Fast structured matrix factorization with an application to football heatmaps

Lorenzo Schiavon & Antonio Canale & Bernardo Nipoti

Statistical methods and models for complex data — Poster Session: 21 September 2022

Football player heatmaps

Football heatmaps are **graphical representations** of the **intensity of a football player action**, measured in different location **over the pitch**.

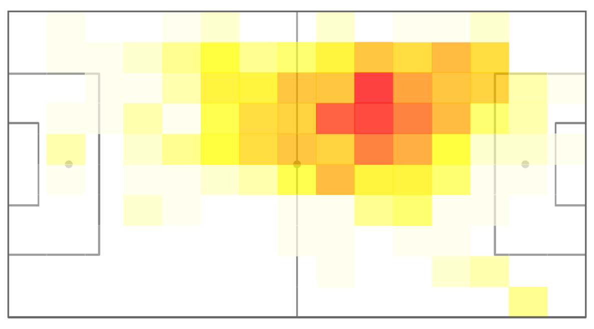

Figure 1. Illustrative heatmap of the distance run by a football player in different areas of the pitch during a match.

We can represent an heatmap as a p-variate **vector**, with p the number of **cells in which the pitch is divided**.

Then, a **collection** of n heatmaps is a $n\times p$ **matrix** y, where **dependence** structure cannot be excluded in any of the **two dimensions**.

Goal...of the project

Modeling the dependence between any couple of elements of the data matrix y_{ij} and y_{ls}, exploiting **exogenous information on the similarity** between players i and l, and the **spatial relation** between the pitch cells j and s.

Data

We have a $n\times p$ data matrix y of $\mathbf{n} = \mathbf{106}$ **heatmaps** of 106 different players collected over five professional football matches by MathAndSport s.r.l.

Each heatmap is represented by a vector of $\mathbf{p} = \mathbf{150}$ **cells** in which we divide the pitch. The element y_{ij} reports the **distance covered** by the player i within the cell j during the match.

A $n \times c$ **covariate matrix** $\boldsymbol{x}$, informing on **player** characteristics, as **expected role and position** during the match, is available.

We also exploit a $p\times m$ **meta covariate matrix** $\boldsymbol{w}$ including information on the **pitch cells location** to induce **spatial dependence**.

Matrix decomposition

Factor models

Factor models express a statistical object of interest in terms of a **collection of simpler objects**. For example, a matrix $\mathbf{y}$ can be expressed as a transformation $f(z)$ of a **sum of** k **rank-one factors**

$$y_{ij} = f(z_{ij}), \quad z_{ij} = \sum_{h=1}^{\infty} \eta_{ih}\,\lambda_{jh} + \epsilon_{ij},$$

In **matrix decomposition** we take care of both **dependence among columns** and **dependence among rows**:

$$z_{ij} = \sum_{h=1}^{\infty} \eta_{ih}(x_i)\,\lambda_{jh}(w_j) + \epsilon_{ij},$$

with x_i a covariate vector and w_j a meta covariate vector, including information on subjects and column entities.

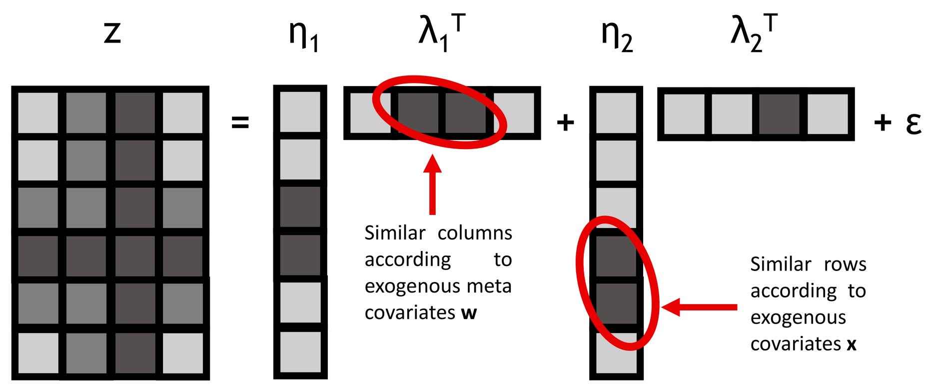

Figure 2. Latent data decomposition in two factors.

X-FILE algorithm

X-FILE, Accelerated Factorization via Infinite Latent Elements, is our novel point-wise estimation algorithm. **Regularized estimates** are obtained by using a **Bayesian prior** as **penalization** and then optimizing the posterior:

$$\underset{\mathcal{P}}{\text{argmin}} - \log\{\mathcal{L}(\mathbf{y};\boldsymbol{\mathcal{P}},\boldsymbol{\Sigma},\mathbf{x},\mathbf{w})\} - \log\{\text{pr}(\mathcal{P})\},$$

−loglikelihood: **loss function**,
−logprior: **penalty function**.

Forward stage-wise additive maximization ($\sim$ **boosting**): given $\mathbf{h-1}$ **terms fixed** we **sequentially** estimate a new factor $\eta_h\lambda_h^\top$, such that

$$\underset{\{\eta,\lambda\}}{\text{argmin}} \log\{\mathcal{L}(z;\sum_{\mathrm{l=1}}^{\mathrm{h-1}}\boldsymbol{\eta}_l\boldsymbol{\lambda}_l^\top + \eta\lambda^\top, x, w)\} + \sum_{\mathrm{l=1}}^{\mathrm{h-1}} \log\{\text{pr}(\boldsymbol{\eta}_l\boldsymbol{\lambda}_l^\top)\} + \log\{\text{pr}(\eta,\lambda)\},$$

.

Structured prior penalty

$$\eta_{ih} \sim N\{0, \psi_{ih}\}, \qquad \lambda_{jh} \sim N\{0, \theta_h\phi_{jh}\},$$
$$E(\psi_{ih} \mid \beta_h) \propto g_x(\boldsymbol{x_i}\beta_h), \qquad E(\phi_{jh} \mid \gamma_h) \propto g_w(\boldsymbol{w_j}\gamma_h),$$

Where ψ_{ih} and ϕ_{jh} are local scales and θ_h is a factor-specific scale.

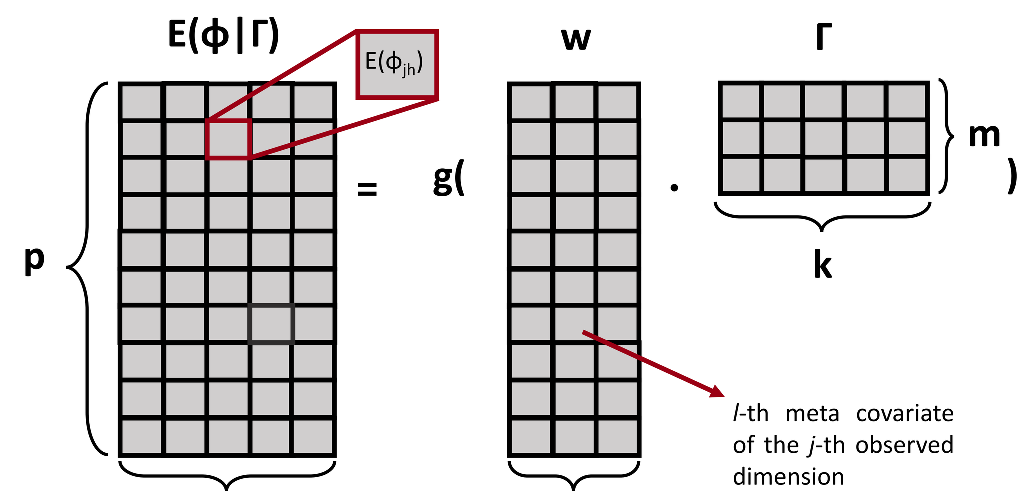

Figure 3. A representation of the mean equation of the loadings local scale.

The **local scale means depend** on exogenous information through a transformation of a linear combination of **covariates** x and **meta covariates** w.

Thus, local shrinkage is tuned also on the basis of prior knowledge.

Factor scale is decomposed as $\theta_h = \vartheta_h\rho_h$.

- $\vartheta_h^{-1} \sim \text{Ga}(a_\theta, b_\theta)$ **i.i.d. power law tail distribution**, such that:
  - λ_{jh} estimation is **robust** to large signals.
  - it represents a **dynamic learning rate** regulating the impact of each additive step of X-FILE, where **small** ϑ_h induces a **better fit** and **large** ϑ_h induces a **fast algorithm** and an easier interpretation.
- $\rho_h \sim \text{Ber}(1-\pi_h)$ with **increasing probability** π_h of being **zero**, such that:
  - the increasing shrinkage allows for **infinite factors**;
  - it provides a simple **stopping rule** for the X-FILE **algorithm**, by adding new factors only if they increase the log-posterior of the model. The algorithm stops at step $h-1$ if

$$\log\{\text{pr}(\rho_h = 1)\} + l_{ij}^{(\rho_h=1)} < \log\{\text{pr}(\rho_h = 0)\} + l_{ij}^{(\rho_h=0)}$$

where $l_{ij}^{(\rho_h=1)}$ and $l_{ij}^{(\rho_h=0)}$ are the maximum log-likelihood of z_{ij} under $\rho_h = 1$ and $\rho_h = 0$, respectively, with $\rho_{h+1} = \rho_{h+2}\ldots = 0$.

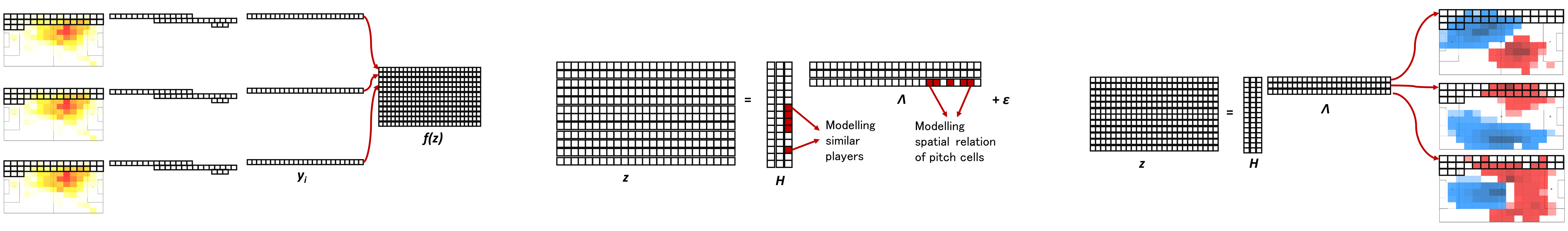

Figure 4. Data elaboration pipeline: **1-** player tracking **data are transformed into heatmaps**, that are **vectorized** and stored in a matrix y; **2-** the latent matrix z is **decomposed** through H and Λ by considering **prior knowledge** to induce **structures**; **3-** the Λ matrix estimated by the **X-FILE algorithm** is represented in the form of a collection of **archetypal heatmaps**.

Application results

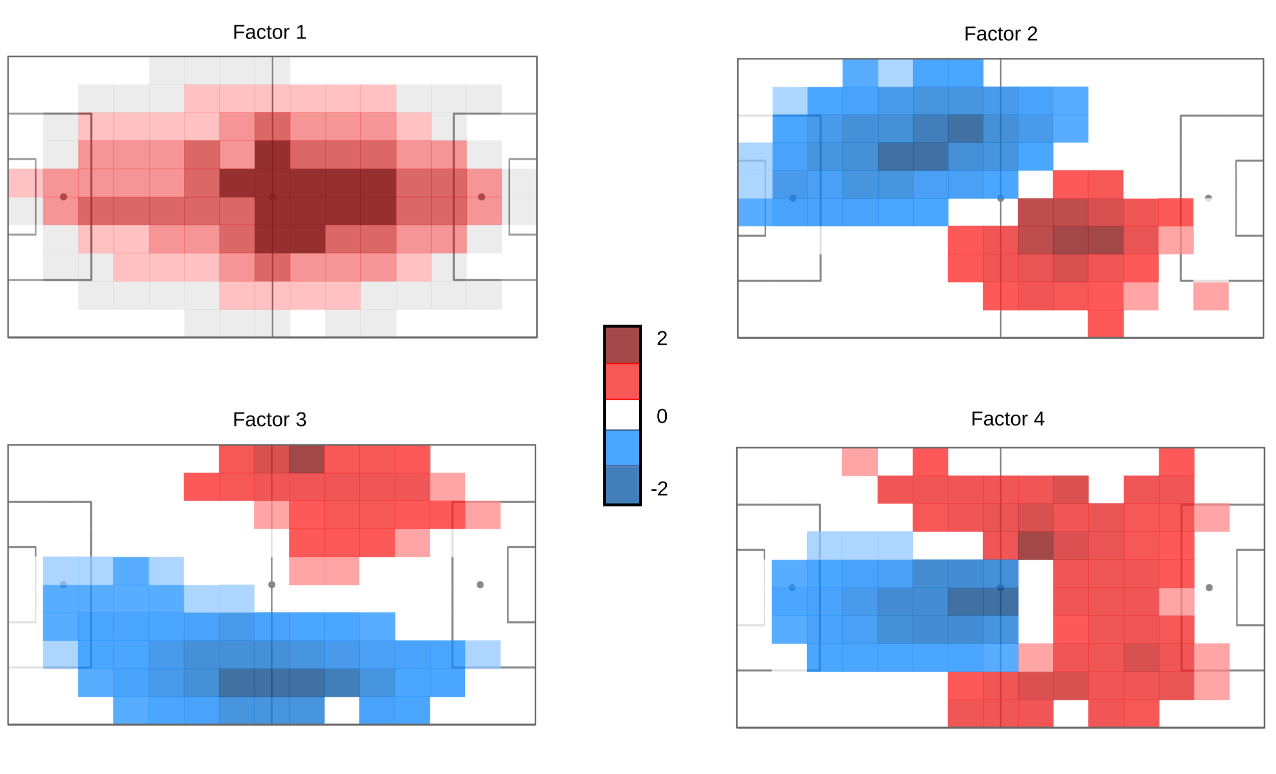

Figure 5. Archetypal heatmaps obtained by representing the columns of the estimated Λ.

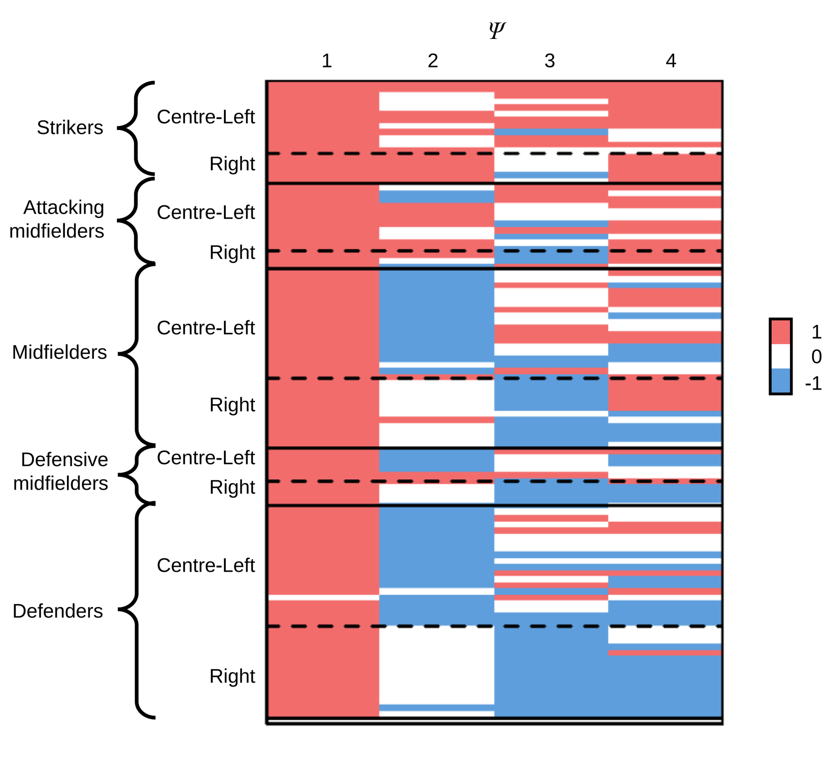

Figure 6. Shrinkage structure of the H matrix, inducing a three-group player clustering in every factor.

Main references

A. Bhattacharya and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika*.

J. Friedman, T. Hastie and R. Tibshirani (2000). Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*.

S. Legramanti, D. Durante, and D. B. Dunson (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika*.

A. Mnih and R. R. Salakhutdinov (2008). Probabilistic matrix factorization. In *Advances in neural information processing systems*.

L. Schiavon, A. Canale, and D. B. Dunson (2022). Generalized infinite factorization models. *Biometrika*.

Contact information

Lorenzo Schiavon, Post-doc research fellow
Department of Statistical Sciences, University of Padova
lorenzo.schiavon@unipd.it
https://lorenzo-schiavon.github.io/
https://github.com/lorenzo-schiavon/