University of BRISTOL

Making Sense of (Multi-)Relational Data

Part V: A global fully-relational approach: Coupled Matrix-Tensor Factorisations

Jefrey Lijffijt

Eirini Spyropoulou

Tijl De Bie

Data model

- ‘Global’ patterns
- Applicable to:
  - Simple tabular (‘non-relational data’)
    - Matrix factorizations
  - Multi-modal (n-ary) data
    - Tensor factorizations
  - Entity-Relationship data model
    - Coupled matrix-tensor factorizations

Research roots

- Research domains:
    - Chemometrics
    - Psychometrics
    - Econometrics
    - Bioinformatics
    - Computational neuroscience
    - Numerical linear algebra (and ICA)
    - Data mining and machine learning
- "Data fusion", "Coupled data", "Linked data", "Multiset data", "Multiblock data", "Integrative data analysis",…

Tensorlab

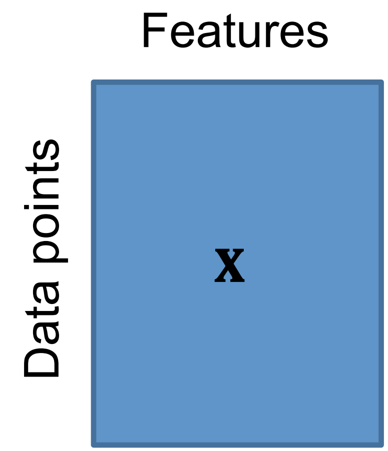
- Examples with the Tensorlab Matlab toolbox
  - Laurent Sorber, Marc Van Barel and Lieven De Lathauwer. *Tensorlab v2.0,* Available online, January 2014.
  - URL: http://www.tensorlab.net/.
  - Highly intuitive, outstanding documentation, insightful demos…

Matrix factorisations

The data

- Tabular data
- Mathematical representation:
  - A matrix $\mathbf{X}$

$$\mathbf{X}\in\mathbb{R}^{n\times m}$$
$$\mathbf{X}\in\{\mathbf{0},\mathbf{1}\}^{n\times m}$$
$$\mathbf{X}\in\mathbb{Z}^{n\times m}$$
$$\ldots$$

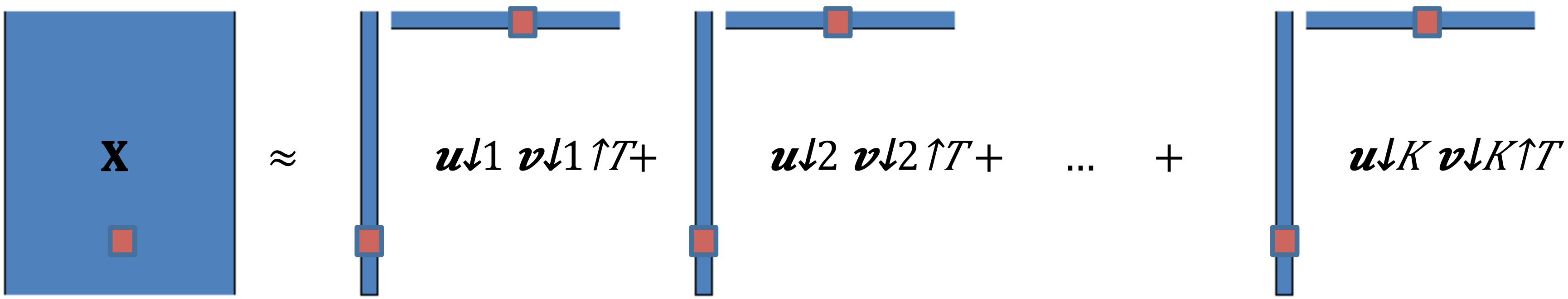


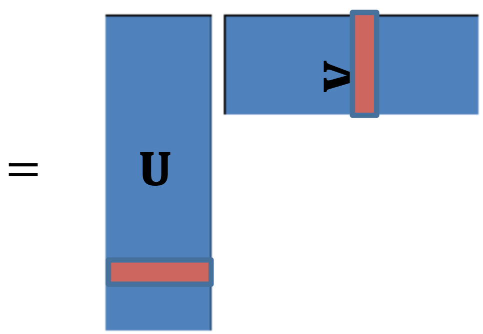
  - Rows of $\mathbf{X}$ often represent ‘data points’
  - Columns of $\mathbf{X}$ often represent ‘features’, ‘attributes’, or ‘dimensions’

University of BRISTOL

Pattern syntax

- Low-rank matrix factorisation:



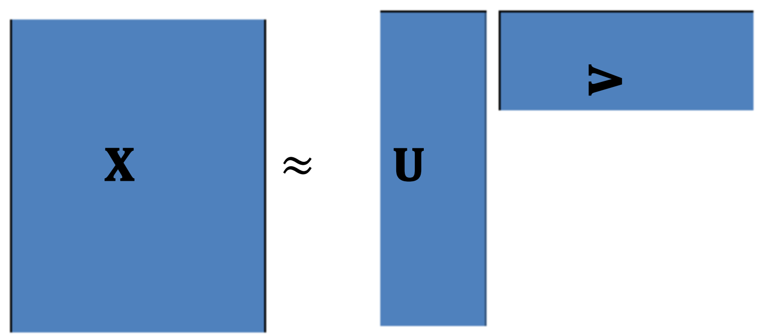


$\mathbf{U} \in \mathbb{R}^{n \times K}$
$\mathbf{V} \in \mathbb{R}^{m \times K}$

$\mathbf{X}|_{i,j} \approx \mathbf{U}\mathbf{V}^{T}|_{i,j} = \sum_{k=1}^{K} u_{i,k} v_{j}$

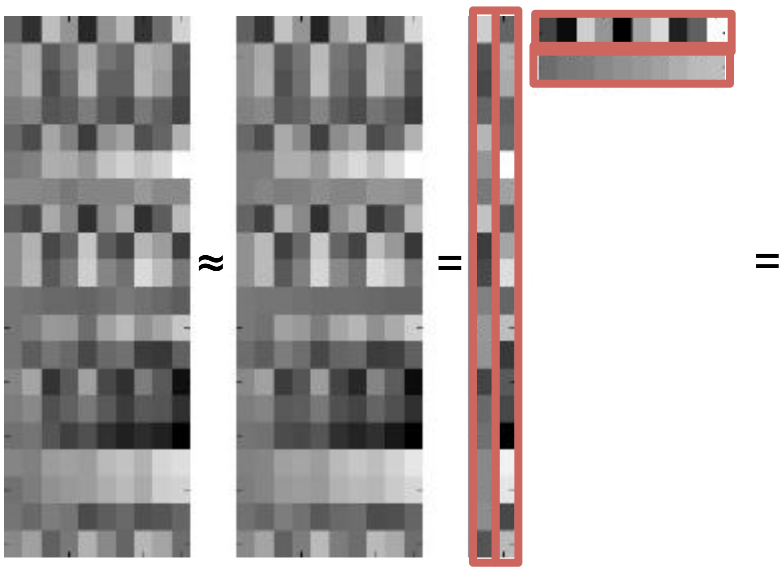
Pattern syntax

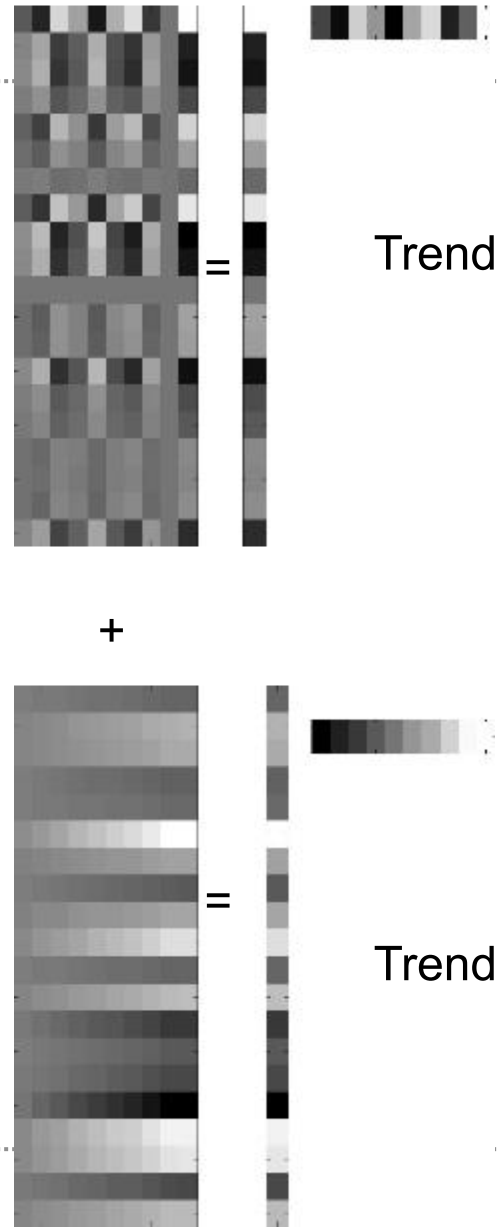
- So, patterns of the form $\mathbf{X} \approx \mathbf{U}\mathbf{V}^{T}$



  - Global patterns
- Constraints on $\mathbf{U}$ *and* $\mathbf{V}$ yield different methods
  - Examples follow…

Linear combinations of ‘trends’: PCA

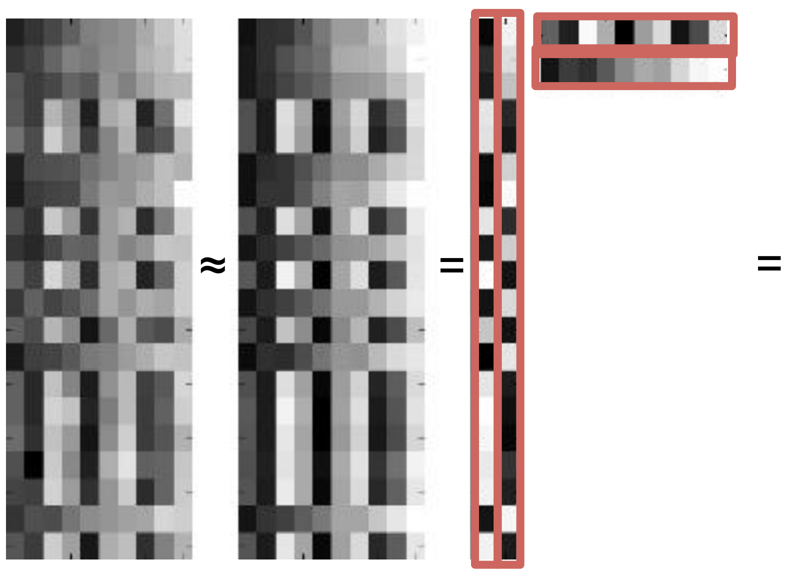


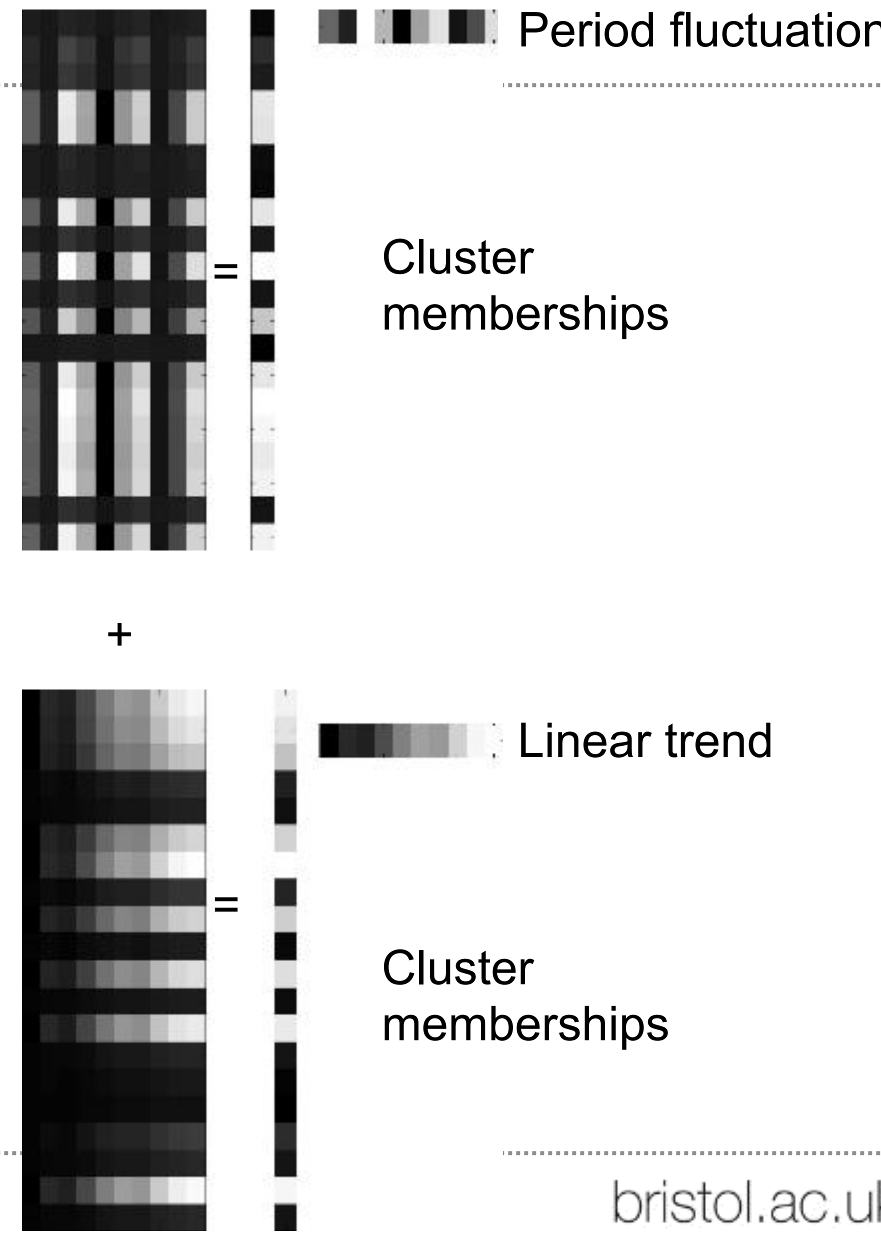


No constraints

University of BRISTOL

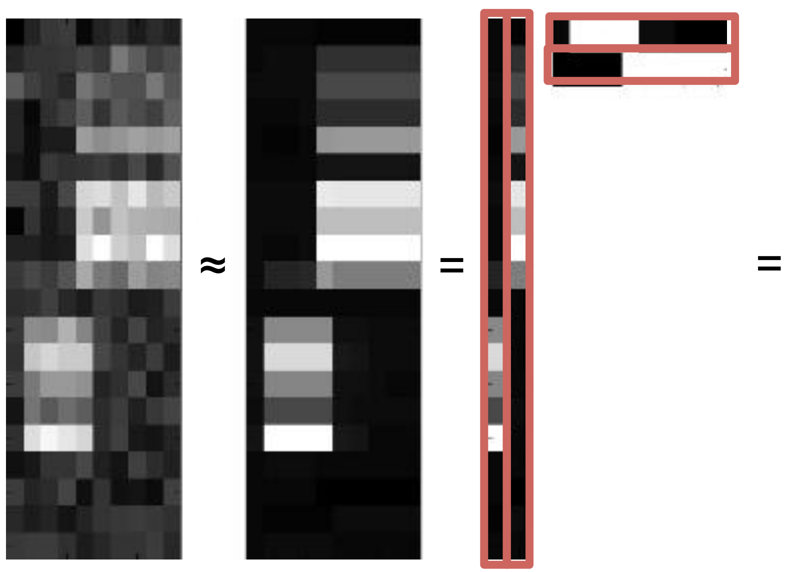
Clustering patterns

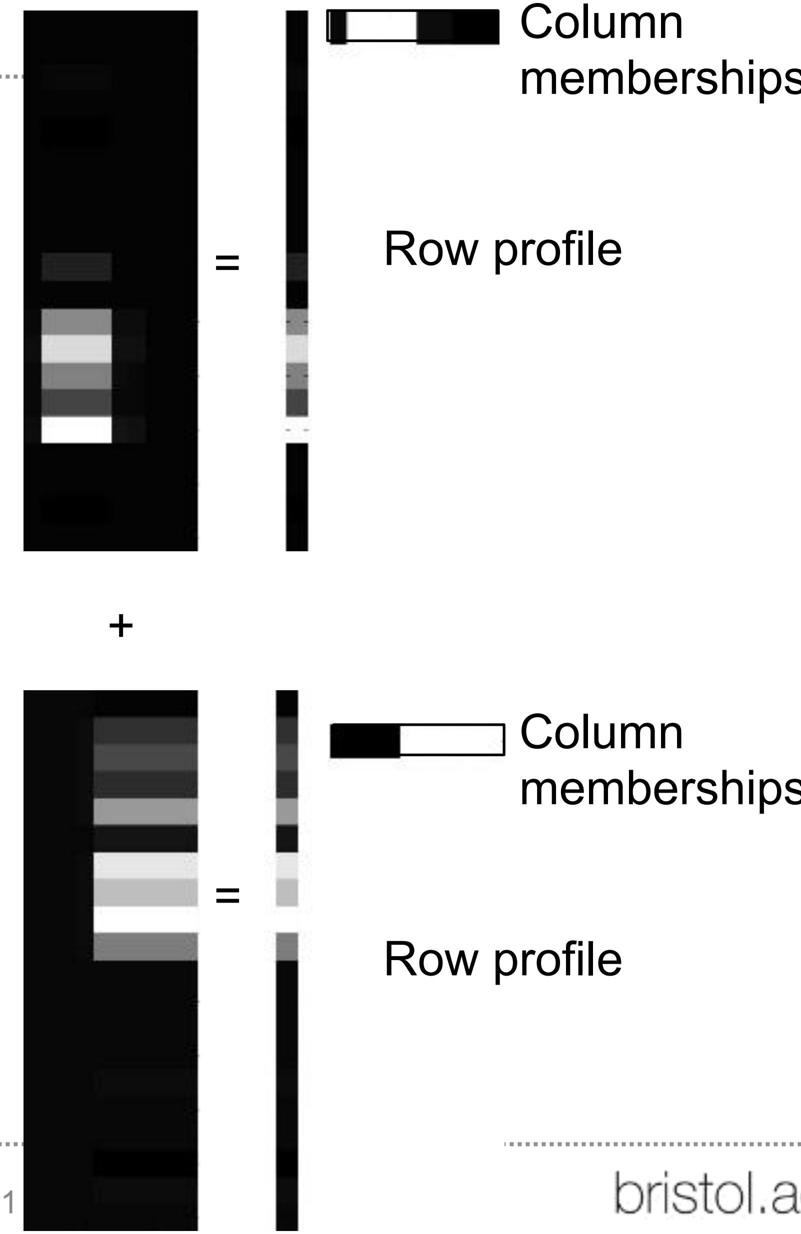




U binary

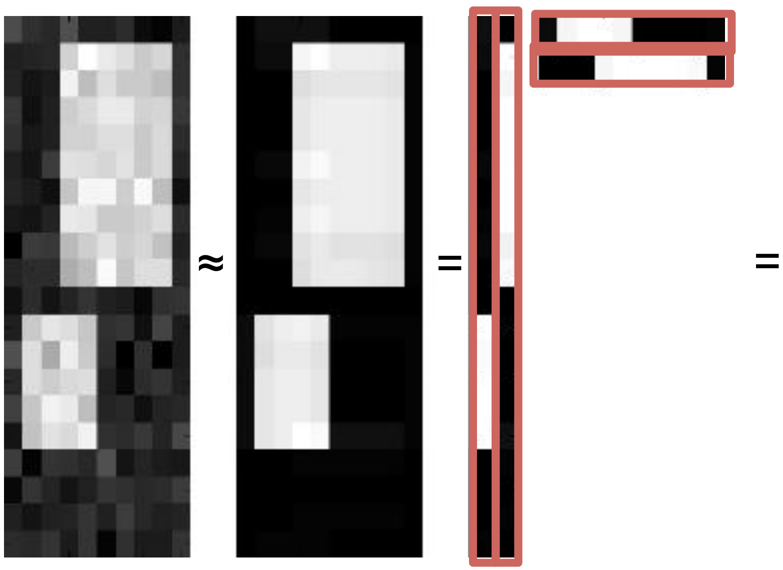
Biclustering patterns

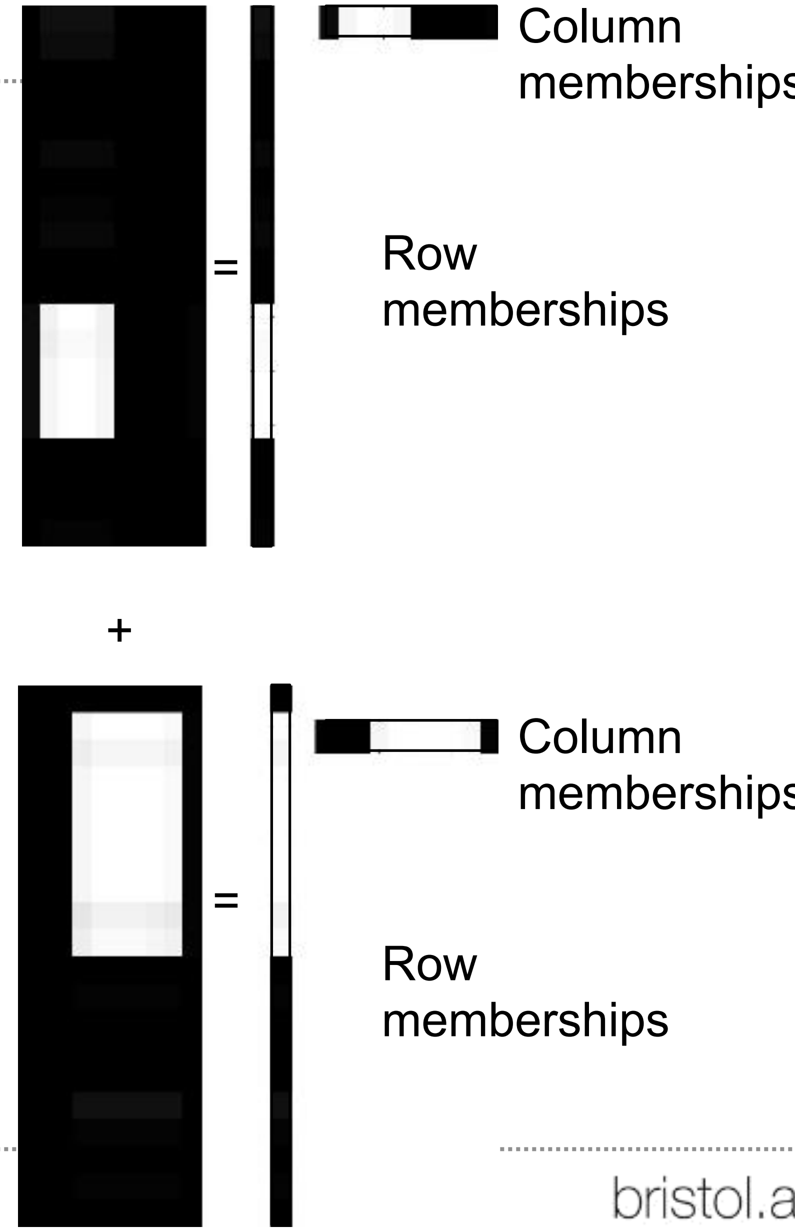




U sparse
V binary

Tile patterns





U binary
V binary

Interestingness

- Almost always: L2-norm
  - More interesting if $\|\mathbf{X}-\mathbf{U}\mathbf{V}^T\|_2$ is smaller
- Alternatives:
  - Robust variants (e.g. L1-norm)
  - Likelihood-based approach
    - Then $\mathbf{U}\mathbf{V}^T$ determines the parameters of a probabilistic model for $\mathbf{X}$ $\mathbf{X}\in\{\mathbf{0},\mathbf{1}\}^{n\times m}$, $\mathbf{X}\in\mathbb{N}^{n\times m}$
    - Especially relevant if

Algorithmic approach

- Aim
  - Pattern set mining
    - With rank K: find best set of K patterns
- User experience
  - Typically one-shot
  - Can be iterative

Algorithmic approach – continuous optimisation

- Eigenvalue problem in simplest case (PCA)
- More complex with
  - constraints on the factors
  - regularisation (e.g. L1 for sparsity)
  - missing values
  - different norm (interestingness)
- Then, often:
  - Alternating least squares (ALS)
  - More generally: coordinate descent
    - Alternate between optimising $\mathbf{U}$ and $\mathbf{V}$ – often individually convex problems
- Prone to local minima, but often reasonably good

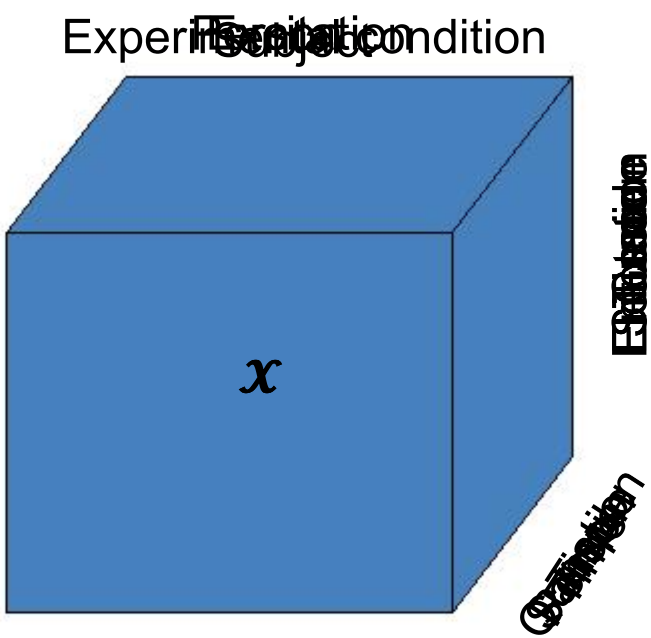
Research issues

- Pattern syntax
  - Estimating the rank K
  - The nature of $\mathbf{U}$ and $\mathbf{V}$
    - Constraints (binary/positive/range-constrained/…)
    - Biases (L1 or L2 regularisation)
  - Other algebras (Pauli Miettinen's work)
- Interestingness
  - Which norm for the error $\|\mathbf{X}-\mathbf{U}\mathbf{V}^T\|$
    - Model-based (Bayesian)?
- Algorithmic
  - More advanced optimisation approaches

Tensor factorisations

The case for tensors

- Data stored in a (hyper)cube: a tensor



Bioinformatics

Computational neuroscience

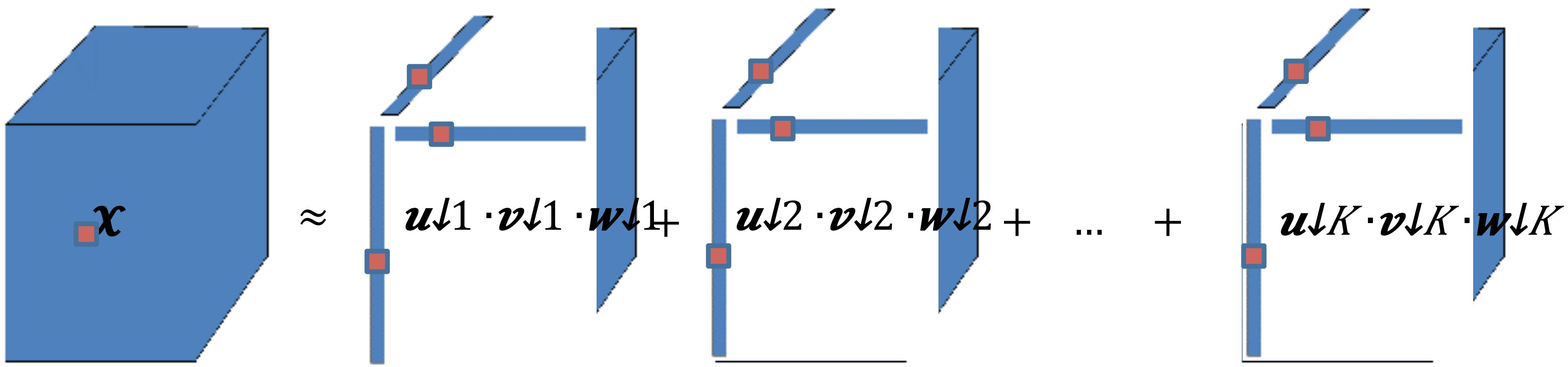
Chemoinformatics / fluorescence spectroscopy

Social network analysis

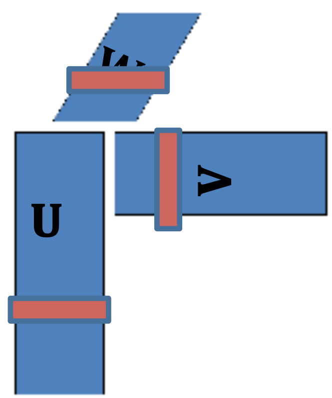
Psychometrics

Pattern syntax: Canonical Polyadic Decomposition (CPD) / Candecomp/Parafac (CP)

- Low-rank tensor factorisation:



=



$\mathbf{U} \in \mathbb{R}^{n \times K}$

$\mathbf{V} \in \mathbb{R}^{m \times K}$

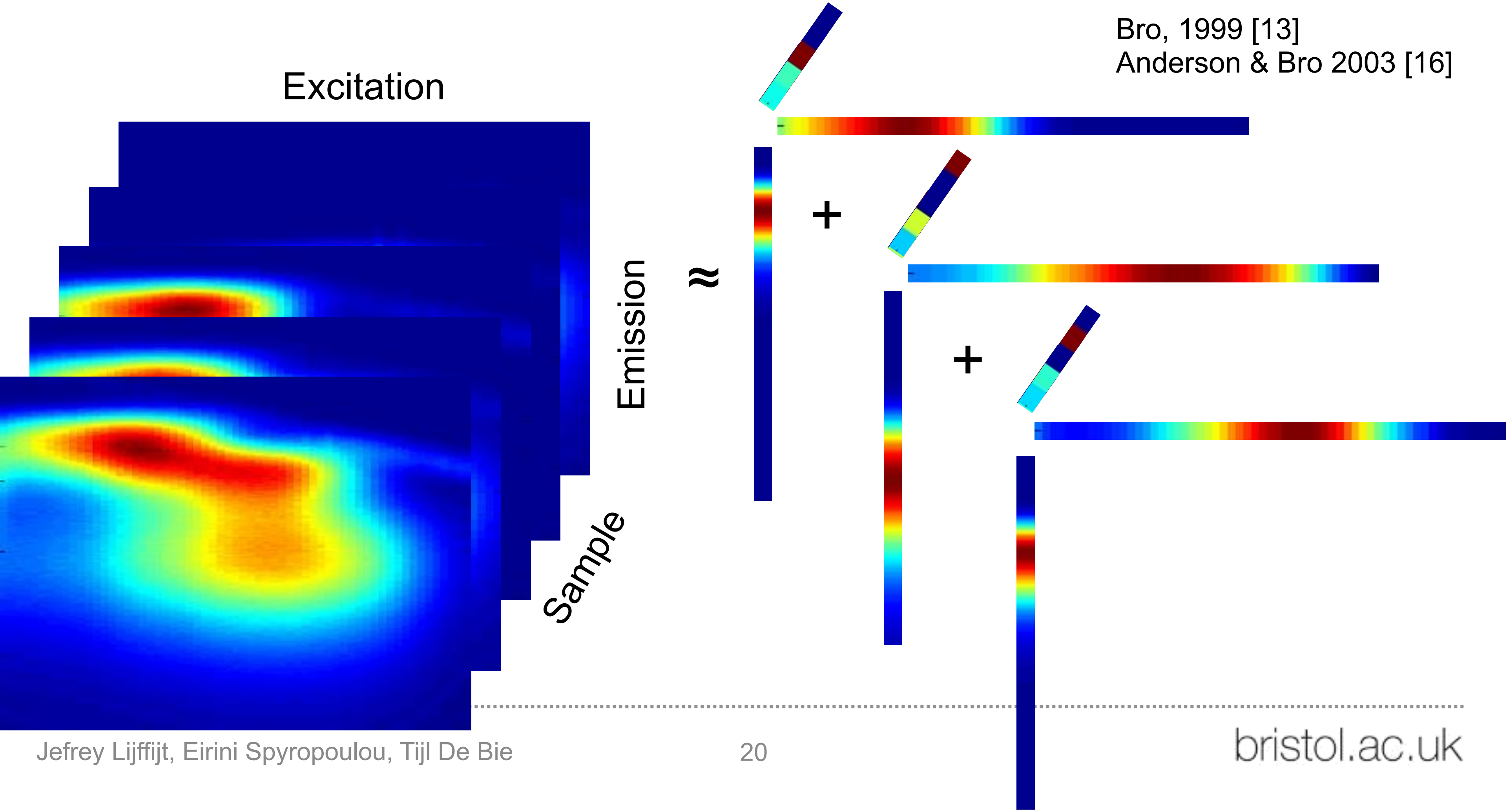
$\mathbf{W} \in \mathbb{R}^{d \times K}$

$\mathbf{X}_{i,j,l} \approx [\![\mathbf{U},\mathbf{V},\mathbf{W}]\!]_{i,j,l} = \sum_{k=1}^{K} u_{i,k}\, v$

Notation: $[\![\mathbf{U},\mathbf{V},\mathbf{W}]\!]$

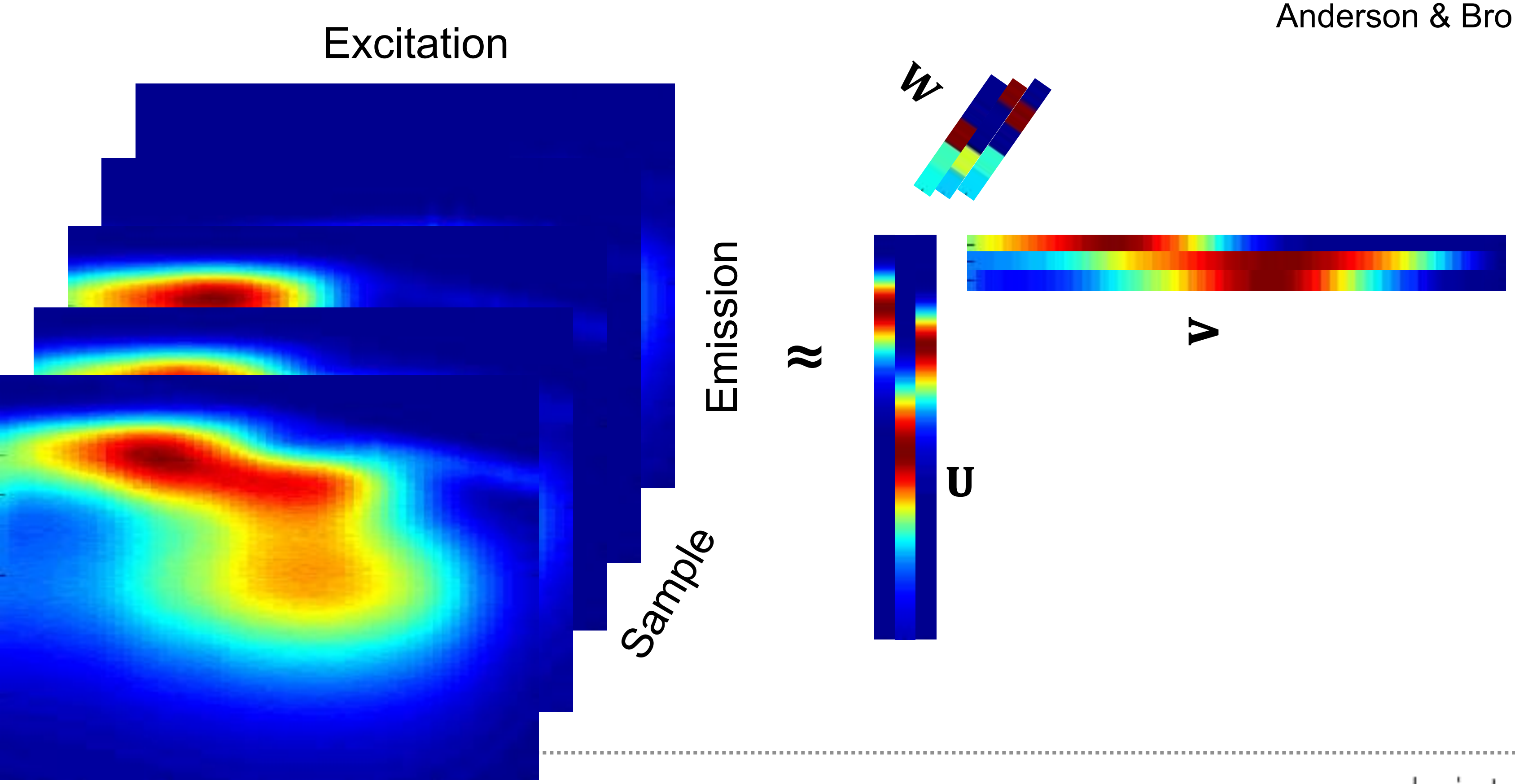
Example: fluorescence spectroscopy

Bro, 1999 [13]
Anderson & Bro 2003 [16]

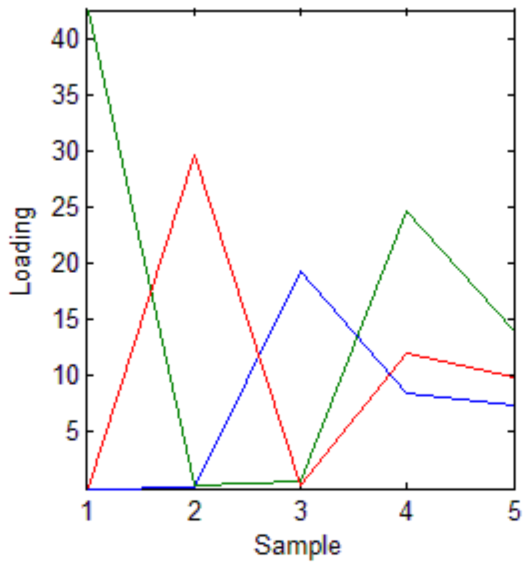


Example: fluorescence spectroscopy

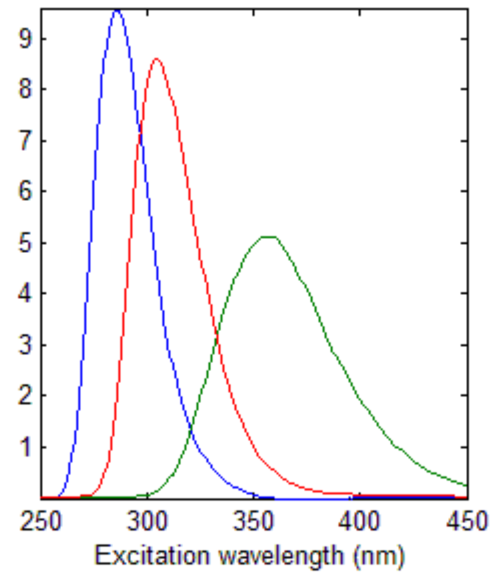
Bro, 1999 [13]
Anderson & Bro 2003 [16]



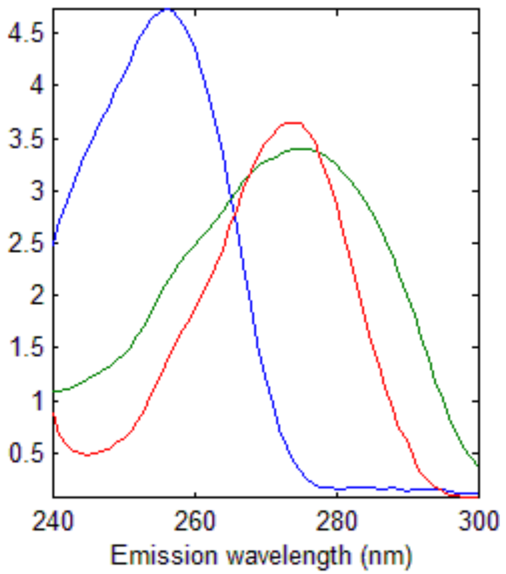
Example: fluorescence spectroscopy



W



V



U

Alternative pattern syntaxes

- Other decompositions:
  - Tucker decomposition (various kinds)
  - Block term decomposition
  - Many others
- Constraints on factors:
  - Positivity, range
  - Identity of factors, e.g.: $\mathbf{U} \equiv \mathbf{V}$ (useful in social network analysis)
  - Sparsity

Interestingness

- Almost always: L2-norm
  - More interesting if $\|\mathcal{X} - [\![\mathbf{U},\mathbf{V},\mathbf{W}]\!]\|_2$ is smaller
- Alternatives:
  - Robust variants (e.g. L1-norm)
  - Likelihood-based approach
    - Then $[\![\mathbf{U},\mathbf{V},\mathbf{W}]\!]$ determines the parameters of a probabilistic model for $\mathcal{X}$
    - Especially relevant if $\mathcal{X} \in \{\mathbf{0},\mathbf{1}\}^{n \times m \times d}$, $\mathcal{X} \in \mathbb{N}^{n \times m \times d}$

Algorithmic approach

- No longer an eigenvalue problem, even in the simplest case
- Typical approach:
  - Alternating Least Squares (ALS)
  - More generally: block coordinate descent
    - Alternate between optimising each of the component matrices
    - Often individually convex
- Prone to local optima, but often reasonably good

Research issues

- Pattern syntax
  - Sometimes: physical models / first principles
    - Otherwise lots of degrees of freedom
  - Estimating the rank K
  - The nature of the factors
    - Constraints (binary / positive / …)
    - Biases (L1/L2 regularisation)
  - Other algebras
- Interestingness
  - Which norm for the error $\|\mathcal{X} - [\![\mathbf{U},\mathbf{V},\mathbf{W}]\!]\|_2$
    - Model-based (Bayesian)?
- Algorithmic

Coupled matrix-tensor factorisations

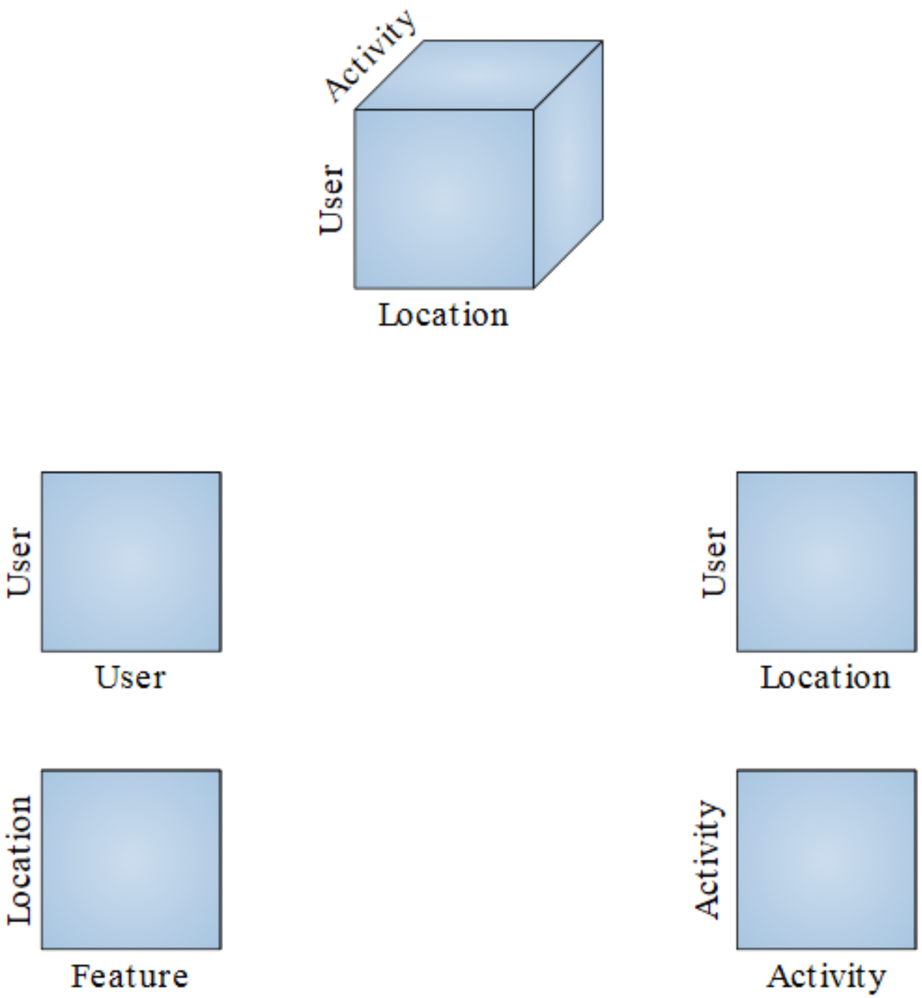
General idea

- Factorising each matrix and tensor, as before
- Constraints to glue factorisations together
  - Equating factors in different factorisations
  - Factorisations 'weakly supervise' / regularise each other

Example: GPS dataset

Zheng et al., 2010 [25]

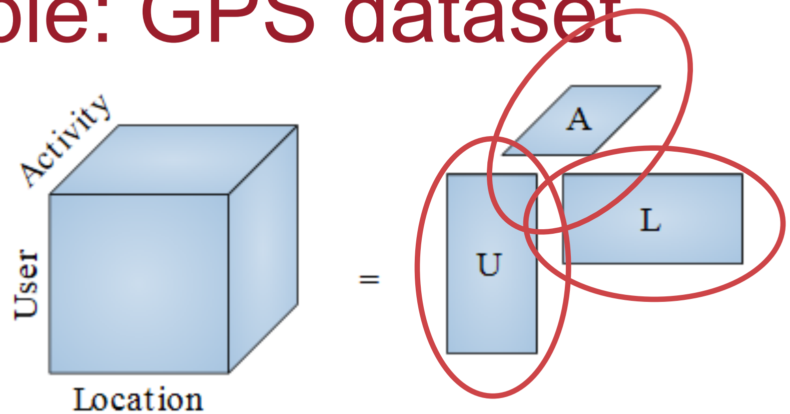
146 users
5 activities
168 locations
14 location features



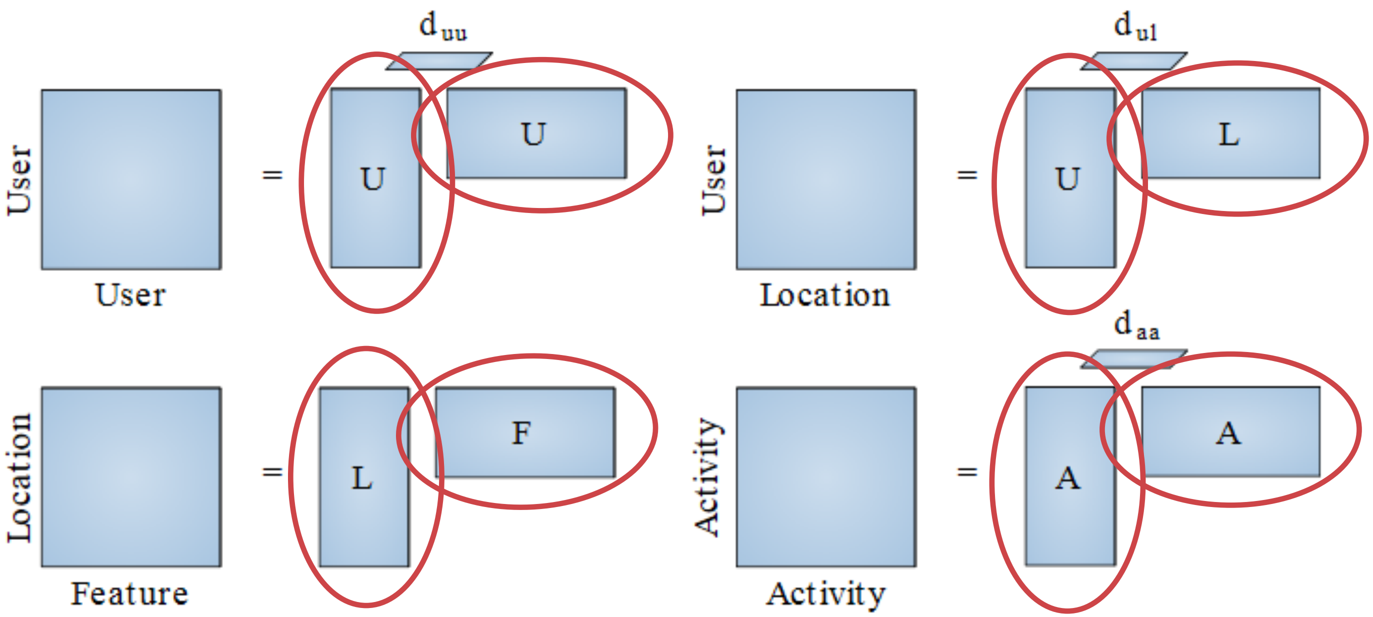
Images taken with permission from the Tensorlab documentation [2,3,4]

Example: GPS dataset

Zheng et al., 2010 [25]



146 users
5 activities
168 locations
14 location features



Images taken with permission from the Tensorlab documentation [2,3,4]

Example: GPS dataset

- Evaluation: Collaborative Filtering
  - Estimate random missing entries in User-Location-Activity tensor
    - Considerable more accurate than using tensor alone
  - Estimate Location-Activity entries for users that are entirely missing in User-Location-Activity tensor
    - Impossible without 'data fusion' (cold start problem)

Research issues

- Pattern syntax
  - Even more degrees of freedom…
  - Regularisation
    - Factorisations regularise each other!
- Interestingness
  - Combined interestingness of different factorisations
- Algorithmic
  - Alternating Least Squares is baseline
  - No guarantees, but computations under control

Summing up…

Summing up…

- Data types
  - Very close to Entity-Relationship data model, and multidimensional data (OLAP!)
- Pattern syntax
  - Very flexible, though *linear* and *global*
  - For completion/prediction, and for insight/exploration
- Interestingness
  - For convenience, often L2-norm
  - Often more appropriate (but harder) alternatives exist
- Algorithmic approach
  - Numerical optimisation
  - No guarantee to find global optimum

References

1. Bro. R. PARAFAC. Tutorial and Applications. *Chemometrics Intell. Lab. Syst.* 1997; 38:149-171.
2. Laurent Sorber, Marc Van Barel and Lieven De Lathauwer. *Tensorlab v2.0*, Available online, January 2014. URL: http://www.tensorlab.net/.
3. Sorber L., Van Barel M., De Lathauwer L., ``Structured data fusion'', *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, Jun. 2015, pp. 586-600.
4. O. Debals, F. Van Eeghem, N. Vervliet, L. De Lathauwer, Tensor computations using Tensorlab, Tech. Report 15-143, ESAT-STADIUS, KU Leuven, Leuven, Belgium, 2015.
5. R.A. Harshman, Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Working Papers in phonetics*, 16 (1970) 1.
6. J.D. Carroll and J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35 (1970) 283.
7. P. Geladi, Analysis of multi-way (multi-mode) data. *Chemom. Intell. Lab. Syst.*, 7 (1989) 11.
8. A.K. Smilde, Three-way analyses. Problems and prospects. *Chemom. Intell. Lab. Syst.*, 5 (1992) 143.
9. P.M. Kroonenburg, Three-mode principal component analysis. Theory and applications. DSWO Press, Leiden, 1983.
10. H.A.L. Kiers, Hierarchical relations among three-way methods. *Psychometrika*, 56 (1991) 449.
11. R.A. Harshman and S.A. Berenbaum, Basic concepts underlying the PARAFAC-CANDECOMP three-way factor analysis model and its application to longitudinal data, in "Present and past in middle life". (Eds. D.H. Eichorn, J.A. Clausen, N. Haan, M.P. Honzik and P.H. Mussen), Academic Press, NY, 1981, 435-459.
12. Kiers, Henk AL. "Hierarchical relations among three-way methods."*Psychometrika* 56.3 (1991): 449-470.
13. Bro, Rasmus. "Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis." *Chemometrics and Intelligent Laboratory Systems* 46.2 (1999): 133-147.

References

14. A. K. Smilde, J. A. Westerhuis, and R. Boque, "Multiway multiblock component and covariates regression models", *Journal of Chemometrics,* 14:301-331, 2000.
15. R. A. Harshman and M. E. Lundy, "Parafac: Parallel factor analysis", *Computational Statistics and Data Analysis,* 18: 39-72, 1994
16. Andersen, C. M. and Bro, R. 2003. Practical aspects of parafac modelling of fluorescence excitation-emission data. J. of Chemometrics 17, 4, 200–215.
17. Miwakeichi, F., Martnez-Montes, E., Valds-Sosa, P., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. 2004. Decomposing eeg data into space-time-frequency components using parallel factor analysis. NeuroImage 22, 3, 1035–1045.
18. Acar, E., Camtepe, S. A., Krishnamoorthy, M., and Yener, B. 2005. Modeling and multiway analysis of chatroom tensors. In Proc. of IEEE International Conference on Intelligence and Security Informatics. Springer, Germany, 256–268.
19. Kolda, T. G., Bader, B. W., and Kenny, J. P. 2005. Higher-order web link analysis using multilinear algebra. In Proc. of The 5th IEEE International Conference on Data Mining. 242–249.
20. Kolda, T. G. and Bader, B. W. 2006. The tophits model for higher-order web link analysis. In Workshop on Link Analysis, Counterterrorism and Security.
21. Sun, J. T., Zeng, H. J., Liu, H., Lu, Y., and Chen, Z. 2005. Cubesvd: A novel approach to personalized web search. In Proc. of the 14th International World Wide Web conference. 382–390.
22. Wang, H. and Ahuja, N. 2004. Compact representation of multidimensional data using tensor rank-one decomposition. In Proc. of International Conference on Pattern Recognition. Vol. 1. 44–47.
23. Vasilescu, M. A. O. and Terzopoulos, D. 2002. Multilinear image analysis for facial recognition. In Proc. of International Conference on Pattern Recognition. Vol. 2. 511–514.
24. Meng, X., Morris, A. J., and Martin, E. B. 2003. On-line monitoring of batch processes using a parafac representation. J. of Chemometrics 17, 1, 65–81.
25. V. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: a user-centered approach. In *AAAI*, volume 10, 236–241. 2010.