

Functional Data Analysis in Unit Interval

Moo K. Chung mkchung@wisc.edu

January 19, 2020

1 Hilbert space theory in $[0, 1]$

Consider the i -th functional time series data

$$\zeta_i(t) = \mu_i(t) + \epsilon_i(t), \quad (1)$$

where t is time variable. We assume the functional data is scaled in such a way that they are defined in $[0, 1]$. Formulating the Fourier analysis in a unit interval makes the numerical implementation more convenient. ϵ_i is a zero mean noise at each fixed t , i.e.,

$$\mathbb{E}\epsilon_i(t) = 0.$$

μ_i is an unknown smooth functional signal to be estimated. It is reasonable to assume that

$$\zeta_i, \mu_i \in \mathcal{L}^2[0, 1],$$

the space of square integrable functions. Note any function $f \in \mathcal{L}^2[0, 1]$ satisfies the condition

$$\int_0^1 f^2(t) dt < \infty.$$

This condition is needed to guarantee the convergence in the Fourier series expansion. Instead of estimating μ_i in $\mathcal{L}^2[0, 1]$, we approximate μ_i as a linear combination of some known basis functions $\psi_0, \psi_1, \dots, \psi_k$ by projecting data to the finite subspace spanned by $\psi_0, \psi_1, \dots, \psi_k$.

Assume ψ_l is orthonormal basis in $[0, 1]$ with respect to the inner product

$$\langle f, g \rangle = \int_0^1 f(t)g(t) dt.$$

Thus, it is required to have

$$\int_0^1 \psi_l(t)\psi_m(t) dt = \delta_{lm},$$

the Kronecker delta with 1 if $l = m$ and 0 if $l \neq m$. With respect to the inner product, the norm $\|\cdot\|$ is then defined as

$$\|f\| = \langle f, f \rangle^{1/2} = \left[\int_0^1 f^2(t) dt \right]^{1/2}.$$

We estimate μ_i in the subspace \mathcal{H}_k spanned by up to the k orthonormal basis functions:

$$\mathcal{H}_k = \left\{ \sum_{l=0}^k c_l \psi_l(t) : c_l \in \mathbb{R} \right\} \subset \mathcal{L}^2[0, 1].$$

The least squares estimation (LSE) of μ_i in \mathcal{H}_k is given by

$$\hat{\mu}_i = \arg \min_{f \in \mathcal{H}_k} \|f - \zeta_i(t)\|^2. \quad (2)$$

Theorem 1 *The minimization of (2) is given by*

$$\hat{\mu}_i = \sum_{l=0}^k \langle \zeta_i, \psi_l \rangle \psi_l, \quad (3)$$

where the l -th degree Fourier coefficient $\langle \zeta_i, \psi_l \rangle$ is given by the inner product

$$\langle \zeta_i, \psi_l \rangle = \int_0^1 \zeta_i(t) \psi_l(t) dt. \quad (4)$$

Proof. We need to find function

$$f(t) = \sum_{l=0}^k c_l \psi_l(t)$$

that is the closest to ζ_i . The distance between f and ζ_i is given by

$$I(c_0, c_1, \dots, c_k) = \int_0^1 \left| \sum_{l=0}^k c_l \psi_l(t) - \zeta_i(t) \right|^2 dt,$$

which is a $k+1$ dimensional function in unknown parameter space $(c_0, c_1, \dots, c_k) \in \mathbb{R}^{k+1}$. Since I is a quadratic function in (c_0, c_1, \dots, c_k) , it has the global minimum at

$$\frac{\partial I}{\partial c_0} = \frac{\partial I}{\partial c_1} = \dots = \frac{\partial I}{\partial c_k} = 0.$$

The algebraic derivation is left as an exercise but we have

$$c_l = \langle \zeta_i, \psi_l \rangle$$

for all $l = 0, 1, \dots, k$. \square

The expansion (3) is called the *Fourier series*. As $k \rightarrow \infty$, the expansion converges to ζ_i , i.e.,

$$\zeta_i(t) = \sum_{l=0}^{\infty} \langle \zeta_i, \psi_l \rangle \psi_l.$$

2 Cosine series representation

The cosine series representation (CSR) is first introduced in [Chung et al. \(2010\)](#). Many incorrectly mistake it as the discrete cosine transform. CSR is based on the least squares estimation technique while the discrete cosine transform is not.

There are infinitely many possible orthonormal basis in interval $[0, 1]$. Here we explain a spectral approach for obtaining orthonormal basis in $[0, 1]$. Consider the space of square integrable functions in $[0, 1]$ denoted by $\mathcal{L}^2[0, 1]$. Let us solve the eigenfunction

$$\Delta\psi + \lambda\psi = 0 \quad (5)$$

in $\mathcal{L}^2[0, 1]$ with 1D Laplacian $\Delta = \frac{d^2}{dt^2}$. We are solving the 2nd order differential equation

$$\frac{d^2\psi(t)}{dt^2} + \lambda\psi(t) = 0. \quad (6)$$

This part is left as an [exercise](#), but we can show that eigenfunctions $\psi_0, \psi_1, \psi_2, \dots$ form an orthonormal basis in $\mathcal{L}^2[0, 1]$. Note that if ψ_l is an eigenfunction, any multiple of ψ_l is also an eigenfunction. Thus, it is expected the eigenfunctions are properly normalized such that

$$\int_0^1 \psi_l^2(t) dt = 1.$$

The eigenfunctions satisfying (5) is then given by the usual Fourier sine and cosine basis

$$\psi_0(t) = 1, \psi_l = \sqrt{2} \sin(l\pi t), \sqrt{2} \cos(l\pi t) \quad (7)$$

with the corresponding eigenvalues $\lambda_l = l^2\pi^2$. This is left as an [exercise](#). We can check (7) are solutions by differentiating the eigenfunctions twice.

There are two eigenfunctions corresponding to the same eigenvalue. The multiplicity of eigenfunctions only happen if there is a symmetry in the domain of the Laplace eigenvalue problem (5). Proving this statement is left as an [exercise](#). The constant $\sqrt{2}$ is introduced to make the eigenfunctions orthonormal in $[0, 1]$.

Using both sine and cosine basis is not computationally efficient. Instead of solving (5) in the domain $[0, 1]$, if we add an additional constraint to eigenfunctions, we can get rid of the sine basis. Consider solving the problem in the larger unbounded domain \mathbb{R} with the periodic constraint

$$\psi(t+2) = \psi(t). \quad (8)$$

Suppose ψ is the solution in $[0, 1]$. The period 2 constraint forces ψ to be only valid in the intervals $\dots, [-2, -1], [0, 1], [2, 3], \dots$. There are gaps in $\dots, (-1, 0), (1, 2), (3, 4), \dots$, where there is no solution. We can fill the gap by padding with some arbitrary function. However, if we pad the gaps with any function, it may result in the Gibbs phenomenon (ringing artifacts) at the boundary of the intervals $\dots, 2, 1, 0, 1, 2, \dots$

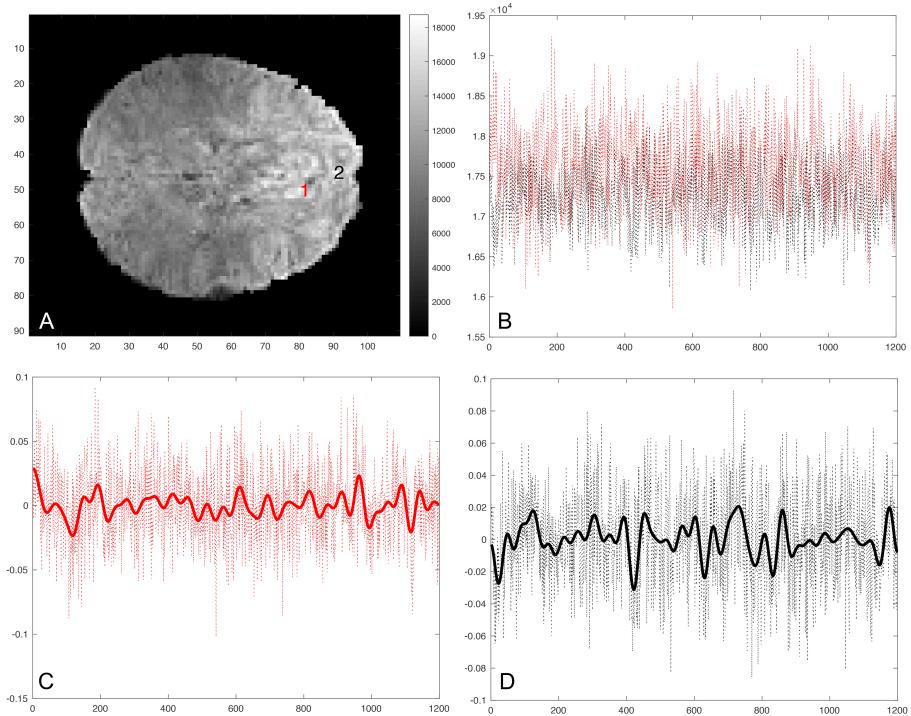


Figure 1: A. fMRI time series at two different voxels at time point t_1 . B. fMRI time series at voxel 1 (red) and 2 (black) shown for all 1200 time points. C. Normalized and scaled time series at voxel 1 and its cosine series representation with degree $k = 59$. D. Normalized and scaled time series at voxel 2 and its cosine series representation with degree $k = 59$. Such high-frequency denoising is often necessary for functional signals like fMRI and EEG.

(Chung et al., 2007). To avoid the Gibbs phenomenon, we force the function to be continuous at the boundary by putting the additional constraint of evenness, i.e.,

$$\psi(t) = \psi(-t). \quad (9)$$

If $\psi(t)$ is the eigenfunction well defined in $\dots, [-2, -1], (-1, 0), [0, 1], (1, 2), [2, 3], \dots$ we must have

$$\dots, \psi(t-2), \psi(-t), \psi(t), \psi(-t+2), \psi(t+2), \dots$$

The only eigenfunctions satisfying the two constraints (8) and (9) are the cosine basis

$$\psi_0(t) = 1, \psi_l(t) = \sqrt{2} \cos(l\pi t)$$

with the corresponding eigenvalues $\lambda_l = l^2\pi^2$ for integers $l > 0$. Then using the cosine basis only, any $f \in L^2[0, 1]$ can be represented as

$$f(t) = \sum_{l=0}^k c_l \psi_l(t) + \epsilon(t),$$

where c_l is the Fourier coefficients and ϵ is the residual error for using only k -th degree expansion.

If we put the constraint of oddness, i.e., $\psi(t) = -\psi(-t)$, we have sine basis

$$\psi_l(t) = \sqrt{2} \sin(l\pi t).$$

This is left as an exercise.

3 Parameter estimation

In practice, functional time series are observed at discrete time points t_1, t_2, \dots, t_n :

$$\zeta_i(t_j) = \mu_i(t_j) + \epsilon_i(t_j), \quad j = 1, \dots, n. \quad (10)$$

The underlying mean functions $\mu_i(t)$ are estimated as

$$\hat{\mu}_i(t) = \sum_{l=0}^k c_{li} \psi_l(t),$$

where the Fourier coefficients $(c_{0i}, c_{1i}, \dots, c_{ki})$ for the i -th time series is estimated using LSE as follows. At each t_j , we have

$$\underbrace{\begin{pmatrix} \zeta_i(t_1) \\ \zeta_i(t_2) \\ \vdots \\ \zeta_i(t_n) \end{pmatrix}}_{Y_i} = \Psi_{n \times (k+1)} \underbrace{\begin{pmatrix} c_{i0} \\ c_{i1} \\ \vdots \\ c_{ik} \end{pmatrix}}_{C_i},$$

where the design matrix of basis functions Ψ is given by

$$\Psi_{n \times (k+1)} = \begin{pmatrix} \psi_0(t_1) & \psi_1(t_1) & \cdots & \psi_k(t_1) \\ \psi_0(t_2) & \psi_1(t_2) & \cdots & \psi_k(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_0(t_n) & \psi_1(t_n) & \cdots & \psi_k(t_n) \end{pmatrix}.$$

The unknown coefficient vector C_i is estimated by multiplying Ψ^\top on the both sides:

$$\Psi^\top Y_i = \Psi^\top \Psi C_i.$$

If $n \gg k$, $\Psi^\top \Psi$ is full rank and invertible. Thus, we have

$$C_i = (\Psi^\top \Psi)^{-1} \Psi^\top Y_i. \quad (11)$$

Although we will not show, it is further possible to discretize and reshape the basis ψ_l in such a way that

$$\Psi^\top \Psi = I$$

the identity matrix. This is left as an [exercise](#). The detail is given in ([Chung, 2012](#)). This will make the numerical implementation of the Fourier series expansion computationally much more efficient for large k . The proposed least squares estimation technique avoids using the often used implicit Fourier transform (FT) ([Batchelor et al., 2006](#); [Bulow, 2004](#); [Gu et al., 2004](#)).

If we have p number of time series, it requires solving equation (11) p number of times. For really large LSE problems, this is computationally very inefficient. A more efficient way is to estimate the all the coefficients using a single LSE and invert $\Psi^\top \Psi$ only once by solving

$$Y_{n \times p} = \Psi_{n \times k} C_{k \times p},$$

where

$$\begin{aligned} Y_{n \times p} &= [Y_1, Y_2, \dots, Y_p], \\ C_{k \times p} &= [C_1, C_2, \dots, C_p]. \end{aligned}$$

Subsequently, all the coefficients are simultaneously estimated in the least squares fashion as

$$\widehat{C} = (\Psi^\top \Psi)^{-1} \Psi^\top Y.$$

4 Additional notes

It is also possible to have a slightly different but equivalent model that is easier to use in statistical inference. Assuming Gaussianness of data, $\epsilon_i(t)$ in (1) is a Gaussian stochastic process, which is simply a collection of random variables. Then $\epsilon_i(t)$ can be expanded using the given basis ψ_l as follows.

$$\epsilon_i(t) = \sum_{l=0}^k Z_l \psi_l(t) + e_i(t),$$

where $Z_l \sim N(0, \tau_l^2)$ are possibly *correlated* Gaussian random variables and e_i is the residual error that can be neglected in practice if enough number of basis are used. This is the consequence of the Karhunen-Loeve expansion (Adler, 1990; Dougherty, 1999; Kwapien and Woyczyński, 1992; Yaglom, 1987).

Note. Karhunen-Loeve expansion states that $\epsilon_i(t)$ can be decomposed as

$$\epsilon_i(t) = \sum_{l=0}^m Z_l \phi_l(t)$$

for uncorrelated Gaussian random variables Z_l and some orthonormal basis $\phi_l(t)$. The algebraic determination of Z_l and $\phi_l(t)$ are left as an exercise. Since ϕ_l and ψ_l are different basis, ϕ_l can be represented as a linear combination of ψ_l . Rewriting ϕ_l in terms of ψ_l will make the Gaussian random variables Z_l correlated.

At the end, we can write model (1) as

$$\zeta_i(t) = \sum_{l=0}^k X_l \psi_l(t) + e_i(t),$$

where X_l are correlated Gaussian random variables.

References

- R.J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS, Hayward, CA, 1990.
- P.G. Batchelor, F. Calamante, J.D. Tournier, D. Atkinson, D.L. Hill, and A. Connelly. Quantification of the shape of fiber tracts. *Magnetic Resonance in Medicine*, 55: 894–903, 2006.
- T. Bulow. Spherical diffusion for 3D surface smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1650–1654, 2004.
- M.K. Chung. *Computational Neuroanatomy: The Methods*. World Scientific, Singapore, 2012.
- M.K. Chung, K.M. Dalton, L. Shen, A.C. Evans, and R.J. Davidson. Weighted Fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26:566–581, 2007.
- M.K. Chung, N. Adluru, J.E. Lee, M. Lazar, J.E. Lainhart, and A.L. Alexander. Cosine series representation of 3d curves and its application to white matter fiber bundles in diffusion tensor imaging. *Statistics and Its Interface*, 3:69–80, 2010.
- E.R. Dougherty. *Random Processes for Image and Signal Processing*. IEEE Press, 1999.

- X. Gu, Y.L. Wang, T.F. Chan, T.M. Thompson, and S.T. Yau. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging*, 23:1–10, 2004.
- S. Kwapien and W.A. Woyczyński. *Random Series and Stochastic Integrals: Single and Multiple*. Birkhauser, 1992.
- A.M. Yaglom. *Correlation Theory of Stationary and Related Random Functions Vol. I: Basic Results*. Springer-Verlag, 1987.