A Novel Approach To Estimate Functional Gaussian Graphical Model Based On Penalized Multivariate Functional Regression Model

Salvatore Latora, Luigi Augugliaro and Gerda Claeskens

Abstract In this article we contribute to the literature on the functional Gaussian graphical model by introducing a new penalty function for the multivariate functional regression model that allows us to have a direct connection between the estimated parameters and conditional dependencies among the functional curves, i.e. the edge set of the graphical model. Introducing the model in infinite dimensional space, we show how to estimate the model in a finite-dimensional subspace. A focus is also given on how the bases and scores for the Karhunen-Loéve expansion of the curves are obtained. The performance of the proposed method is evaluated by a simulation study and compared with that of other better-known models in the literature, showing good performance.

1 Introduction

Many authors have contributed to the literature of Gaussian Graphical Models (GGM) for functional data. One of the first works is those of [4] in which, since the covariance operator is compact and therefore non-invertible, the connection between the inverse covariance operator and conditional independence is lost, the authors assume that the functional curves are concentrated to a finite-dimensional subspace by working with an $K-$trunkated version of the Karnhunen-Loève expansion. Therefore,

Salvatore Latora
Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze Ed. 13, Palermo - Italy, e-mail: salvatore.latora@unipa.it

Luigi Augugliaro
Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze Ed. 13, Palermo - Italy, e-mail: luigi.augugliaro@unipa.it

Gerda Claeskens
ORSTAT - Faculty of Economics and Business, KU Leuven, Naamsestraat 69, Leuven - Belgium, e-mail: gerda.claeskens@kuleuven.be

they estimate a finite-dimensional graphical model by a modified lasso criterion. The main problem with this approach is that the authors prove the asymptotic properties in the finite setting, but they are unsatisfied with the infinite-dimensional space of curves. This approach is overcome by [3] which introduces the assumption of *partial separability* of the covariance operator. This allows to have a direct connection between conditional independence and conditional covariance function. Consequently, the functional graphical model is given by a sequence of finite-dimensional Gaussian graphical models. However, the partial separability assumption is a very strong assumption that is not verified in many real-world applications. In [2] the authors provide an edge set learning strategies based on neighborhood selection for functional GGM. After obtaining the neighbourhood set of each node, the final edge set will be obtained by using an **AND/OR** scheme. In this article we contributed to the literature on functional GGM by introducing a novel penalty function that, by working with a neighbourhood selection framework for functional data, allows a direct connection between the matrix of estimated parameters of the regression model and the edge set.
The paper is organized as follows: in the second section the methodological proposal is explained; in the third section a simulation study will be shown; finally, conclusions are provided in the final section.

2 Proposal Methodology

Before introducing the methodological innovation brought by this paper, it is appropriate to give some notations and definitions.

2.1 Notation and Definitions

Let $f_j \in L^2_{[0,1]}$, for $j = 1, \dots, p$ be a random function, where p denote the number of random functions and $L^2_{[0,1]}$ denote the space of square-integrable functions endowed with the inner product $\langle f_1, f_2 \rangle = \int_0^1 f_1(t) f_2(t) dt$ and associated norm $\|f\| = \sqrt{\langle f, f \rangle}$. We denote with $\boldsymbol{f} = (f_1, \dots, f_p)^T \in \mathcal{F} = \left[L^2_{[0,1]}\right]^p$ the curves in a p-dimensional Hilbert space equipped with inner product $\langle \boldsymbol{f}, \boldsymbol{g} \rangle_p = \sum_{j=1}^p \langle f_j, g_j \rangle$ and associated norm $\|\boldsymbol{f}\|_p = \sqrt{\langle \boldsymbol{f}, \boldsymbol{f} \rangle_p}$. A Functional Gaussian Graphical Model is defined as $\{\boldsymbol{f}, MPG(\mu, \mathcal{G}), G\}$, where the multivariate functional data $\boldsymbol{f}$ follows a *multivariate Gaussian process* $MPG(\mu, \mathcal{G})$ specified by a mean function $\mu(t)$ and a covariance operator $\mathcal{G}(t, s)$, and $G = (V, E)$ is an undirected graph defined by V the node set in which each node is a functional curve f_j, and $E \subset V \times V$ is the edge set. E encodes the presence or absence of conditional independence among the functional curves: $f_j \perp\!\!\!\perp f_h \mid f_l$ if $(h, k) \in E$. The multivariate Gaussian process is an assumption we will carry throughout the paper. We also assume that $\mu_j(t) = E\left[f_j(t)\right] = 0$ and $V\left[f_j(t)\right] < \infty$, for $j = 1, \ldots, p$.

2.2 Proposed Methodology

The goal of the proposed model is to recover the edge set E by using a neighbourhood selection approach similar in [2], based on *multivariate functional linear regression* ([6]) by introducing a novel penalty function that allows us to have e direct connection between the parameter estimates and E. In particular, assuming we have n units, for each target node $j \in V$ we have that

$$f_{ij}(t) = \sum_{k \neq j}^{p} \int_0^1 f_{ik}(t)\beta_{jk}(t,t')dt' + \epsilon_{ij}(t), \quad \forall\, j = 1, \ldots, p, \; i = 1, \ldots, n, \tag{1}$$

where $\beta_{jk}(t,t')$ is a functional parameter that captures the effect of the functional curve k to the response j, and $\epsilon_{ij}(t)$ is a functional error term. As in [6] the functional parameter $\beta_{jk}(t,t')$ admits the following expansion

$$\beta_{jk}(t,t') = \sum_{l,l'=1}^{+\infty} b_{jkll'}\phi_l(t)\phi_{l'}(t'), \tag{2}$$

where $b_{jkll'} = \int \beta_{jk}(t,t')\phi_l(t)\phi_{l'}(t')$ and $\{\phi_l\}_{l=1}^{\infty}$ is a *complete orthonormal system* (CONS) of $\left[L^2_{[0,1]}\right]^p$, with $\phi_l = \left(\phi_{1l}, \ldots, \phi_{pl}\right)^T$, such that $\mathcal{G}(t,t') = \sum_{l=1}^{\infty} \gamma_l \phi_l(t)\phi_{l'}(t')$, where $\{\gamma_l\}_{l=1}^{\infty}$ are its non-negative and decreasign eigenvalues. By setting (1) in a multivariate penalized functional regression, the problem becomes

$$\hat{\boldsymbol{\beta}}_{.1}, \ldots, \hat{\boldsymbol{\beta}}_{.p} = \underset{\boldsymbol{\beta}_{.1},\ldots,\boldsymbol{\beta}_{.p}}{\arg\min} \sum_{i=1}^{n}\sum_{j}^{p} \|f_{ij}(t) - \sum_{k\neq j}^{p}\langle f_{ik}, \beta_{jk}(t,\cdot)\rangle\|^2 + \lambda_n P(\boldsymbol{\beta}_{.k}), \tag{3}$$

where $P(\boldsymbol{\beta}_{.k})$ is a proper penalty function. In particular, we introduce the following penalty function

$$P(\boldsymbol{\beta}_{.k}) = \sum_{\substack{j,k \\ j\neq k}} \sqrt{\|\boldsymbol{\beta}_{jk}\|^2 + \|\boldsymbol{\beta}_{kj}\|^2}, \tag{4}$$

which allows us to impose a direct connection between parameter functions and E such that if $\|\beta_{jk}\|^2 = \|\beta_{kj}\|^2 = 0$ then there is not an edge between functional curves j and k. However, the problem in (3) is an infinite-dimensional optimization problem, so a way to solve it is to work with an finite-dimensional approximation.

2.3 Vector-on-Vector Multivariate Regression

In order to recover the edge set E is useful to work with the Karnhunen-Loève theorem, which says that:

$$f_{ij}(t) = \sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(t), \tag{5}$$

where $\xi_{ijl} = \langle f_{ij}; \phi_{jl}\rangle$, for $i = 1, \dots, n$, is an element of $\boldsymbol{\xi}_{jl}$, a r.v. that we assume has a Gaussian distribution with $E\left[\boldsymbol{\xi}_{jl}\right] = 0$ $V\left[\boldsymbol{\xi}_{jl}\right] = \gamma_{jl}$ and $E\left[\boldsymbol{\xi}_{jl}\boldsymbol{\xi}_{jr}\right] = \delta_{lr}$, where δ_{lr} is the Kronecker's delta such that $\delta_{lr} = 1$ if $l = r$. In particular we will work with an L truncated version of (5), so for the generic curve j we denote with $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \dots, \xi_{ijL})^T$ the vector of scores. As in [2], each vector score can be represented as:

$$\boldsymbol{\xi}_{ij} = \sum_{k\neq j}^{p} \boldsymbol{B}_{kj}\boldsymbol{\xi}_{ik} + \boldsymbol{r}_{ij}, \tag{6}$$

where $\boldsymbol{r}_{ij}$ is an error term and $\boldsymbol{B}_{kj}$ is a block of a matrix $\boldsymbol{B}$ of parameters related to the effect each functional curves $k \neq j$ to the response curve j for all the basis L. In particular, the generic term of $\boldsymbol{B}_{kj}$ is given by b_{kjl} scores of the expasion of $\beta_{kj}(t, t')$ in (2). The finite version of (3) will be

$$\hat{\boldsymbol{B}} = \underset{\boldsymbol{B}}{\arg\min} \sum_{i=1}^{n}\sum_{j=1}^{p} \|\boldsymbol{\xi}_{ij} - \boldsymbol{\xi}_{i\setminus j}\boldsymbol{B}_{kj}\|_F^2 + \lambda_n \sum_{\substack{j,k\\ j\neq k}} \sqrt{\|\boldsymbol{B}_{jk}\|_F^2 + \|\boldsymbol{B}_{kj}\|_F^2}, \tag{7}$$

where $\|\cdot\|_F^2$ is the Frobenius norm, which can be solved using the ADMM algorithm. Proposed by [7], the ADMM algorithm allows us to write the optimization problem as

$$\begin{aligned}\hat{\boldsymbol{B}}, \hat{\boldsymbol{Z}} = &\underset{\boldsymbol{B},\boldsymbol{Z}}{\arg\min} \sum_{i=1}^{n}\sum_{j=1}^{p} \|\boldsymbol{\xi}_{ij} - \boldsymbol{\xi}_{i\setminus j}\boldsymbol{B}_{kj}\|_F^2 + \lambda_n \sum_{\substack{j,k\\ j\neq k}} \sqrt{\|\boldsymbol{Z}_{jk}\|_F^2 + \|\boldsymbol{Z}_{kj}\|_F^2} \\ &\text{s.t. } \boldsymbol{B}_{jk} = \boldsymbol{Z}_{jk} \quad \forall\, j, k = 1, \dots, p,\end{aligned} \tag{8}$$

where $\boldsymbol{\xi}_{i\mathrm{j}} = (\xi_{i1}, \dots, \xi_{ip})\setminus\xi_{ij}$. It is possible to show that the solution of (8), at step $t + 1$, will be:

$$\begin{aligned}\hat{\boldsymbol{B}}_{kj}^{(t+1)} &= \left(\left(\boldsymbol{\xi}_{\setminus j}\right)^T \boldsymbol{\xi}_{\setminus j} + \rho\mathcal{I}_L\right)^{-1}\left(\left(\boldsymbol{\xi}_{\setminus j}\right)^T \boldsymbol{\xi}_j + \rho\left(\boldsymbol{Z}_{kj}^{(t)} + \boldsymbol{U}_{kj}^{(t)}\right)\right), \\ \hat{\boldsymbol{Z}}_{jk}^{(t+1)} &= \left(1 - \frac{\lambda}{\rho\|\boldsymbol{B}_{jk}^{(t+1)} + \boldsymbol{U}_{jk}^{(t)}\|_F^2}\right)_+ \left(\boldsymbol{B}_{jk}^{(t+1)} + \boldsymbol{U}_{jk}^{(t)}\right), \\ \hat{\boldsymbol{U}}_j^{(t+1)} &= \boldsymbol{U}_j^{(t)} + \boldsymbol{B}_j^{(t+1)} - \boldsymbol{Z}_j^{(t+1)}.\end{aligned} \tag{9}$$

2.4 Choice of the Basis and Scores Estimation

The choice of the basis $\{\phi_{jl}\}_{l=1}^{+\infty}$ used to estimate the vector of scores $\boldsymbol{\xi}_{jl}$ is made by using the PACE method proposed in [5] curve by curve. In particular, for the generic $j-$curve we assume that each curve is observed with a measurement error on a finite number of random temporal points N_{ij}, i.e., $Y_{ij}(T) = f_{ij}(T_{ij}) + \epsilon_{ij} = \mu_j(T_{ij}) + \sum_{l=1}^{L} \xi_{ijl}\phi_{jl}(T_{ij}) + \epsilon_{ij}, \quad \forall\, T_{ij} \in \mathcal{T}$, with $E\left[\epsilon_{ij}\right] = 0$ and $V\left[\epsilon_{ij}\right] = \sigma_j^2$, where ϵ_{ij} are independent of ξ_{ijl}. First, a local linear smoother is used to obtain a smooth mean function estimate $\hat{\mu}_j$ and a smooth covariance estimate $\hat{\mathcal{G}}_j(f_j(t), f_j(s))$ such that the eigenfunctions and eigenvalues is estimate by the standard eigenequations $\int_{\mathcal{T}} \hat{\mathcal{G}}_j(s,t)\hat{\phi}_{jl}(s)\, ds = \hat{\gamma}_{jl}\hat{\phi}_{jl}(s)$. Assuming that ξ_{ijl} and ϵ_{ij} are jointly Gaussian, and denoted with $\tilde{\boldsymbol{Y}}_{ij} = \left(Y_{ij1}, \dots Y_{ijN_i}\right)^T$, $\tilde{\boldsymbol{X}}_{ij} = \left(X_{ij}(T_{ij1}), \dots, X_{ij}(T_{ijN_i})\right)^T$ and $\boldsymbol{\mu}_{ij} = \left(\mu_j(T_{ij1}), \dots \mu_j(T_{ijN_i})\right)^T$, the best prediction of the scores is given by the conditional expectation

$$\hat{\xi}_{ijl} = E\left[\xi_{ijl} \mid \tilde{\boldsymbol{Y}}_{ij}\right] = \hat{\gamma}_{jl}\hat{\phi}_{jl}^T \hat{\Sigma}_{\mathbf{Y}_{ij}}^{-1} \left(\tilde{\boldsymbol{Y}}_{ij} - \hat{\boldsymbol{\mu}}_{ij}\right), \tag{10}$$

where $\hat{\Sigma}_{\boldsymbol{Y}_{ij}}$ is the estimate of $\Sigma_{\boldsymbol{Y}_{ij}} = cov\left(\tilde{\boldsymbol{Y}}_{ij}\tilde{\boldsymbol{Y}}_{ij}\right) + \sigma^2 \mathcal{I}_{N_{ij}}$, in which σ^2 is the variance error term, estimated as $\hat{\sigma^2} = \frac{2}{|\mathcal{T}|}\int_{\mathcal{T}_1}\{V\hat{(}t) - G\hat{(}t)\}dt$.

3 Simulation Study

The simulation study was carried out by setting the number of curves $p = 17$, the number of equispaced temporal points $d = 60$ in the domain $\mathcal{T} = [0,1]$ and taking into account five sample sizes: $n = 25, 50, 75, 100, 200$. For the generation of the true scores, a variance-covariance matrix of dimension $pL \times pL$ was generated with the following structure as in [4]

$$\Sigma = \left[\begin{array}{c|c|c|c} \Lambda_1 & \Sigma_{12} & \dots & \Sigma_{1p} \\ \hline \Sigma_{21} & \Lambda_2 & \dots & \Sigma_{2p} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \Sigma p1 & \dots & \dots & \Lambda_p \end{array}\right], \tag{11}$$

where Λ_j is a diagonal matrix of variances of $\left(\xi_{j1}, \dots, \xi_{jL}\right)^T$, given by $\gamma_{j1} \geq \gamma_{j2} \geq \dots \geq \gamma_{jL}$. To simulate Σ matrix we follow a similar strategy used in [2]: we generate a first Toepliz matrix $\boldsymbol{T}$ of dimension $L \times L$ such that $T_{jk} = 1$ for $j = k$ and $T_{jk} = 1/2 \mid j - k \mid$ if $j \neq k$. Then a tridiagonal matrix R of dimension $l \times L$ such that $R_{jj} = 1$, $R_{j+1,j} = R_{j,j+1} = 0.5$ and the other entries are set to 0. A first precision matrix $\tilde{\Theta}$ is obtained by setting $\tilde{\Theta}_{jj} = T$, $\tilde{\Theta}_{j,j+1} = \tilde{\Theta}_{j+1,j} = 0.4R$ and $\tilde{\Theta}_{j,j+2} = \tilde{\Theta}_{j+2,j} = 0.2\mathcal{I}_L$. Then a singular value decomposition of each block in the diagonal of $\tilde{\Theta}^{-1}$ was made such that $\Lambda_j^{-1} = U_j M_j V_j$ and a diagonal block matrix $U = diag\left(U_1, \dots, U_p\right)$ is build. The final Σ is obtained as $\Sigma = U\Theta^{-1}U^T$.

We set the number of basis $L^* = 5$, and we generate the true scores $\boldsymbol{\xi}^*$ from a multivariate Gaussian distribution with zero mean and variance-covariance matrix Σ. By using a set of L^* Fourier basis, denoted as ϕ_l^*, the observed curves are obtained as $Y_{ij}(T_{ij}) = \sum_{l=1}^{5} \boldsymbol{\xi}_{il}^{X_1*} \phi_l^* + \epsilon_{ij}^2$ where $\epsilon_{ij}^2 \sim \mathcal{N}(0, 1)$ is an error term. The performance of our model has been evaluated, on 100 simulations, via the *ROC* curve for 100 different values of λ_n: from λ_{max} such that all parameters will be equal to zero and therefore the edge set E is empty, to λ_{min} such that all parameters will be different from zero and therefore the edge set E is full. These performances have been compared with those of other competitors, in particular, the model proposed by [4] and those proposed by [2], using both **AND/OR** schemes and, as the authors propose, two different methods for the estimation of the scores: a first one by projecting the curves onto their own spaces (denoted by gX), and a second one by projecting the curves onto the response curve's space (denoted by gY). Table 1 shows the average

Table 1: Average AUC for the five sample size.

model	$n = 25$	$n = 50$	$n = 75$	$n = 100$	$n = 200$
proposal	0.8274	0.9224	0.9667	0.9783	0.9976
Zhao.and.gX	0.8222	0.9257	0.9684	0.9848	0.9985
Zhao.and.gY	0.8307	0.9241	0.9675	0.9850	0.9985
Zhao.or.gX	0.7884	0.8964	0.9453	0.9705	0.9961
Zhao.or.gY	0.7924	0.8928	0.9442	0.9705	0.9961
Qiao	0.7396	0.8288	0.8782	0.9091	0.9599

AUC values, from which it can be noted that our proposal has excellent performances for sample sizes of 50 and above and is similar to the model proposed by [2], in particular using an **AND** scheme, while it has slightly better performances than the same model but using an **OR** scheme, and superior to the model proposed by [4].

4 Conclusions

The methodology shown in this article allows to estimate a Gaussian Graphical model for functional data that, remaining within the framework of penalized multivariate functional regression models, represents an excellent alternative to the models already known in the literature: While the advantages over the model proposed by [4] are obvious in that our proposal has superior performance, the advantages over using the methodology proposed by [2] are computational. With our proposal, parameter estimation involves the estimation of the adjacency matrix E, whereas in the method proposed by [2], parameter estimation is followed by the estimation of the neighbourhood set for each functional curve, and then by the estimation of the adjacency matrix E by combining the neighbourhood sets based on an **AND/OR** scheme, the choice of which is subjective and may influence the estimation.

Acknowledgements The authors are grateful to the anonymous reviewers for their valuable comments.

Funding

L.A. gratefully acknowledges financial support from the University of Palermo (FFR2022, FFR2023 and FFR2024). L.A. was also financially supported by the European Union - Next Generation EU - Mission 4 Component 2 - CUP: B53D23009480006. G.C. acknowledges the support from project C16/20/002 of the Research Fund KU Leuven, Belgium.

References

1. J. O. Ramsay, B. W. Silverman. Functional Data Analysis. 1997. Springer Series in Statistics. Springer New York, NY, https://doi.org/10.1007/b98888
2. B. Zhao. Percy S. Zhai. Y. Samuel Wang. Mladen Kolar. "High-dimensional functional graphical model structure learning via neighborhood selection approach." Electron. J. Statist. 18 (1) 1042 - 1129, 2024. https://doi.org/10.1214/24-EJS2219
3. J. Zapata, S. Y. Oh, A. Petersen, Partial separability and functional graphical models for multivariate Gaussian processes, Biometrika, Volume 109, Issue 3, September 2022, Pages 665–681, https://doi.org/10.1093/biomet/asab046
4. X. Qiao, S. Guo, & G. M. James (2018). Functional Graphical Models. Journal of the American Statistical Association, 114(525), 211–222. https://doi.org/10.1080/01621459.2017.1390466
5. Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional Data Analysis for Sparse Longitudinal Data. Journal of the American Statistical Association, 100(470), 577–590. https://doi.org/10.1198/016214504000001745
6. J.M. Chiou, Y.F. Yang, Y.T. Chen, Multivariate functional linear regression and prediction, Journal of Multivariate Analysis, Volume 146, 2016, 301-312, ISSN 0047-259X, https://doi.org/10.1016/j.jmva.2015.10.003.
7. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Found. Trends Mach. Learn. 3, 1 (January 2011), 1–122. https://doi.org/10.1561/2200000016
8. T. Hsing, & R. Eubank (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators.