# Time Series Models: From Statistics to AI

lecturer: Daniel Durstewitz
tutors: Lukas Eisenmann, Christoph Hemmer, Alena Brändle, **Florian Hess**, Elias Weber, Florian Götz
SS2025

## Exercise Sheet 3

### Regulations

Please submit your solutions via Moodle in teams of **2 students**, before the exercise group on **Wednesday, May 14th, 2025**. Each submission must include **exactly** two files:

- A `.pdf` file containing both your Jupyter notebook and solutions to analytical exercises. The Jupyter notebook can be exported to pdf by selecting **File → Download as → pdf** in JupyterLab. If this method does not work, you may print the notebook as a pdf instead. Your analytical solutions can be either scanned handwritten solutions or created using LaTeX.
- A `.ipynb` file containing your code as Jupyter notebook.

Both files must follow the naming convention:
`Lastname1-Lastname2-sheet03.pdf`
`Lastname1-Lastname2-sheet03.ipynb`

**Task 1. Kullback-Leibler divergence of two multivariate normal distributions**
The Kullback–Leibler (KL) divergence of two continuous distributions $\mathcal{P}$ and $\mathcal{Q}$ is defined as:

$$KL(\mathcal{P} \parallel \mathcal{Q}) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx.$$

Assume $\mathcal{P}$ and $\mathcal{Q}$ are two multivariate normal distributions $\mathcal{P} = \mathcal{N}(\mu_a, \Sigma_a)$ and $\mathcal{Q} = \mathcal{N}(\mu_b, \Sigma_b)$ where $\mu_a, \mu_b \in \mathbb{R}^n$ and $\Sigma_a, \Sigma_b \in \mathbb{R}^{n\times n}$.

1. Derive the analytical expression for the KL divergence given the distributions above.
(**Hint**: Consult The Matrix Cookbook for useful identities to simplify the expressions.)
2. How does the KL divergence simplify if we assume the covariances to be diagonal, i.e. $\Sigma_a = diag([\sigma_{a,1}, \sigma_{a,2}, \ldots, \sigma_{a,n}])$ and $\Sigma_b = diag([\sigma_{b,1}, \sigma_{b,2}, \ldots, \sigma_{b,n}])$?

**Task 2. M-Step in a linear Gaussian state space model**

Consider a linear Gaussian state space model of the form

$$\begin{aligned} z_t &= Az_{t-1} + \epsilon, & \epsilon &\sim \mathcal{N}(0, \Sigma) \\ x_t &= Bz_t + \eta, & \eta &\sim \mathcal{N}(0, \Gamma) \end{aligned}$$

with parameters $\{A, B, \Sigma, \Gamma\}$. In the lecture we derived the M-step to determine the transition matrix $A$.

1. Derive the expressions for all remaining parameters, i.e. $\{B, \Sigma, \Gamma\}$. Reduce all expressions such that they only contain expectations in $z$ (e.g. $\mathbb{E}[z_t], \mathbb{E}[z_t z_{t-1}^T]$, ...) and known quantities such as observations $x_t$.
(**Hint**: You may want to consult The Matrix Cookbook for this exercise, too.)