

Functional Data Analysis

Moo K. Chung
University of Wisconsin-Madison
`mkchung@wisc.edu`

February 13, 2026

Abstract

We present a functional data analysis (FDA) framework based on explicit orthonormal basis expansion for modeling and denoising complex biomedical signals. Observed functional data are represented as smooth functions in a Hilbert space, and statistical inference is performed directly on their basis coefficients. This formulation provides a transparent and flexible approach to smoothing, regularization, and hypothesis testing. Applications to diffusion tensor imaging tract modeling and EEG denoising demonstrate the advantages of explicit basis representations for scalable and interpretable functional modeling.

1 Introduction

Functional data analysis (FDA) studies data that are represented as functions rather than finite-dimensional vectors. Instead of treating observations as multivariate measurements, FDA models each sample as a smooth function defined on a continuum ([Ramsay and Silverman, 1997](#); [Ramsay, 2000](#)). This Hilbert space perspective provides a principled foundation for smoothing, regression, principal component analysis, and differential equation modeling. A central idea in FDA is basis expansion: a function $f(t)$ on $[0, 1]$ is represented as a linear combination of known basis functions, and statistical inference is performed on the coefficient vectors rather than directly on discretized measurements ([Chung et al., 2010](#); [Wang et al., 2018](#)). This separates representation from estimation and enables flexible regularization and shrinkage.

In this paper, we emphasize explicit orthonormal basis constructions, in particular the cosine basis obtained from Laplacian eigenfunctions. Unlike implicit transform-based procedures commonly used in signal processing, such as the discrete cosine transform, our formulation is grounded in least squares projection in

$\mathcal{L}^2[0, 1]$. This Hilbert space perspective allows direct statistical modeling at the coefficient level, naturally incorporating shrinkage methods such as the Wiener filter and stochastic representations through the Karhunen–Loëve expansion.

The methodology builds on earlier applications of cosine basis representations to three-dimensional white matter tract modeling in diffusion tensor imaging (Chung et al., 2010) and to EEG signal denoising (Wang et al., 2018). The resulting framework integrates projection theory, basis-based least squares estimation, optimal shrinkage, and stochastic process modeling into a unified and computationally efficient FDA pipeline. MATLAB codes and sample data are available at <http://brainimaging.waisman.wisc.edu/~chung/tracts>.

2 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 straightforward and 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.,

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

3 Least squares estimation

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

$$\zeta_i(t_j) = \mu_i(t_j) + \epsilon_i(t_j), \quad j = 1, \dots, n. \quad (5)$$

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

$$C_i = (\Psi^\top \Psi)^{-1} \Psi^\top Y_i. \quad (6)$$

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. 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 (6) p number of times. For really large LSE problems, this is computationally very inefficient. A more efficient way is to estimate 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

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

4 Wiener filter

In FDA, observed signals are often contaminated by observed measurements. After representing the signal in a suitable orthonormal basis, denoising reduces to estimating the basis coefficients from noisy observations. A natural criterion for estimation is the minimization of mean squared error (MSE). The Wiener filter (Wiener, 1949) provides the optimal linear estimator under this criterion by shrinking each coefficient according to its signal-to-noise ratio.

Suppose the observed signal satisfies

$$y(t) = f(t) + \varepsilon(t),$$

where $\varepsilon(t)$ is additive noise. In the functional data analysis framework, we represent the signal using an orthonormal basis $\{\psi_k(t)\}$ on $[0, 1]$:

$$f(t) = \sum_{k=0}^{\infty} \beta_k \psi_k(t).$$

Expanding all components in the same basis gives

$$y(t) = \sum_{k=0}^{\infty} c_k \psi_k(t), \quad \varepsilon(t) = \sum_{k=0}^{\infty} \eta_k \psi_k(t).$$

Matching coefficients at each frequency yields

$$c_k = \beta_k + \eta_k.$$

Assume further

$$\begin{aligned} E[\beta_k] &= 0, & E[\eta_k] &= 0, \\ \text{Var}(\beta_k) &= \sigma_f^2(k), & \text{Var}(\eta_k) &= \sigma_\varepsilon^2(k), \end{aligned}$$

and that β_k and η_k are uncorrelated.

We estimate the signal f by estimating its coefficients β_k through linear shrinkage of the observed coefficients:

$$\hat{\beta}_k = a_k c_k,$$

where the shrinkage factor a_k is chosen to minimize the mean squared error. The mean squared error at frequency k is

$$R(a_k) = \mathbb{E}[(\beta_k - a_k c_k)^2].$$

Substituting $c_k = \beta_k + \eta_k$ yields

$$R(a_k) = \mathbb{E}[((1 - a_k)\beta_k - a_k \eta_k)^2].$$

Because signal and noise are uncorrelated,

$$R(a_k) = (1 - a_k)^2 \sigma_f^2(k) + a_k^2 \sigma_\varepsilon^2(k).$$

Minimizing $R(a_k)$ with respect to a_k gives

$$a_k = \frac{\sigma_f^2(k)}{\sigma_f^2(k) + \sigma_\varepsilon^2(k)}.$$

Therefore, the Wiener estimator is

$$\hat{\beta}_k = \frac{\sigma_f^2(k)}{\sigma_f^2(k) + \sigma_\varepsilon^2(k)} c_k,$$

and the reconstructed signal becomes

$$\hat{f}(t) = \sum_{k=0}^{\infty} \frac{\sigma_f^2(k)}{\sigma_f^2(k) + \sigma_\varepsilon^2(k)} c_k \psi_k(t).$$

Thus, in the chosen orthonormal basis, the Wiener filter performs coefficient-wise shrinkage, scaling each coefficient by the ratio of signal variance to total variance. This estimator minimizes the mean squared error among all linear estimators of the form $\hat{\beta}_k = a_k c_k$.

5 Karhunen–Loève Expansion

It is often convenient to adopt a stochastic representation that is more suitable for statistical inference. Assume that the observational noise $\varepsilon_i(t)$ in (1) is a zero-mean Gaussian stochastic process on $[0, 1]$ with covariance function

$$C(s, t) = \text{Cov}\{\varepsilon_i(s), \varepsilon_i(t)\}.$$

A stochastic process is simply a collection of random variables indexed by t , and its second-order structure is fully characterized by $C(s, t)$.

By the Karhunen–Loève (KL) expansion (Adler, 1990; Dougherty, 1999; Kwapien and Woyczyński, 1992; Yaglom, 1987), any square-integrable zero-mean Gaussian process admits the representation

$$\varepsilon_i(t) = \sum_{l=0}^{\infty} Z_{il} \phi_l(t),$$

where $\{\phi_l\}$ are the orthonormal eigenfunctions of the covariance operator defined by

$$\int_0^1 C(s, t) \phi_l(s) ds = \tau_l^2 \phi_l(t),$$

and the coefficients satisfy

$$Z_{il} \sim N(0, \tau_l^2), \quad \text{Cov}(Z_{il}, Z_{im}) = 0 \text{ for } l \neq m.$$

Thus, the KL expansion diagonalizes the covariance operator, and the random coefficients are uncorrelated and independent under Gaussian assumptions. In practice, the infinite expansion is truncated to

$$\varepsilon_i(t) = \sum_{l=0}^K Z_{il} \phi_l(t) + e_i(t),$$

where $e_i(t)$ is the truncation error, which becomes negligible when K is sufficiently large.

If the chosen basis $\{\psi_k\}$ coincides with or approximates the eigenbasis $\{\phi_l\}$, the stochastic representation becomes particularly simple and facilitates inference on the coefficient vectors. The KL expansion therefore provides a probabilistic foundation for basis representations in functional data analysis.

6 Cosine basis

The cosine series representation (CSR) was introduced in Chung et al. (2010). Unlike the discrete cosine transform used in signal processing, CSR is formulated through least squares estimation within a functional data analysis framework.

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

$$\Delta\psi + \lambda\psi = 0 \quad (7)$$

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

$$\frac{d^2\psi(t)}{dt^2} + \lambda\psi(t) = 0. \quad (8)$$

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 (7) is then given by the usual Fourier sine and cosine basis

$$\psi_0(t) = 1, \psi_l = \sqrt{2} \sin(l\pi t), \sqrt{2} \cos(l\pi t) \quad (9)$$

with the corresponding eigenvalues $\lambda_l = l^2\pi^2$. We can check (9) 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 (7). 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 (7) 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

$$\psi(t+2) = \psi(t). \quad (10)$$

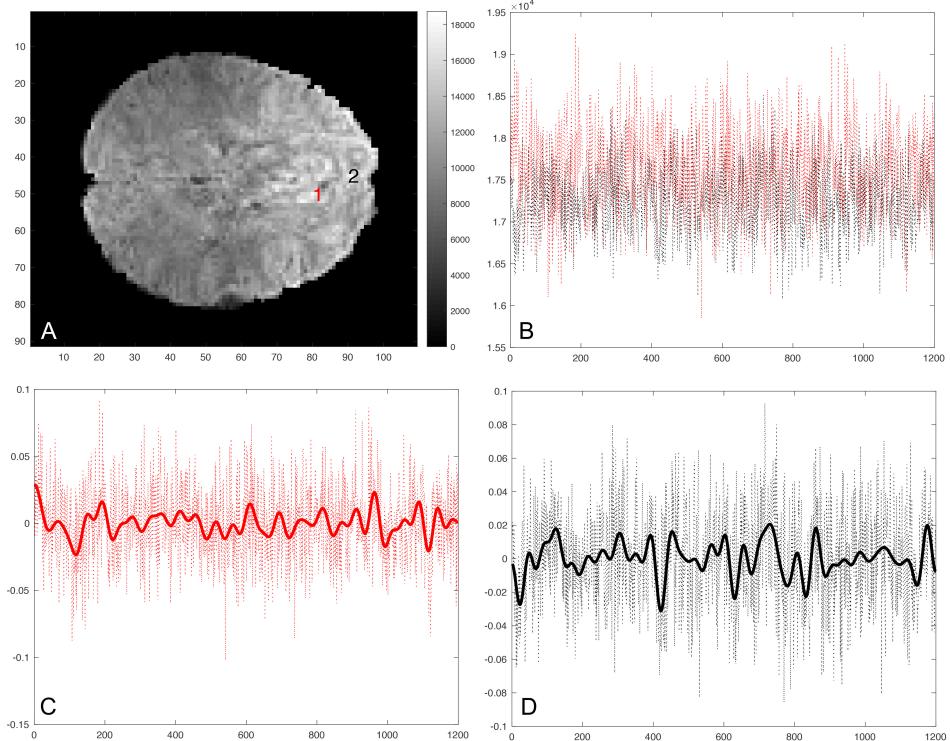


Figure 1: A. fMRI time series at two different voxels at time point t_1 . B. fMRI time series at voxel 1 (red) and 2 (black) shown for all 1200 time points. C. Normalized and scaled time series at voxel 1 and its cosine series representation with degree $k = 59$. D. Normalized and scaled time series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

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 $\dots, 2, 1, 0, 1, 2, \dots$ (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.,

$$\psi(t) = \psi(-t). \quad (11)$$

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 (10) and (11) 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 $f \in \mathcal{L}^2[0, 1]$ 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

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

7 Acknowledgment

I would like to thank Professor James Ramsay of McGill University for introducing me to this wonderful field of functional data analysis when I was a student there. This study was supported by NIH grants R01 EB022856 and R01 EB028753, and NSF grant MDS-2010778.

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.
- J.O Ramsay. Differential equation models for satistical functions. *The Canadian Journal of Statistics*, 28:225–240, 2000.
- J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer-Verlag, 1997.
- Y. Wang, H. Ombao, and M.K. Chung. Topological data analysis of single-trial electroencephalographic signals. *Annals of Applied Statistics*, 12:1506–1534, 2018.

N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949.

A.M. Yaglom. *Correlation Theory of Stationary and Related Random Functions Vol. I: Basic Results*. Springer-Verlag, 1987.