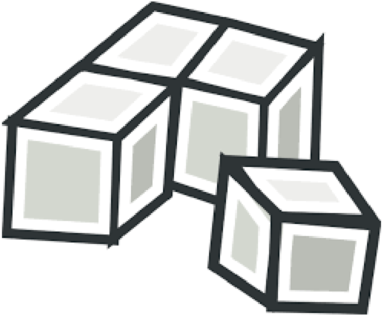
Joint decompositions with flexible couplings

Jeremy E.Cohen
Rodrigo Cabral Farias
Pierre Comon



TDA 2016
01/2016

ERC DecoDa
AdG 320594

Joint decompositions with flexible couplings

Jeremy E.Cohen
Rodrigo Cabral Farias
Pierre Comon

TDA 2016
01/2016

ERC DecoDa
AdG 320594

Joint decompositions with flexible couplings

Jeremy E.Cohen
Rodrigo Cabral Farias
Pierre Comon

TDA 2016
01/2016

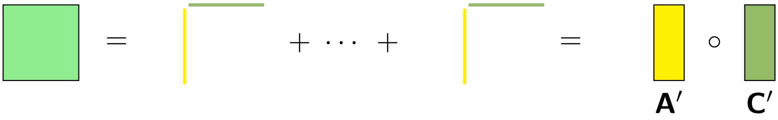
ERC DecoDa
AdG 320594

Plan

1. Joint tensor decompositions
   Motivation, similar factors setting
2. Joint tensor decompositions with flexible couplings
   Bayesian setting, examples, algorithm
3. Simulations
   Similar factors, sampling, nonnegative coupling
4. Joint Compression

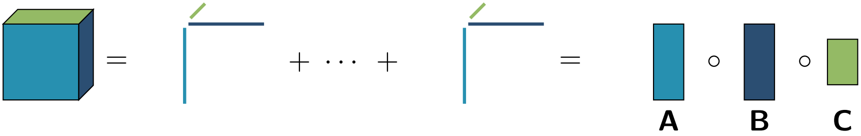
Motivation I - Ill-posed problems

Matrix factorization



Non unique - except under constraints

Tensor decomposition - Canonical polyadic (CP) decomposition



Unique - Kruskal's condition - $2R + D - 1 \leq \sum_d \kappa_d$

$$\mathcal{T}_{EEG} = (\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C})\mathcal{I}_R + \mathcal{E}$$

$$\mathcal{T}_{MEG} = \left(\boldsymbol{A}' \otimes \boldsymbol{B}' \otimes \boldsymbol{C}\right)\mathcal{I}_{R'} + \mathcal{E}$$

Motivation II - Noise removal

Scenario : $\boldsymbol{C} = \boldsymbol{C}'$, $SNR = 25dB \quad SNR' = 65dB$, Uncorrelated $\mathbf{A}$

	Uncoupled (MSE)		Coupled	
	CRB	Sim.	CRB	Sim.
A	0.008	0.008	0.007	0.007
B	0.016	0.016	0.010	0.015
C	0.018	0.018	1.25×10^{-6}	1.09×10^{-6}
A′	1.13×10^{-6}	1.13×10^{-6}	1.13×10^{-6}	1.13×10^{-6}
B′	9.23×10^{-7}	8.88×10^{-7}	9.23×10^{-7}	8.88×10^{-7}
C′	1.25×10^{-6}	1.10×10^{-6}	1.25×10^{-6}	1.10×10^{-6}

Similar factors

Until recently $\mathbf{C} = \mathbf{C}'$

Harshman1984, Banerjee2007, Singh2008, Lin2009, Acar2011/2013a/2013b, Yilmaz2011, Sørensen2013, Sorber2013

What if $\mathbf{C} \approx \mathbf{C}'$?

Seichepine2014 - Penalization $\|\mathbf{C} - \mathbf{C}'\|_F$ or $\|\mathbf{C} - \mathbf{C}'\|_1$
Approximation setting

What if $\mathbf{C}$ is similar to $\mathbf{C}'$ but in a broader sense ?

Bayesian setting

Plan

1. Joint tensor decompositions
   Motivation, hard couplings setting

2. Joint tensor decompositions with flexible couplings
   Bayesian setting, examples, algorithm

3. Simulations
   Similar factors, sampling

4. Joint Compression

Bayesian approach

1. Parameters $\boldsymbol{\theta} = \begin{bmatrix} \text{vec}(\mathbf{A}) \\ \text{vec}(\mathbf{B}) \\ \text{vec}(\mathbf{C}) \end{bmatrix}$ and $\boldsymbol{\theta}' = \begin{bmatrix} \text{vec}(\mathbf{A}') \\ \text{vec}(\mathbf{B}') \\ \text{vec}(\mathbf{C}') \end{bmatrix}$ are random

2. Known prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\theta}')$ and likelihoods $p(\mathcal{Y}|\boldsymbol{\theta})$ and $p(\mathcal{Y}'|\boldsymbol{\theta}')$

MAP estimation under conditionnal independance,

$$\underset{\boldsymbol{\theta},\boldsymbol{\theta}'}{\arg\max}\ p(\boldsymbol{\theta}, \boldsymbol{\theta}'|\mathcal{Y}, \mathcal{Y}') = \underset{\boldsymbol{\theta},\boldsymbol{\theta}'}{\arg\min}\ \Upsilon(\boldsymbol{\theta}, \boldsymbol{\theta}')$$

$$\begin{aligned} \Upsilon(\boldsymbol{\theta}, \boldsymbol{\theta}') &= -\log p(\mathcal{Y}|\boldsymbol{\theta}) - \log p(\mathcal{Y}'|\boldsymbol{\theta}') - \log p(\boldsymbol{\theta}, \boldsymbol{\theta}') \\ &= \quad \text{data fitting terms} \quad + \quad \text{coupling} \end{aligned}$$

Joint Gaussian modeling

Joint model

$$\mathbf{M}\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\theta}'\end{bmatrix} = \boldsymbol{\Sigma}\boldsymbol{u} + \boldsymbol{\mu}$$

$\mathbf{M}$ has left-inverse, $\boldsymbol{u} \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma}$ is a diagonal covariance matrix

Joint Gaussian distribution

$$\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\theta}'\end{bmatrix} \sim \mathcal{N}\{{}^{\dagger}\boldsymbol{M}\boldsymbol{\mu}, \boldsymbol{\Gamma}\}, \text{ where } \boldsymbol{\Gamma} = ({}^{\dagger}\boldsymbol{M})\boldsymbol{\Sigma}\boldsymbol{\Sigma}({}^{\dagger}\boldsymbol{M}^{\top})$$

MAP

$$\Upsilon\left(\boldsymbol{\theta}, \boldsymbol{\theta}'\right) = \sum_{k=1}^{2} \|\boldsymbol{\mathcal{T}}_k - (\boldsymbol{A}_k \otimes \boldsymbol{B}_k \otimes \boldsymbol{C}_k)\mathcal{I}_{R_k}\|^2 + \|\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\theta}'\end{bmatrix} - {}^{\dagger}\boldsymbol{M}\boldsymbol{\mu}\|^2_{\boldsymbol{\Gamma}^{-1}}$$

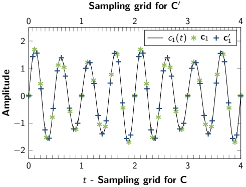
Joint Gaussian modeling : Example

Coupling the third factors, with some non-informative priors on the other factors

$$\begin{bmatrix} \frac{1}{\sigma_A^2}\boldsymbol{I} & & & & & \\ & \frac{1}{\sigma_B^2}\boldsymbol{I} & & & & \\ & & \frac{1}{\sigma_c^2}\boldsymbol{I} & & & -\frac{1}{\sigma_c^2}\boldsymbol{I} \\ & & & \frac{1}{\sigma_{A'}^2}\boldsymbol{I} & & \\ & & & & \frac{1}{\sigma_{B'}^2}\boldsymbol{I} & \\ & & & & & \frac{1}{\sigma_{C'}^2}\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \text{vec}(\boldsymbol{A}) \\ \text{vec}(\boldsymbol{B}) \\ \text{vec}(\boldsymbol{C}) \\ \text{vec}(\boldsymbol{A}') \\ \text{vec}(\boldsymbol{B}') \\ \text{vec}(\boldsymbol{C}') \end{bmatrix} = \boldsymbol{u}$$

This yields $\text{vec}(\boldsymbol{C}) \sim \mathcal{N}\left(\text{vec}(\boldsymbol{C}'), \frac{1}{\sigma_c^2}\boldsymbol{I}\right)$, i.e. $\boldsymbol{C} = \boldsymbol{C}' + \epsilon$.

Hybrid Gaussian modeling



Coupling with linear transformations

$$\boldsymbol{HC} \approx \boldsymbol{H}'\boldsymbol{C}'$$

Sampling the same continuous function

$$c^r(t) \approx \sum_{k=1}^{K} \mathbf{c}_k^r h(t, t_k) \approx \sum_{k'=1}^{K'} \mathbf{c'}_{k'}^r h'(t, t_{k'})$$

where $\boldsymbol{H}_{lk} = h(t_l, t_k)$, $\boldsymbol{H}'_{lk'} = h(t_l, t_{k'})$ and $\{t_l, l \in 1, \cdots, L\}$

Gaussian model

$$\begin{bmatrix} & \cdots & \cdots \\ \mathbf{0} & \textbf{diag blk}(\boldsymbol{H}) & -\textbf{diag blk}(\boldsymbol{H}') \end{bmatrix} \begin{bmatrix} \vdots \\ \text{vec}(\boldsymbol{C}) \\ \text{vec}(\boldsymbol{C}') \end{bmatrix} = \boldsymbol{\Sigma u}$$

ALS - Hybrid Gaussian case

$$\Upsilon(\theta,\theta') = \sum_{k=1}^{2} \frac{1}{\sigma_k^2} \|\mathcal{Y}_k - (\boldsymbol{A}_k \otimes \boldsymbol{B}_k \otimes \boldsymbol{C}_k)\mathcal{I}_{R_k}\|_F^2 + \frac{1}{\sigma_c^2}\|\boldsymbol{H}_1\boldsymbol{C}_1 - \boldsymbol{H}_2\boldsymbol{C}_2\|_F^2$$

Uncoupled Factors Estimation

$$\hat{\mathbf{A}}_1 = \boldsymbol{Y}_1^{[1]}(\hat{\mathbf{B}}_1 \odot \hat{\mathbf{C}}_1)^\dagger \quad \hat{\mathbf{A}}_1 = \boldsymbol{Y}_2^{[1]}(\hat{\mathbf{B}}_2 \odot \hat{\mathbf{C}}_2)^\dagger,$$

$$\hat{\mathbf{B}}_1 = \boldsymbol{Y}_1^{[2]}(\hat{\mathbf{A}}_1 \odot \hat{\mathbf{C}}_1)^\dagger \quad \hat{\mathbf{B}}_1 = \boldsymbol{Y}_2^{[2]}(\hat{\mathbf{A}}_2 \odot \hat{\mathbf{C}}_2)^\dagger,$$

Coupled Factors Estimation : Sylvester Equations

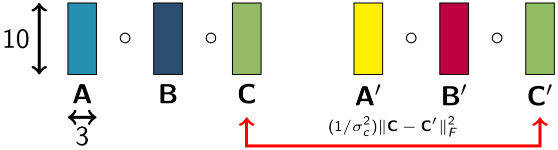
$$\begin{cases} \boldsymbol{H}_1^T\boldsymbol{H}_1\hat{\boldsymbol{C}}_1 & + & \hat{\boldsymbol{C}}_1\boldsymbol{F}_1^T\boldsymbol{F}_1 & - & \boldsymbol{H}_1^T\boldsymbol{H}_2\hat{\boldsymbol{C}}_2 & = & \boldsymbol{Y}_1^{[3]}\boldsymbol{F}_1 \\ \boldsymbol{H}_2^T\boldsymbol{H}_2\hat{\boldsymbol{C}}_2 & + & \hat{\boldsymbol{C}}_2\boldsymbol{F}_2^T\boldsymbol{F}_2 & - & \boldsymbol{H}_2^T\boldsymbol{H}_1\hat{\boldsymbol{C}}_1 & = & \boldsymbol{Y}_2^{[3]}\boldsymbol{F}_2 \end{cases}$$

where $\boldsymbol{F}_1 = (\hat{\mathbf{A}}_1 \odot \hat{\mathbf{B}}_1)$ and $\boldsymbol{F}_2 = (\hat{\mathbf{A}}_2 \odot \hat{\mathbf{B}}_2)$

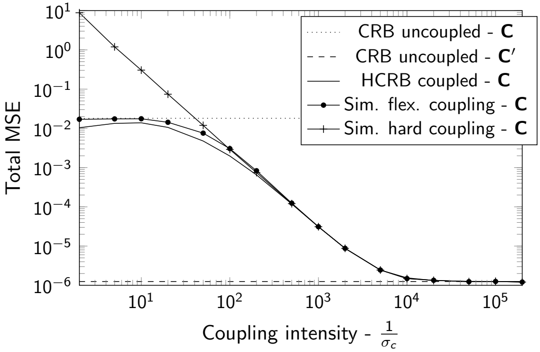
Plan

Similar factors

Low SNR
$\sigma_n = 0.1$

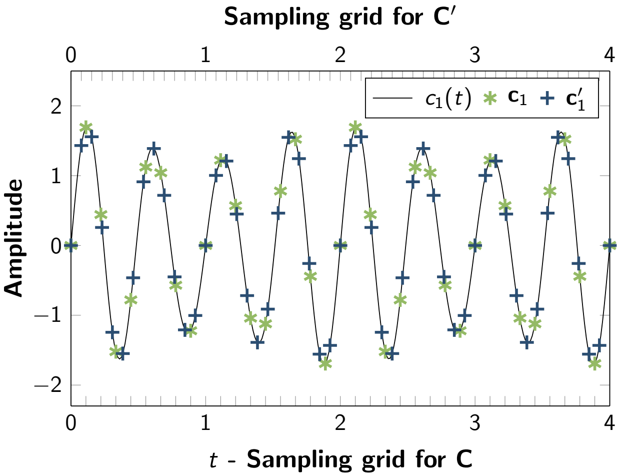


High SNR
$\sigma'_n = 0.001$

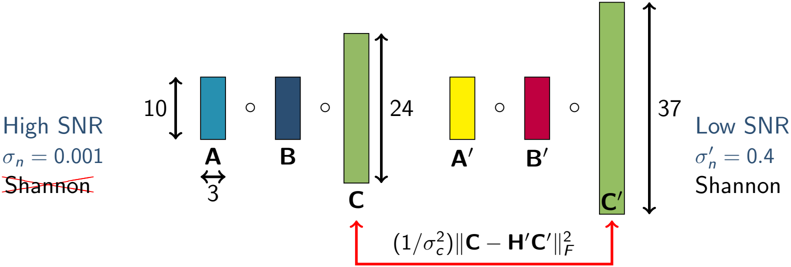


Sampling

Bandlimited periodic signal



Sampling III



Total MSE on the continuous functions (numerical integration)

	C ~~Shannon~~	**C′** noisy
Uncoupled	33.4968	2.6581
Coupled	33.4968	1.0375

Plan

1. Joint tensor decompositions
   Motivation, similar factors setting

2. Joint tensor decompositions with flexible couplings
   Bayesian setting, examples, algorithm

3. Simulations
   Similar factors, sampling, nonnegative coupling

4. Joint Compression

Disjoint Compression

$$\boldsymbol{C} = \boldsymbol{C}'$$
$$\boldsymbol{C}_c := \hat{\boldsymbol{W}}^T \boldsymbol{C}$$
$$\boldsymbol{C}'_c := \hat{\boldsymbol{W}}'^T \boldsymbol{C}'$$
$$\hat{\boldsymbol{W}} \boldsymbol{C}_c \neq \hat{\boldsymbol{W}}' \boldsymbol{C}'_c$$

Indeed, $\boldsymbol{C}$ and $\boldsymbol{C}'$ do not respectively belong to the column space of $\hat{\boldsymbol{W}}$ and $\hat{\boldsymbol{W}}'$.

Joint compression

By choosing the same basis for the coupled mode, the coupling relationship is still exact even though the basis is estimated from noisy data.

$$\boldsymbol{C} = \boldsymbol{C}'$$

$$\boldsymbol{C}_c := \hat{\boldsymbol{W}}_j^T \boldsymbol{C}$$

$$\boldsymbol{C}'_c := \hat{\boldsymbol{W}}_j^T \boldsymbol{C}'$$

$$\boldsymbol{C}_c = \boldsymbol{C}'_c$$

$\boldsymbol{W}_j$ can be defined as the basis of the concatenated mode $[\boldsymbol{C}, \boldsymbol{C}']$:

$$TrunSVD\left(\left[\frac{\boldsymbol{T}_{(3)}}{\sigma_n}, \frac{\boldsymbol{T}'_{(3)}}{\sigma'_n}\right]\right) = \boldsymbol{W}_j \boldsymbol{N}_j$$

Conclusions

- Bayesian setting allows naturally to define flexible couplings
- Flexible couplings: explore transition between exactly coupled and uncoupled models
- Flexible couplings: allow to fuse heterogeneous data
  - Different noise levels (Trivial)
  - Different subsets of latent variables (Standard Coupled CP)
  - Similar but different coupled factors ([Seichepine2014])
  - Coupled factors with different nature (Tweedie)
  - Coupled factors with different sizes (Sampling rates)
- Other points:
  - Algorithm for nonnegative coupling
  - Compression of coupled data
  - Performance (Bayesian and Hybrid CRB)

Exploring multimodal data fusion through joint decompositions with flexible couplings - arXiv:1505.07717

Thank you for your attention!

Examples: non Gaussian coupling

When $\mathbf{C} > 0$ and $\mathbf{C}' > 0$, why Gaussian coupling ?

Coupling with Tweedie's distribution (strong coupling)

$$p(\boldsymbol{C}_{ij}|\boldsymbol{C}'_{ij}) \approx (2\pi\phi\boldsymbol{C}_{ij}^{\beta})^{-1/2}\exp[-d_\beta(\boldsymbol{C}_{i,j}|\boldsymbol{C}'_{i,j})/\phi]$$

where $d_\beta(\boldsymbol{C}_{i,j}|\boldsymbol{C}'_{i,j})$ is the β-divergence

Types of coupling

$\beta = 1$ - Poisson, $\beta = 2$ - Gamma, $\beta = 3$ - Inverse Gaussian
($\beta \to 0$ - Gaussian as a degenerated case)

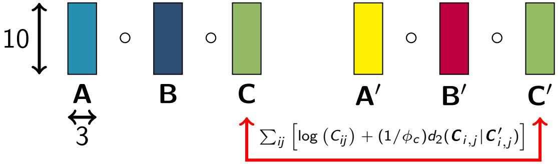
MAP

$$\Upsilon\left(\boldsymbol{\theta},\boldsymbol{\theta}'\right) =$$
$$-\log p(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}) - \log p(\boldsymbol{\mathcal{Y}}'|\boldsymbol{\theta}') + \sum_{ij}\left[(\beta/2)\log\left(C_{ij}\right) + (1/\phi)d_\beta(\boldsymbol{C}_{i,j}|\boldsymbol{C}'_{i,j})\right]$$

Nonnegative coupling

Gamma × Gamma × Gamma model

Gamma $\phi_n = 0.5$



Gamma $\phi'_n = 0.05$

Multiplicative update algorithm - $\phi_c = 0.05$

	Uncoupled (MSE)	Coupled
	Sim.	Sim.
A	0.041	0.015
B	0.054	0.021
C	4.904	0.803
A′	0.001	0.001
B′	0.001	0.001
C′	0.129	0.127

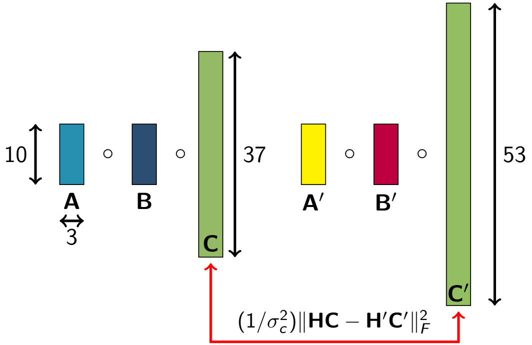
Sampling

Interpolation kernels - Dirichlet kernels

$$\boldsymbol{H}_{lk} = \frac{\sin\{K\pi[(l-1)T_i-(k-1)T]/[(L-1)T_i\}}{K\sin\{\pi[(l-1)T_i-(k-1)T]/[(L-1)T_i\}} \text{ - } K = 37 \text{ - } T = 1/9s$$

$$\boldsymbol{H}'_{lk'} = \frac{\sin\{K'\pi[(l-1)T_i-(k'-1)T']/[(L-1)T_i\}}{K'\sin\{\pi[(l-1)T_i-(k'-1)T']/[(L-1)T_i\}} \text{ - } K' = 53 \text{ - } T' = 1/13s$$

Same duration - 4s



References

- [Harshman1984] R. A. Harshman and M. E. Lundy. Data preprocessing and the extended PARAFAC model. Research Methods for Multimode Data Analysis, pages 216-284, 1984.
- [Banerjee2007] A. Banerjee, S. Basu, and S. Merugu. Multi-way clustering on relation graphs. In SDM, volume 7, pages 225-334. SIAM, 2007.
- [Singh2008] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. In Proc. 14th ACM SIGKDD Conf., pages 650-658. ACM, 2008.
- [Lin2009] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, and A. Kelliher. Metafac: community discovery via relational hypergraph factorization. In Proc. ACM SIGKDD Conf., pages 527-536. ACM, 2009.

References

- [Acar2011] E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. arXiv preprint arXiv:1105.3422, 2011.
- [Acar2013a] E. Acar, A. J. Lawaetz, M. A. Rasmussen, and R. Bro. Structure-revealing data fusion model with applications in metabolomics. In Conf. Proc. IEEE Eng. Med. Biol. Soc., pages 6023-6026. IEEE, 2013.
- [Acar2013b] E. Acar, M. A. Rasmussen, F. Savorani, T. Ns, and R. Bro. Understanding data fusion within the framework of coupled matrix and tensor factorizations. Chemometr. Intell. Lab., 129:53-63, 2013.
- [Yilmaz2011] K. Y. Yilmaz, A. T. Cemgil, and U. Simsekli. Generalised coupled tensor factorisation. In Adv. Neural. Inf. Process. Syst., pages 2151-2159, 2011.

References

- [Sørensen2013] M. Sørensen and L. De Lathauwer. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-($L_{r,n}$, $L_{r,n}$, 1) terms. Part I: Uniqueness/Part II: Algorithms. Technical report, KU Leuven (Belgium), 2013.
- [Sorber2013] L. Sorber, M. Van Barel, and L. De Lathauwer. Structured data fusion. Tech. Rep., KU Leuven (Belgium), pages 13-177, 2013.

Figures

- Tofu cubes: http://4vector.com/free-vector/tofu-cubes-clip-art-105437.
- Monolith discovery in "2001 a space odissey": MGM, Stanley Kubrick Productions.
- MEG: http://www.psy.cmu.edu/ lholt/php/researchMethods.php.