IMPERIAL

Department of Mathematics

Probabilistic Sequential Matrix Factorisation for 12-lead ECG Data

Joana Levtcheva

CID: 01252821

Supervised by Dr Deniz Akyildiz

August 4, 2024

Submitted in partial fulfilment of the requirements for the
MSc in Machine Learning and Data Science of
Imperial College London

The work contained in this thesis is my own work unless otherwise stated.

Signed: Joana Levtcheva Date: August 4, 2024

Abstract

Matrix factorisation (MF) techniques are highly effective and widely used in unsupervised machine learning. By decomposing the original matrix into multiple simpler lower-dimensional matrices, MF aims to uncover latent structures that are not immediately obvious in the original matrix. MF finds applications in areas such as image processing, natural language processing, missing data imputation, and recommendation systems. Despite considerable progression in the probabilistic versions, there is demand for such methods in applications such as uncertainty quantification, managing time-series data, and executing efficient probabilistic components.

In this thesis, we show novel applications of the probabilistic sequential MF algorithm Probabilistic Sequential Matrix Factorisation (PSMF) (Ömer Deniz Akyildiz et al. (2021)) to 12-lead ECG data. We explore three tasks related to this complex high-dimensional time-series data with nonlinear subspace: missing data imputation with PSMF and the robust version of PSMF (rPSMF) (Ömer Deniz Akyildiz et al. (2021)) and compare their performance with other probabilistic sequential MF algorithms, R-peaks detection, and forecasting an ECG component based on previous normal heart beats using a Fourier basis with multiple terms and rank higher than one. We perform our experiments using the high-quality comprehensive dataset "A Large Scale 12-lead Electrocardiogram Database for Arrhythmia Study" (Zheng (2022), Zheng et al. (2020), Goldberger et al. (2000)). We describe and outline the experimenting process and the challenges we encountered modelling the complex ECG data, and summarise the experiments results. We find that PSMF performs well when used for imputing missing data, but when it comes to forecasting there are certain challenges which open the door for further research on extending the PSMF algorithm to better handle the complex structure of ECG data.

Acknowledgements

TODO

Contents

1 Introduction

Matrix factorisation (or also matrix decomposition) in the context of linear algebra is simply a factorisation of a matrix into a product of multiple matrices. Many different decompositions exist, and they find various applications in mathematical problems such as solving systems of linear equations, matrix inversion, determinant computation, eigenvalues problems, solving systems of first order ODEs, etc.

In this thesis, we are interested in matrix factorisation (MF) in the context of machine learning. Nowadays, MF techniques are highly effective and widely used in unsupervised machine learning. These methods aim to decompose the original matrix into multiple lower-dimensional matrices. By breaking the matrix into these simpler components MF aims to uncover latent structures that are not immediately apparent in the original matrix. Some applications are in image processing: for reducing dimensionality and noise in images, NLP for topic modelling, missing data imputation, recommendation systems, etc.

Formally, we are interested in the general problem of factorising a data matrix $Y \in \mathbb{R}^{m \times n}$ as

$$Y \approx CX, \tag{1.1}$$

where $C \in \mathbb{R}^{m \times r}$ is the *dictionary matrix*, $X \in \mathbb{R}^{r \times n}$ is the *coefficients matrix* (with columns the coefficients), and r is the *approximation rank* (Akyildiz and Míguez (2019)). Visually we can present the problem as

$$\underbrace{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}}_{\mathrm{Y} \in \mathbb{R}^{m \times n}} \approx \underbrace{\begin{bmatrix} \times & \times \\ \times & \times \\ \times & \times \end{bmatrix}}_{\mathrm{C} \in \mathbb{R}^{m \times r}} \underbrace{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}}_{\mathrm{X} \in \mathbb{R}^{r \times n}}. \tag{1.2}$$

There are also probabilistic versions of MF which incorporate probabilistic models to better handle uncertainty and variability in the data, leading to more accurate predictions. Such methodologies postulate a prior distribution over the latent factors and necessitate the computation of the posterior distribution to derive updated estimates. With that the matrix is not only decomposed but probabilistic interpretations of the factors are also possible.

We should also note that some algorithms are suitable for sequential data - updating C

and X incrementally as new data points are observed and thus incorporating temporal dynamics and sequential dependencies into the factorisation process, and others are non-sequential - treating the dataset as a batch, independent of the time varying component.

Throughout this thesis we are going to focus on probabilistic sequential MF algorithms, along with their application to 12-lead ECG data, targeting the problem of managing high-dimensional time-series data with nonlinear subspace. Some examples of probabilistic MF algorithms are Probabilistic Matrix Factorisation (PMF) (Mnih and Salakhutdinov (2007)) - non-sequential, Dictionary filtering (Akyildiz and Míguez (2019)), Probabilistic Sequential Matrix Factorization (PSMF) (Ömer Deniz Akyildiz et al. (2021)).

The paper "Probabilistic matrix factorisation" (PMF) (Mnih and Salakhutdinov (2007)) introduces an efficient and scalable probabilistic model for collaborative filtering. The algorithm performs well on large, sparse and imbalanced datasets. This is demonstrated by using a Netflix dataset, where PMF models the user preference matrix R as a product of two lower-dimensional matrices: user feature matrix U and movie feature matrix V. The conditional distribution over observed ratings is modeled using Gaussian noise, and zero-mean spherical Gaussian priors are placed on the user and movie feature vectors. The paper also presents two extensions to the initial PMF model: incorporating *adaptive priors* to automatically control the model complexity through these priors over the model parameters, and a *constrained PMF* version to handle and improve predictions for users with few ratings by incorporating constraints based on the assumption that users with similar movie ratings have similar preferences. The authors show that PMF significantly outperforms traditional Singular Value Decomposition (SVD) (Stewart (1993)) models and scales linearly with the number of observations. It's worth noting that PMF treats each rating as an independent event meaning the time varying component is not taken into consideration, making PMF a batch learning model designed to process large datasets in a non-sequential manner.

Later, in the paper "Dictionary Filtering: A Probabilistic Approach to Online Matrix Factorization" (DF) (Akyildiz and Míguez (2019)), the authors introduce a novel online MF algorithm known as dictionary filtering. It leverages probabilistic models, specifically using recursive linear filters, and efficiently factorises the original data matrix into a dictionary matrix and a coefficients matrix. This is an online and sequential algorithm, meaning it is suitable for high-dimensional and time-varying data, and it also has easy to tune parameters. DF is efficient for high-dimensional data with its computational complexity of $O(mr^2)$ independent of the number of data points. Although the model can learn non-stationary and dynamic data, it is developed for linear and Gaussian state space models (SSM). Particularly for ECG data, ECG has a nonlinear SSM which doesn't suit the dictionary filtering.

Two years later, Akyildiz et.al. develop Probabilistic Sequential Matrix Factorization (PSMF) (Ömer Deniz Akyildiz et al. (2021)). This method is tailored to time-varying and non-stationary datasets consisting of high-dimensional time-series. Nonlinear Gaus-

sian SSMs are considered, decomposing the original matrix into a dictionary matrix and time-variying coefficient matrix. This time, the matrices are with potentially nonlinear dependencies, with PSMF efficiently capturing temporal dependencies through Markovian structures on the coefficients, making it possible to encode the dependencies into a lower dimensional latent space. The model is demonstrated to work on tasks such as forecasting, changepoint detection, missing data imputation, and is shown to work on real-world data with a periodic subspace. There is also a robust version rPSMF using Student-t filters to handle model misspecification, and a version for imputing missing data. Although the model is suitbale for reducing high-dimensional data with periodic subspaces to lower-dimensional latent space, PSMF might struggle with very large datasets, having many data points.

1.1 Contributions

Using probabilistic methods, and specifically probabilistic sequential MF ones, on ECG data is not widely explored. Hence, we aim to introduce a few novel real-world applications of the PSMF method to 12-lead ECG data. The contributions made in this thesis are:

- We show three novel applications of PSMF to complex 12-lead ECG data.
- Application 1: We apply both PSMF and rPSMF for imputing missing data in the ECG signals, and compare the results with other sequential probabilistic MF models.
- Application 2: We show that PSMF can be used for R-peaks detection by introducing a simple approach. We remove the reconstructed signal, which has modelled the R-peaks smoother than the real ones, from the original data, and determine a suitable threshold for selecting the peaks.
- Application 3: We forecast an ECG component based on previous normal heart beats by incorporating a Fourier basis with multiple Fourier terms and rank higher than 1.

1.2 Notation

We are going to denote the original data matrix as Y, and let $Y \in \mathbb{R}^{m \times n}$. Let $C \in \mathbb{R}^{m \times r}$ be the dictionary matrix, $X \in \mathbb{R}^{r \times n}$ be the coefficients matrix, and r be the approximation rank.

With $I_d \in \mathbb{R}^{d \times d}$ we are going to denote the identity matrix, with $MN(X; M, U, V)$ the matrix normal distribution with M the mean-matrix, U the row-covariance, and V the column covariance. Further, with $N(x; \mu, \Sigma)$ let's denote the Gaussian density with

mean μ and Σ the covariance matrix, $IG(s; \alpha, \beta)$ is the inverse gamma distribution with shape α and scale β.

TODO:

- the mapping f_θ
- vectorisation
- multivariate normal distribution

2 Background

2.1 Preliminaries

2.1.1 Matrix Normal Distribution

TODO

2.1.2 Kronecker Product

Definition 2.1.1. Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$ be matrices. Then their *Kronecker product* denoted as $A \otimes B \in \mathbb{R}^{mp \times nq}$ is given by (Harville (1997))

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}, \tag{2.1}$$

where $\{a_{ij}\}$, $i \in \{1, ..., m\}$, $j \in \{1, ..., n\}$ are the elements of A.

2.1.3 Fourier Series

TODO

2.2 PSMF

Intro

2.2.1 Model

For observations $(y_k)_{k \geq 1} \in \mathbb{R}^m$, latent coefficients $(x_k)_{k \geq 1} \in \mathbb{R}^r$, and a dictionary matrix $C \in \mathbb{R}^{m \times r}$ the PSMF model can be described with the following state-space equations:

$$p(C) = \mathcal{MN}(C; C_0, I_d, V_0) \tag{2.2}$$

$$p(x_0) = \mathcal{N}(x_0; \mu_0, P_0) \tag{2.3}$$

$$p_\theta(x_k \mid x_{k-1}) = \mathcal{N}(x_k; f_\theta(x_{k-1}), Q_k) \tag{2.4}$$

$$p(y_k \mid x_k, C) = \mathcal{N}(y_k; Cx_k, R_k) \tag{2.5}$$

where ...

We are going to refer to Equation (2.2) as the dictionary prior, Equation (2.3) as the initial state of the coefficients, Equation (2.4) as the transition density, and Equation (2.5) as the observation model.

2.2.2 Parameter Estimation?

2.2.3 Inference

Prediction ...

$$p(x_k \mid y_{1:k-1}) = \int p(x_{k-1} \mid y_{1:k-1}) p(x_k \mid x_{k-1}) dx_{k-1} \tag{2.6}$$

Update: After computing the predictive distribution of x_k, the update steps for x_k and C are given by:

(Incremental) marginal likelihood:

$$p(y_k \mid y_{1:k-1}) = \int \int p(y_k \mid C, x_k) p(x_k \mid y_{1:k-1}) p(C \mid y_{1:k-1}) dx_k dC \tag{2.7}$$

Dictionary update:

$$p(C \mid y_{1:k}) = p(C \mid y_{1:k-1}) \frac{p(y_k \mid C, y_{1:k-1})}{p(y_k \mid y_{1:k-1})}, \tag{2.8}$$

where ...

Coefficient update:

$$p(x_k \mid y_{1:k}) = p(x_k \mid y_{1:k-1}) \frac{p(y_k \mid x_k, y_{1:k-1})}{p(y_k \mid y_{1:k-1})}, \tag{2.9}$$

where ...

TODO: approximate seq. inference

2.2.4 PSMF Algorithm

WIP

Algorithm 1 Iterative PSMF

Initialize γ, θ_0, C_0, V_0, μ_0, P_0, $(Q)_k \geq 1$, $(R)_k \geq 1$.
for $i \geq 1$ **do**
  $C_0 = C_n$, $\mu_0 = \mu_n$, $P_0 = P_n$, $V_0 = V_n$
  for $1 \leq k \leq n$ **do**
    Compute predictive mean of x_k:
      $\bar{\mu}_k = f_{\theta_{i-1}}(\mu_{k-1})$ or $\bar{\mu}_k = f_{\theta_k - 1}(\mu_{k-1})$
    Compute predictive covariance of x_k:
      $\bar{P}_k = F_k P_{k-1} F_k^\top + Q_k$, with $F_k = \frac{\partial f(x)}{\partial x}\Big|_{x=\bar{\mu}_{k-1}}$
    Update dictionary mean C_k using
    Update dictionary covariance V_k with
    Update coefficient mean μ_k using
    Update coefficient covariance P_k with
  Update parameters with

2.3 rPSMF

Intro

2.3.1 Model

The robust variant, rPSMF, incorporates robustness to outliers and model misspecifications by integrating a heavy-tailed t-distribution into the framework.

$$p(s) = \mathcal{IG}(s; \lambda_0/2, \lambda_0/2) \tag{2.10}$$

$$p(C \mid s) = \mathcal{MN}(C; C_0, I_d, sV_0) \tag{2.11}$$

$$p(x_0 \mid s) = \mathcal{N}(x_0; \mu_0, sP_0) \tag{2.12}$$

$$p_\theta(x_k \mid x_{k-1}, s) = \mathcal{N}(x_k; f_\theta(x_{k-1}), sQ_0) \tag{2.13}$$

$$p(y_k \mid x_k, C, s) = \mathcal{N}(y_k; Cx_k, sR_0), \tag{2.14}$$

where ...

2.3.2 Inference

Inference in rPSMF incorporates the handling of the robust parameters and scale mixing variable s introduced to manage outliers and model misspecifications.

TODO: with or wotihout derivation?

2.3.3 rPSMF Algorithm

WIP

Algorithm 2 Iterative rPSMF

Initialize $\gamma, \theta_0, C_0, V_0, \mu_0, P_0, Q_0, R_0$.
for $i \geq 1$ **do**
  $C_0 = C_T$, $\mu_0 = \mu_T$, $P_0 = P_T$, $V_0 = V_T$
  for $1 \leq k \leq T$ **do**
    Predictive mean of x_k: $\bar{\mu}_k = f_{\theta_{i-1}}(\mu_{k-1})$ or $\bar{\mu}_k = f_{\theta_{k-1}}(\mu_{k-1})$
    Predictive covariance of x_k:
    $\bar{P}_k = F_k P_{k-1} F_k^\top + Q_k, \quad$ where $\quad F_k = \frac{\partial f(x)}{\partial x}\Big|_{x=\bar{\mu}_{k-1}}$
  Compute scaling factor for the dictionary update
  $\varphi_k = \frac{\lambda_{k-1}}{\lambda_{k-1}+d} + \frac{(y_k - C_{k-1}\bar{\mu}_k)^\top (y_k - C_{k-1}\bar{\mu}_k)}{\bar{\mu}_k^\top V_{k-1}\bar{\mu}_k + \eta_k}$
    where $\eta_k = \mathrm{Tr}(C_{k-1}\bar{P}_k C_{k-1}^\top + R_{k-1})/d$.
  Mean and covariance updates of the dictionary
  $C_k = C_{k-1} + \frac{(y_k - C_{k-1}\bar{\mu}_k)\bar{\mu}_k^\top V_{k-1}}{\bar{\mu}_k^\top V_{k-1}\bar{\mu}_k + \eta_k}$ and $\quad V_k = \varphi_k \left(V_{k-1} - \frac{V_{k-1}\bar{\mu}_k\bar{\mu}_k^\top V_{k-1}}{\bar{\mu}_k^\top V_{k-1}\bar{\mu}_k + \eta_k}\right)$
  Compute scaling factor for the coefficient update
  $\omega_k = \lambda_{k-1} + (y_k - C_{k-1}\bar{\mu}_k)^\top S_k^{-1}(y_k - C_{k-1}\bar{\mu}_k)$
    where $S_k = C_{k-1}\bar{P}_k C_{k-1}^\top + \bar{R}_k$ and $\quad \bar{R}_k = R_{k-1} + \bar{\mu}_k^\top V_{k-1}\bar{\mu}_k \otimes I_d$.
  Mean and covariance updates of coefficients
  $\mu_k = \bar{\mu}_k + \bar{P}_k C_{k-1}^\top S_k^{-1}(y_k - C_{k-1}\bar{\mu}_k)$ and $\quad P_k = \omega_k(\bar{P}_k - \bar{P}_k C_{k-1}^\top S_k^{-1} C_{k-1}\bar{P}_k)$
  Update noise covariances: $Q_k = \omega_k Q_{k-1}$ and $R_k = \omega_k R_{k-1}$
  Update degrees of freedom: $\lambda_k = \lambda_{k-1} + d$
  Parameter update: $\theta_k = \theta_{k-1} + \gamma \nabla_\theta \log p_\theta(y_k|y_{1:k-1})|_{\theta=\theta_{i-1}}$
  Parameter update: $\theta_i = \theta_{i-1} + \gamma \sum_{k=1}^{T} \nabla_\theta \log p_\theta(y_k|y_{1:k-1})|_{\theta=\theta_{i-1}}$

2.4 PSMF for Handling Missing Data

2.4.1 Model

$$p(C) = \mathcal{MN}(C; C_0, I_d, V_0) \quad \text{(Matrix-variate Gaussian prior on dictionary)} \tag{2.15}$$

$$p(x_0) = \mathcal{N}(x_0; \mu_0, P_0) \quad \text{(Gaussian prior on initial coefficients)} \tag{2.16}$$

$$p_\theta(x_k|x_{k-1}) = \mathcal{N}(x_k; f_\theta(x_{k-1}), Q_k) \quad \text{(State transition model)} \tag{2.17}$$

$$p(z_k|x_k, C) = \mathcal{N}(z_k; M_k C x_k, M_k R_k M_k^T), \quad \text{(Observation model)} \tag{2.18}$$

where ...

TODO

2.4.2 Inference

TODO

2.4.3 Algorithm

TODO

2.5 ECG/medical introduction

WIP

The ECG is a crucial diagnostic tool in clinical practice. It is especially useful in diagnosing rhythm disturbances, changes in electrical conduction, and myocardial ischemia and infarction.

A 12-Lead ECG (Electrocardiogram) is a standard diagnostic tool used to assess the electrical activity of the heart. It provides a comprehensive view of the heart's electrical activity from different angles. The 12 leads consist of:

- 6 precordial leads (V1-V6) placed on the chest
- 3 limb leads (I, II, III)
- 3 augmented limb leads (aVR, aVL, aVF)

The ECG machine typically records data for about 10 seconds, producing a graph of voltage versus time.

add images sources

TODO:

- introduce PQRST
- mention R-peaks and their importance
- general information about the 12-leads
- add images for ECG, leads placements, sources

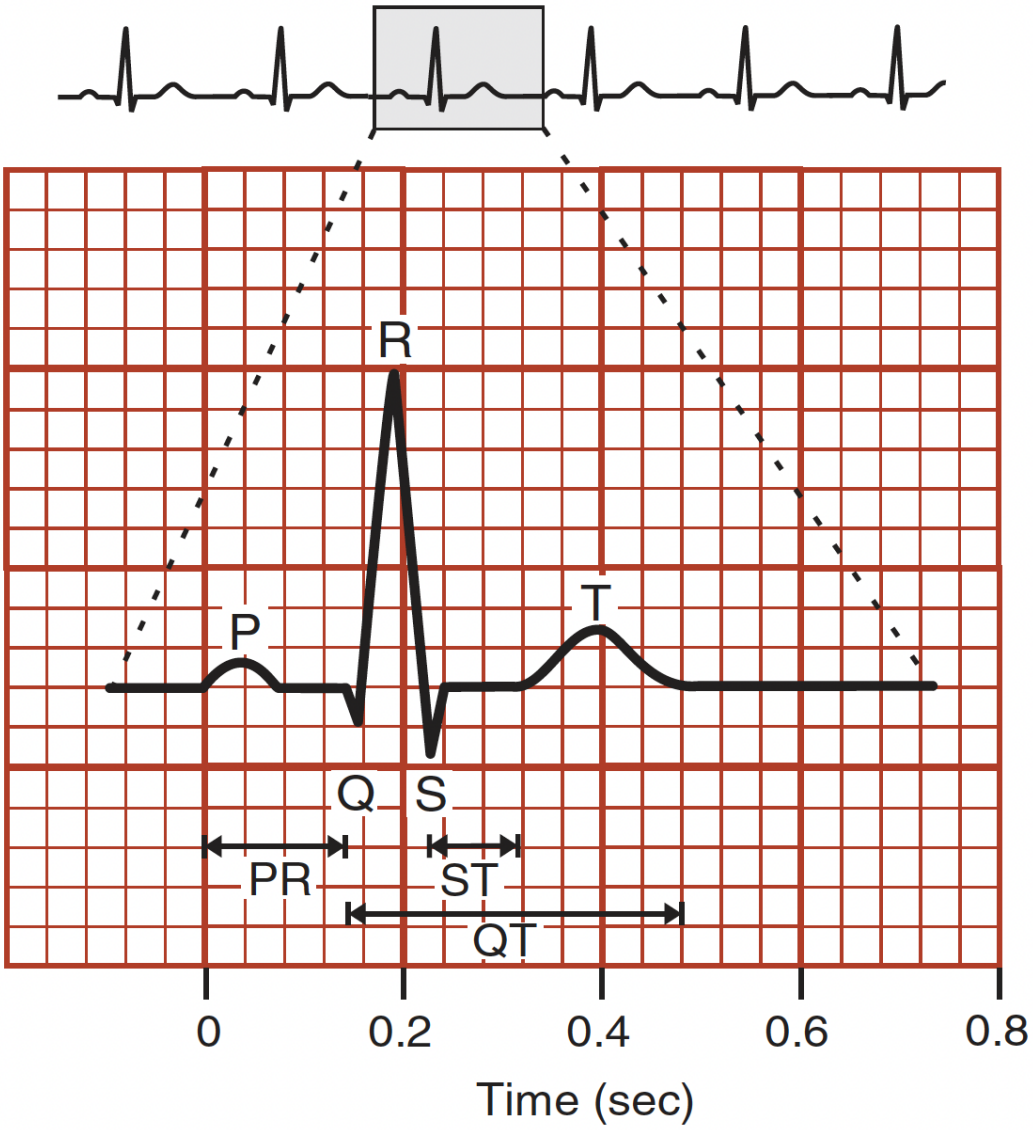


Figure 2.1: Components of the ECG trace. An enlargement of one of the repeating waveform units in the rhythm strip shows the P wave, QRS complex, and T wave, which represent atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively. The PR interval represents the time required for the depolarization wave to transverse the atria and the AV node; the QT interval represents the period of ventricular depolarization and repolarization; and the ST segment is the isoelectric period when the entire ventricle is depolarized. Each small square is 1 mm. (Klabunde (2012 - 2012))

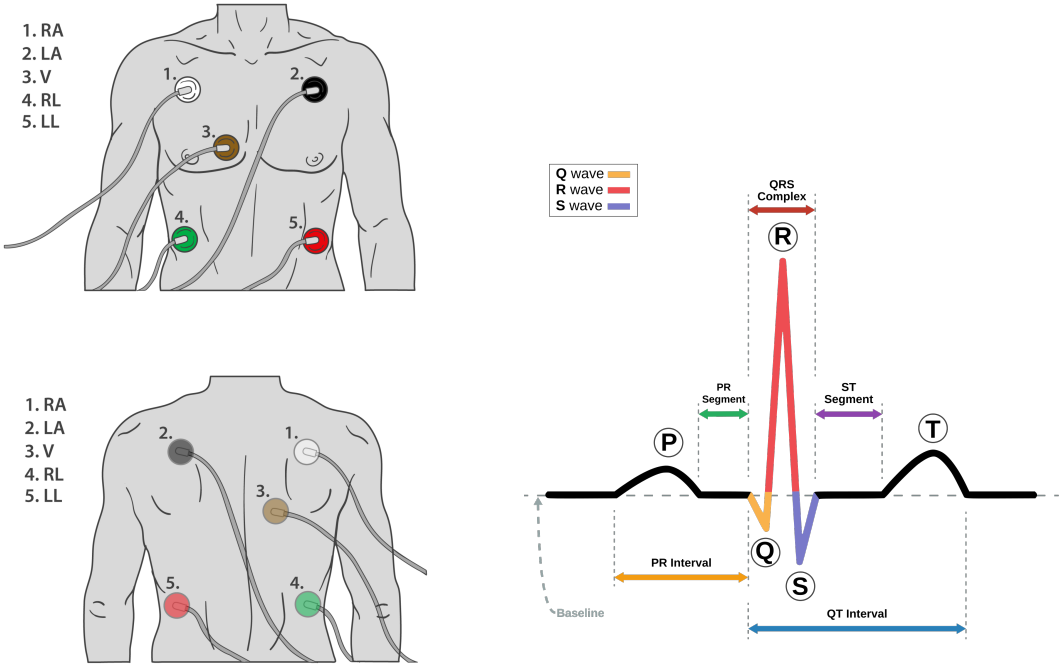


Figure 2.2: Left: . Right:

3 Experiemnts and Results

The research aims to use the already available A Large Scale 12-lead Electrocardiogram Database for Arrhythmia Study (Zheng (2022), Zheng et al. (2020), Goldberger et al. (2000)). This is a comprehensive database of high-quality 12-lead ECG signals collected from 45,152 patients. Each signal is with length of 10 seconds corresponding to 5000 data points. The dataset is designed to support arrhythmia research, containing labeled data for various cardiac conditions such as atrial fibrillation, premature ventricular contractions, and bundle branch blocks. This large dataset has high-quality labels from professional experts, diverse arrhythmia types, and additional cardiovascular conditions, making it suitable for performing different research tasks on it without spending time on gathering, cleaning and processing data.

3.1 Missing Data Imputation

TODO:

- data used - processing
- algorithms tested: psmf, rpsmf, mle-smf, tmf
- coverage, results
- add figures showing the original data, missing parts and imputed parts

	20%	30%	40%
PSMF	0.39	0.32	0.24
rPSMF	**0.85**	**0.78**	**0.71**
MLE-SMF	0.18	0.16	0.15

Table 3.1: Average coverage proportion of the missing data by the 2σ uncertainty bars of the posterior predictive estimates, averaged over 100 repetitions.

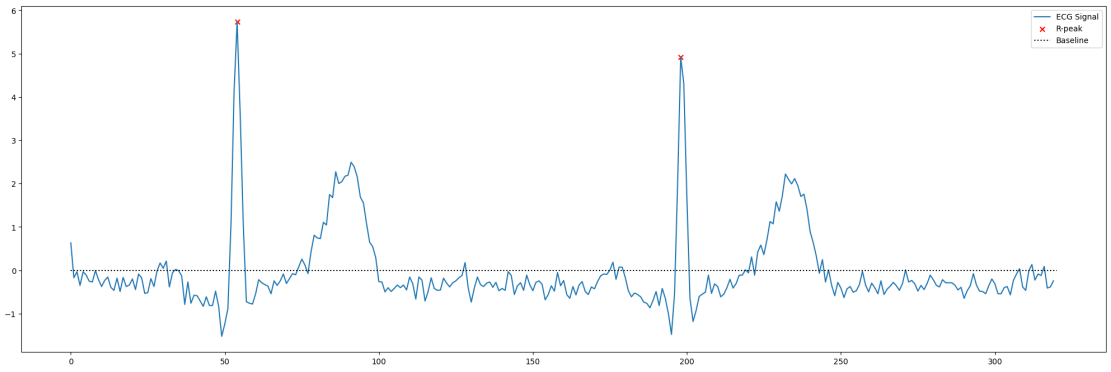
	Imputation RMSE			Runtime (s)		
	20%	30%	40%	20%	30%	40%
PSMF	63.93 (20.80)	80.90 (22.15)	104.68 (21.30)	1.05	1.25	1.19
rPSMF	**60.58** (17.29)	**74.56** (18.50)	**93.22** (20.12)	1.22	1.31	1.17
MLE-SMF	255.87 (22.15)	263.76 (67.00)	271.45 (81.31)	6.17	1.00	1.12
TMF	169.83 (15.78)	167.91 (16.33)	155.58 (17.91)	2.11	0.50	0.58

Table 3.2: Imputation error and runtime using 20%, 30% and 40% missing values, averaged over 100 random repetitions.

3.2 R-peaks Detection

TODO:

- intro: why it is an important task, mention current methods?
- introduce the idea: smoother peaks, remove reconstruction from the original data
- threshold
- (potential) issues, improvements, results
- figure caption, more result figures



3.3 Forecasting

TODO:

- original data (5000) vs subsample (1500)

- smoothing vs no smoothing
- Fourier basis and number of terms, higher rank
- encountered issues, conclusions
- add figures for the basis, loss curves, forecast (show different experiments results)

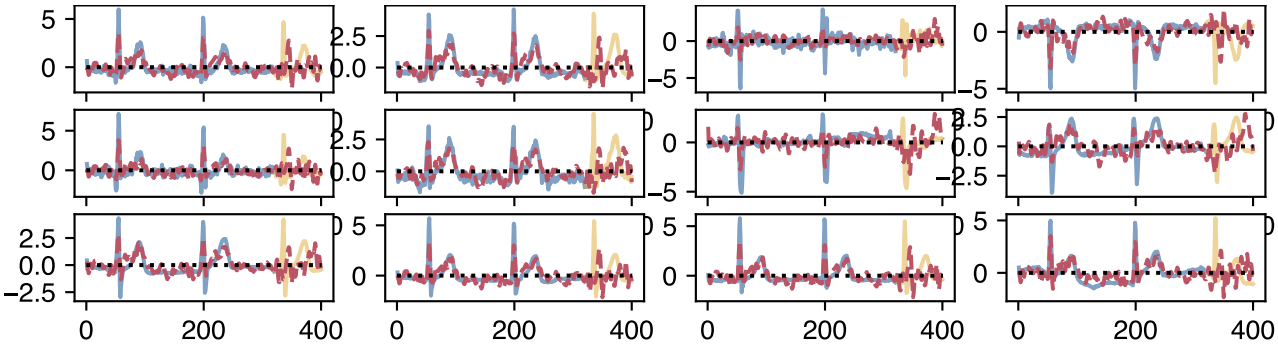


Figure 3.1: My caption here

4 Conclusion

TODO: summarise results

4.1 Future work

TODO:

- More complex model to better suit the complexity of the ECG data
- Better computational efficiency

Bibliography

Ömer Deniz Akyildiz and Joaquín Míguez. Dictionary filtering: a probabilistic approach to online matrix factorisation. *Signal, Image and Video Processing*, 13(4):737–744, 2019.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*, 101(23): e215–e220, 2000.

David A. Harville. *Matrix algebra from a statistician's perspective*. Springer, New York ;, 1997. ISBN 038794978X.

Richard E. Klabunde. *Cardiovascular physiology concepts*. Wolters Kluwer Health, Philadelphia, second edition. edition, 2012 - 2012. ISBN 9781451113846.

Andriy Mnih and Russ R Salakhutdinov. Probabilistic matrix factorization. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

G. W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35(4):551–566, 1993. ISSN 00361445, 10957200. URL `http://www.jstor.org/stable/2132388`.

Guo H. Chu H. Zheng, J. A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0.0). *PhysioNet*, 2022. URL `https://doi.org/10.13026/wgex-er52`.

Jianwei Zheng, Huimin Chu, Daniele Struppa, Jianming Zhang, Sir Magdi Yacoub, Hesham El-Askary, Anthony Chang, Louis Ehwerhemuepha, Islam Abudayyeh, Alexander Barrett, Guohua Fu, Hai Yao, Dongbo Li, Hangyuan Guo, and Cyril Rakovski. Optimal multi-stage arrhythmia classification approach. *Scientific Reports*, 10(1):2898, 2020.

Ömer Deniz Akyildiz, Gerrit J. J. van den Burg, Theodoros Damoulas, and Mark F. J. Steel. Probabilistic sequential matrix factorization, 2021.