare/

JODA: Joint Data Analysis for Enhanced Knowledge Discovery
http://www.models.life.ku.dk/joda