Latent Factor Gaussian Processes with Log-Euclidean Metric for Dynamic Covariance Modeling

Lingge Li [1] Dustin Pluta [1] Babak Shahbaba [1] Pierre Baldi [1]

Abstract

To capture the dynamic covariance of non-stationary multivariate time series, we propose a latent factor Gaussian process model using the Log-Euclidean metric for symmetric positive definite matrices. From multivariate time series, the empirical covariance process is estimated with the sliding window method and mapped to Euclidean space by matrix logarithm. Gaussian process latent factors are then fitted to determine a parsimonious representation of the non-stationary dynamics and quantify the uncertainty. The proposed model naturally incorporates the positive definite constraints of covariance matrices, has desirable Bayesian properties, and shows competitive performance on simulated data.

1. Introduction & Related Work

Studying dynamic covariance of multivariate time series is of great interest in neuroscience and economics but very challenging from a modeling perspective, especially when the time series is non-stationary. For a p-variate time series, the covariance at each time t will have $(p+1)p/2$ terms to estimate. Without further constraints on the covariance process, the number of terms to estimate is likely to be practically prohibitive. In most dynamic covariance models, there is a common framework with three key ingredients: estimation of covariance from observed time series, identification of non-stationary covariance dynamics, and dimensionality reduction.

In the neuroimaging community, many studies first apply a dimensionality reduction technique such as principal component analysis (PCA) on the time series and use a hidden Markov model (HMM) or k-means clustering to find a number of distinct covariance states (Cabral et al., 2017)(Stevner et al., 2019). While many variants offer different improvements, they are nevertheless non-probabilistic and can be problematic for uncertainty quantification. This is a serious drawback when analyzing data that typically have low signal-to-noise ratio such as fMRI measurements.

There are time series extensions of Bayesian factor models of this form $X(t) = f(t)\Lambda(t) + \varepsilon(t)$ such as latent factor stochastic volatility (LFSV) (Kastner et al., 2017)(Chib et al., 2006). The LFSV model expresses the error terms as $\varepsilon(t) = U_t(h_t^U)^{1/2}\epsilon(t)$, where $U_t(h_t^U) = \text{diag}\{\exp(h_{1t}^U, \ldots, \exp(h_{pt}^U)\}$, and the factors as $f(t) = V_t(h_t^V)^{1/2}\zeta(t)$, where $h_t = (h_t^U, h_t^V)$ are the latent volatilities. The log-volatilities are assumed to follow an autoregressive process of order one, $h_{jt}^U = \mu_j + \phi_j(h_{j,t-1} - \mu_j) + \sigma_j\eta_j(t)$. The innovations $\epsilon(t), \zeta(t)$ and $\eta(t)$ are all assumed to follow standard multivariate Gaussian distributions. Conditional on the volatilities, the covariance of the time series is $\text{Cov}(X(t)|h_t) = \Lambda V_t(h_t^V)\Lambda' + U_t(h_t^U)$ where factor loadings $\Lambda(t)$ are fixed at Λ.

An alternative latent factor model proposed in (Fox & Dunson, 2015) instead allows both factors $f(t)$ and factor loadings $\Lambda(t)$ to vary over time. The factors $f(t)$ are modeled with independent Gaussian processes ($\mathcal{GP}$) similar to (Yu et al., 2009). In this setting, the dynamics of the factor loadings must be modeled as a matrix process, with some structure imposed to ensure the model is tractable. The resulting conditional covariance is $\text{Cov}(X(t)|f_t) = \Lambda(t)\Lambda(t)' + \Sigma_\varepsilon$. While this model is considerably more flexible than the LFSV model, the factor matrix process adds substantial complexity.

Aiming to bridge the gap between aforementioned models, we propose the latent factor Gaussian process (LFGP) model with Log-Euclidean metric. Rather than on the observed time series, we place the factor structure on the covariance process, as consistently estimated by tapered sliding window. This has the advantage of decoupling the mean and covariance models. As covariance matrices lie on the manifold of symmetric positive-definite (SPD) matrices, the Log-Euclidean metric allows unconstrained modeling of the upper triangle of the covariance elements. The LFGP model is fully probabilistic and can be used for practical inference on characteristics of the covariance process.

[1]University of California, Irvine, USA. Correspondence to: Lingge Li <linggel@uci.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

Sliding-window (SW) methods have been extensively researched for the estimation of dynamic functional connectivity and shown to be consistent; see (Preti et al., 2017) for a recent detailed review of existing literature. Studies of the performance of sliding window estimates recommend the use of a tapered kernel to decrease the impact of outlying measurements and to improve the spectral properties of the estimate (Leonardi & Van De Ville, 2015).

The intrinsic space of $p \times p$ SPD matrices is a Riemannian manifold, which, after transformation by matrix logarithm, is isomorpic to $\mathbb{R}^q$ with the usual Euclidean norm, where $q = (p+1)p/2$. Methods for modeling covariances in regression contexts via matrix logarithm were first introduced in (Chiu et al., 1996). Further applications of the Log-Euclidean framework in neuroimaging have been developed in recent years (Zhu et al., 2009), but our application of the Log-Euclidean framework is novel for modeling dynamic covariance.

2. Latent Factor Gaussian Process Model

2.1. Formulation

We consider estimation of dynamic covariance from a sample of n independent time series with p variables and T time points. Denote the ith observed p-variate time series by $X_i(t)$, $i = 1, \cdots, n$. We assume that each $X_i(t)$ follows an independent distribution $\mathcal{D}$ with zero mean and stochastic covariance process $K_i(t)$. To model the covariance process, we first compute the Gaussian tapered sliding window covariance estimates for each $X_i(t)$, with fixed window size L and taper τ to obtain $\hat{K}_{\tau,i}$. We then apply the matrix logarithm to obtain the $q = p(p+1)/2$ length vector $Y_i(t)$ specified by $\hat{K}_{\tau,i} = \text{Log}(\vec{\mathbf{u}}(Y_i))$, where $\vec{\mathbf{u}}$ maps a matrix to its vectorized upper triangle. We refer to $Y_i(t)$ as the "log-covariance" at time t.

The resulting $Y_i(t)$ can be modeled as an unconstrained q-variate time series. The LFGP model represents $Y_i(t)$ as a linear combination of r latent factors $F_i(t)$ through an $r \times q$ loading matrix B and independent Gaussian errors ϵ_i. The loading matrix B is held constant across observations and time. Here $F_i(t)$ is modeled as a product of independent Gaussian processes. Placing priors on the loading matrix B, Gaussian noise variance σ^2, and Gaussian process hyperparameter θ, gives a fully probabilistic latent factor model on the covariance process:

$$X_i(t) \sim \mathcal{D}(0, K_i(t)) \text{ where } K_i(t) = \exp\left(\vec{\mathbf{u}}(Y_i(t))\right) \quad (1)$$

$$Y_i(t) = F_i(t) \cdot B + \epsilon_i \text{ where } \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, I\sigma^2) \quad (2)$$

$$F_i(t) \sim \mathcal{GP}(0, \kappa(t;\theta)) \quad (3)$$

$$B \sim p_1, \sigma^2 \sim p_2, \theta \sim p_3. \quad (4)$$

The LFGP model employs a latent distribution of curves $\mathcal{GP}(0, \kappa(t;\theta))$ to capture temporal dependence between observations, thus inducing a $\mathcal{GP}$ on log-covariance $Y(t)$. This conveniently allows multiple observations to be modeled as different realizations of the same induced $\mathcal{GP}$ as done in (Lan et al., 2017). The model posteriors are conditioned on different observations despite sharing the same kernel. For better identifiability, the $\mathcal{GP}$ variance scale is fixed so that the loading matrix can be unconstrained. The model can learn a wide range of temporal dynamics in the latent space due to the flexibility of $\mathcal{GP}$ and we will demonstrate this advantage on simulated data in Section 4.2.

2.2. Properties

Stationarity of LFGP log-covariance process The covariance of the log-covariance process $Y(t)$ depends only on the static loading matrix B and the factor covariance kernels. Consequently, when stationary kernels are chosen for the factor priors, the log-covariance process estimated by the LFGP model is stationary. Explicitly, for factor kernels $\kappa(s,t;\theta_k), k = 1, \ldots, r$, and assuming $\varepsilon_i(t) \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$, with $\Sigma = (\sigma^2_{jj'})_{j,j' \le q}$ constant across observations and time, the covariance of elements of $Y(t)$ is

$$\text{Cov}(Y_{ij}(s), Y_{ij'}(t)) = \text{Cov}(\sum_{k=1}^{r} F_{ik}(s)\beta_{kj} + \varepsilon_{ij}(t), \quad (5)$$

$$\sum_{k=1}^{r} F_{ik}(t)\beta_{kj'} + \varepsilon_{ij'}(t)) \quad (6)$$

$$= \sum_{k=1}^{r} \beta_{kj}\beta_{kj'}\kappa(s,t;\theta_k) + \sigma^2_{jj'} \quad (7)$$

which is weakly stationary when κ is weakly stationary.

Large support The prior distribution of the log-covariance process $Y(t)$ is a linear combination of r independent GPs each with mean 0 and kernel $\kappa(s,t;\theta_j), j = 1, \cdots, r$. That is, each log-covariance element will have prior $Y_j(t) = \sum_{k=1}^{r} \beta_{jk} F_k(t) \sim \mathcal{GP}(0, \sum \beta^2_{jk}\kappa(s,t;\theta_k))$. This provides an extremely flexible representation of covariance elements. Even in the relatively simple case of $\kappa(t;\theta_k)$ chosen as a squared exponential kernel, the linear combination of GPs allows the model to capture dynamics at multiple frequencies. Because the LFGP model can be interpreted as a regression on the sliding window estimates (which are consistent for $K(t)$), and under the assumption of weak stationarity for the covariance process $K(t)$, the large support of the LFGP model inherits from large support of the latent factor model for multivariate stationary time series.

Posterior contraction To consider posterior contraction of the proposed model, we can extend results given in (van der Vaart et al., 2008) on the posterior contraction rate of univariate GP regression. For this, we assume that

$Y(t) : [0, 1] \to Sym_p$ is a smooth function in $\ell^\infty([0, 1])$ with respect to the Euclidean norm, which implies that $K(t) : [0, 1] \to \mathbb{P}_p$ is a smooth $\ell^\infty([0, 1])$ function with respect to the Log-Euclidean norm.

Assume the true log-covariance process $w = \vec{\mathbf{u}}(\log(K(t))$ is in the support of the product GP $W \sim F(t)B$, for $F(t)$ and B defined above. Further, assume that the prior p_2 for σ^2 has support on a given interval $[a, b] \subset (0, \infty)$. Then, since $W \in \ell^q_\infty([0, 1])$, an extension of Theorem 3.3 in (van der Vaart et al., 2008) gives $E_0\Pi_n((w, \sigma) : \|w - w_0\|_n + |\sigma - \sigma_0| > M\varepsilon_n|Y_1, \cdots, Y_n) \to 0$ for sufficiently large M and $\varepsilon_n \to 0$ satisfying a specific lower bound condition. We will verify posterior contraction empirically in Section 4.1.

3. Scalable Computation

The LFGP model can be fitted with Gibbs sampling as commonly done for Bayesian latent variable models. In every iteration, we first sample latent $\mathcal{GP}$ factors $F|B, \sigma^2, \theta, Y$ from the conditional $p(F|Y)$ as F, Y are jointly multivariate Gaussian, where the covariance can be written in terms of B, σ^2, θ. However, it is worth noting that the joint distribution has a large covariance matrix, which could be computationally expensive to invert naïvely. Given latent factors F, the parameters B, σ^2 and θ become conditionally independent. Using conjugate priors for Bayesian linear regression, the posterior $p(B, \sigma^2|F, Y)$ is directly available. For the posterior of $\mathcal{GP}$ length scales, $p(\theta|F)$, either Metropolis random walk or slice sampling (Neal et al., 2003) can be used within each Gibbs step because the parameter space is low dimensional.

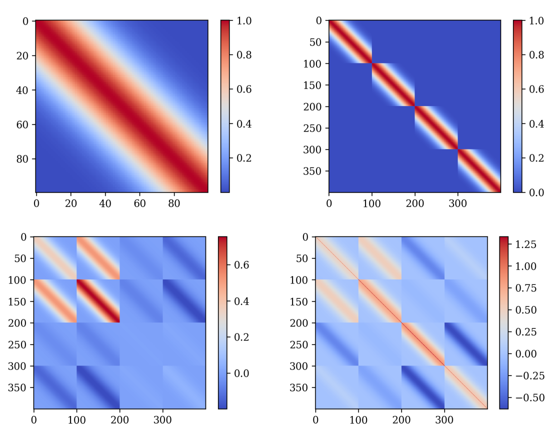

Figure 1. Toeplitz covariance matrix with squared exponential kernel, Kronecker covariance matrix of multiple observations, Kronecker covariance matrix of induced process by one factor, final covariance matrix with diagonal variance.

To make computation of the $\mathcal{GP}$ posteriors tractable, it is important to exploit the structure of the covariance matrix. For each independent latent $\mathcal{GP}$ factor F_j, there are n sets of observations at T time points. Therefore, the $\mathcal{GP}$ covariance matrix Σ_{F_j} has dimensions $nt \times nt$. We notice that the covariance Σ_{F_j} can be decomposed with Kronecker product $\Sigma_{F_j} = I \otimes K_T(t)$. K_T is the kernel over time t; it is a $t \times t$ matrix so the inversion cost is $\mathcal{O}(t^3)$ instead of $\mathcal{O}((nt)^3)$. With either the squared-exponential or Matern kernel, $K_T(t)$ has a Toeplitz structure and can be approximated with inducing points or interpolation, further reducing the computational cost (Wilson & Nickisch, 2015).

With all the latent $\mathcal{GP}$ factors F (dimensions $n \times t \times r$) and loading matrix B (dimensions $r \times q$), we have an induced $\mathcal{GP}$ on Y. The dimensionality of Y is $n \times t \times q$ so the full $(ntq) \times (ntq)$ covariance matrix is prohibitive to invert. As every column of Y is a weighted sum of the $\mathcal{GP}$ factors, the covariance matrix Σ_Y can be written as a sum of Kronecker products

$$\sum_{j=1}^{r} A_j \otimes \Sigma_{F_j} + I\sigma^2 \tag{8}$$

where Σ_{F_j} is the covariance matrix of the jth latent $\mathcal{GP}$ factor and A_j is a $q \times q$ matrix based on the factor loadings. Iteratively fitting latent factors only involves $A_j \otimes \Sigma_{F_j} + I$, which can be inverted in a computationally efficient way (Stegle et al., 2011).

4. Experiments

4.1. Empirical Posterior Contraction

To verify posterior contraction in the LFGP model, we simulate data with various sample sizes and numbers of observation time points. The covariance dimensions are 5×5 with two latent factors. On each simulated data set, we then fit a LFGP model and obtain posterior draws. For model bias, we consider the mean squared error of posterior median of the reconstructed log-covariance series. For posterior uncertainty, the posterior sample variance is used. As shown in Table 1, both sample size n and number of observation time points t contribute to posterior contraction; there is less bias and uncertainty as n and t increase.

Table 1. Simulation results with various sample sizes and numbers of time points

	$t = 25$	$t = 50$	$t = 100$
$n = 1$	12.21 (20.22)	6.911 (7.588)	3.728 (5.218)
$n = 10$	7.845 (8.743)	4.123 (5.836)	1.682 (2.582)
$n = 20$	7.089 (7.714)	3.273 (3.989)	1.672 (2.659)
$n = 50$	5.869 (7.358)	3.237 (3.709)	1.672 (1.907)

MSE of posterior median (sample variance) $\times 10^{-2}$

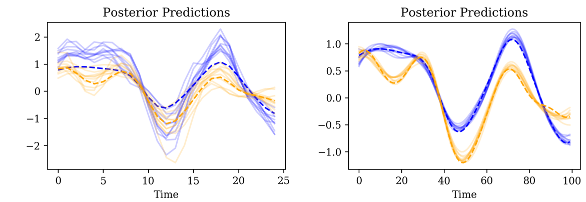

Figure 2. Diffused and biased posterior predictions of LFGP model ($n = 1, t = 25$) with truth shown in dashed lines (left); concentrated and unbiased posterior predictions ($n = 50, t = 100$) (right).

4.2. Model Comparisons on Simulated Data

Here we consider three benchmark models: sliding window with principal component analysis (SW-PCA), hidden markov model (HMM), and latent factor stochastic volatility model (LFSV). SW-PCA and HMM are commonly used in dynamic functional brain connectivity studies but have limitations. For instance, the estimated covariance matrices are sensitive to the choice of sliding window size and PCA does not take the estimation error into account. While HMM is a probabilistic model and can be used in conjunction with a time series model, it is not able to capture smoothly varying covariance dynamics. In contrast to the proposed model, none of these benchmark models can handle multiple observations.

To compare the performance of different models, we simulate time series data $X_t \sim N(0, K(t))$ with time varying covariance $K(t)$. The covariance $K(t)$ has a low-rank structure and follows deterministic dynamics, that are given by $\vec{\mathbf{u}}(\log(K(t))) = T(t) \cdot A$. We consider three different scenarios of latent dynamics $T(t)$: square waves, piece-wise linear functions, and smooth splines. For each scenario, we randomly generate 100 time series data sets and fit all the models. Each time series has 10 variables with 1000 observations and the latent dynamics are 4-dimensional. The evaluation metric is reconstruction loss of the covariance as measured by the Log-Euclidean metric. Table 2 displays the simulation results and we can see that the proposed LFGP model outperforms benchmark models.

Table 2. Model comparison results on simulated data

	Square	Linear	Spline
SW-PCA	0.693 (0.499)	0.034 (0.093)	0.037 (0.016)
HMM	1.003 (1.299)	0.130 (0.124)	0.137 (0.113)
LFSV	4.458 (2.416)	0.660 (0.890)	0.532 (0.400)
LFGP	0.380 (0.420)	0.027 (0.088)	0.028 (0.123)

Median log-covariance MSE (standard deviation)

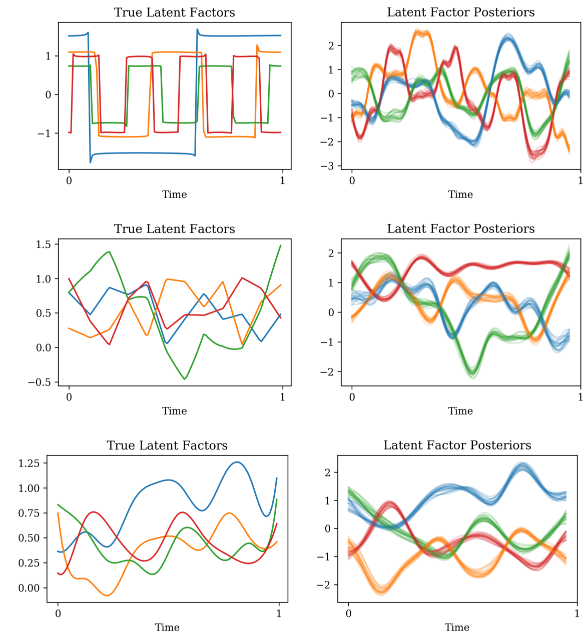

Figure 3. With the jagged dynamics of discrete states, the LFGP model fails to capture the "jumps" but approximates the overall trend (top). When the underlying dynamics are smooth, the LFGP model can accurately recover the shape up to some scaling constant (bottom).

5. Discussion & Future Work

We have presented a novel model for dynamic covariance of non-stationary multivariate time series. Based on sliding window estimates for covariances, the model utilizes latent Gaussian process factors on the SPD matrix manifold with Log-Euclidean metric. Determining the number of latent factors in the model is crucial in practice. As seen in Bayesian factor analysis literature, sparsity inducing priors could be used to achieve automatic factor selection. A more important practical consideration is model interpretability. To this end, the factor loadings can be used to identify clusters and potentially reveal common structures in the covariance process.

Currently, the model limitations are: (1) the covariance estimation and model fitting are done separately; (2) the model input is $(p+1)p/2$-dimensional and computation is challenging when p is large. A possible solution to address both limitations is joint covariance estimation and dimensionality before the log-Euclidean transformation. Given the strong preliminary model performance on simulated data, we think the possibility is well worth exploring.

References

Cabral, J., Vidaurre, D., Marques, P., Magalhães, R., Moreira, P. S., Soares, J. M., Deco, G., Sousa, N., and Kringelbach, M. L. Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific reports*, 7(1): 5135, 2017.

Chib, S., Nardari, F., and Shephard, N. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371, 2006.

Chiu, T. Y., Leonard, T., and Tsui, K.-W. The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, 91(433):198–210, 1996.

Fox, E. B. and Dunson, D. B. Bayesian nonparametric covariance regression. *The Journal of Machine Learning Research*, 16(1):2501–2542, 2015.

Kastner, G., Frühwirth-Schnatter, S., and Lopes, H. F. Efficient bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics*, 26(4):905–917, 2017.

Lan, S., Holbrook, A., Fortin, N. J., Ombao, H., and Shahbaba, B. Flexible bayesian dynamic modeling of covariance and correlation matrices. 2017.

Leonardi, N. and Van De Ville, D. On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage*, 104:430–436, 2015.

Neal, R. M. et al. Slice sampling. *The annals of statistics*, 31(3):705–767, 2003.

Preti, M. G., Bolton, T. A., and Van De Ville, D. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160:41–54, 2017.

Stegle, O., Lippert, C., Mooij, J. M., Lawrence, N. D., and Borgwardt, K. Efficient inference in matrix-variate gaussian models with\iid observation noise. In *Advances in neural information processing systems*, pp. 630–638, 2011.

Stevner, A., Vidaurre, D., Cabral, J., Rapuano, K., Nielsen, S. F. V., Tagliazucchi, E., Laufs, H., Vuust, P., Deco, G., Woolrich, M., et al. Discovery of key whole-brain transitions and dynamics during human wakefulness and non-rem sleep. *Nature communications*, 10(1):1035, 2019.

van der Vaart, A. W., van Zanten, J. H., et al. Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, 36(3):1435–1463, 2008.

Wilson, A. and Nickisch, H. Kernel interpolation for scalable structured gaussian processes (kiss-gp). In *International Conference on Machine Learning*, pp. 1775–1784, 2015.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L. (eds.), *Advances in Neural Information Processing Systems 21*, pp. 1881–1888. Curran Associates, Inc., 2009.

Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C., and Lin, W. Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *Journal of the American Statistical Association*, 104(487): 1203–1212, 2009.