Geometry-driven functional boxplots

Karthik Bharath

University of Nottingham

Partially supported by NSF DMS 1613054

With: W. Xie, S. Kurtek (OSU) and Y. Sun (KAUST).

- **Paper**:

  Geometric Approach to Visualization of Variability in Functional Data.

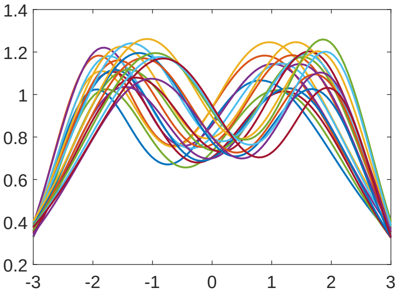
  JASA (to appear). arXiv:1702.01183

- **Software**:

  `Matlab`: http://www.stat.osu.edu/ kurtek.1/code.html

  `R`: Package ‘fdasrvf’.

Variability in functional data

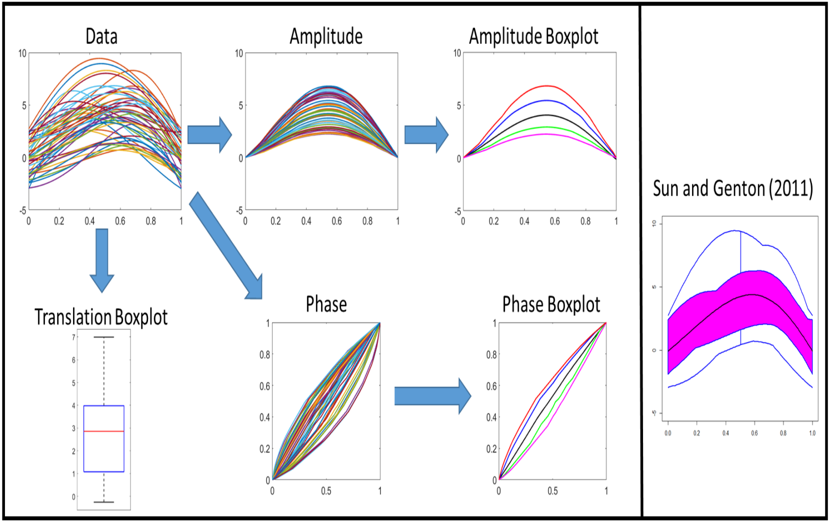


- Translation.
- Amplitude.
- Phase.

Depth, functional quantiles and outliers

1. A. Chakraborty and P. Chaudhuri (2014). *On data depth in infinite dimensional spaces*. AISM.
2. G. Claeskens, M. Hubert, L. Slaets, and K. Vakili (2014). *Multivariate functional halfspace depth*. JASA.
3. R. Fraiman and B. Lopez. *Quantiles for finite and infinite dimensional data* (2012). JMA.
4. M. Hubert, P. J. Rousseeuw, and P. Segaert. *Multivariate functional outlier detection* (2015). Statistical Methods and Applications.
5. S. L-Pintado and J. Romo. *On the concept of depth for functional data* (2009). JASA.
6. Y. Sun and M. G. Genton (2011). *Functional boxplots*. JCGS.

Our approach



Geometry of phase-amplitude separation

Function $f_i : [0,1] \to \mathbb{R}$ are from $\mathcal{F}$, the class of absolutely continuous functions.

Assume that one observes

$$f_i \circ \gamma_i(t) + a_i, \quad 0 \le t \le 1, \quad i = 1, \ldots, n,$$

where $\gamma_i \in \Gamma := \{\gamma : [0,1] \to [0,1],$ increasing, smooth and $\gamma(0) = 0, \gamma(1) = 1\}$. (Group with composition)

Geometry of phase-amplitude separation

Function $f_i : [0,1] \to \mathbb{R}$ are from $\mathcal{F}$, the class of absolutely continuous functions.

Assume that one observes

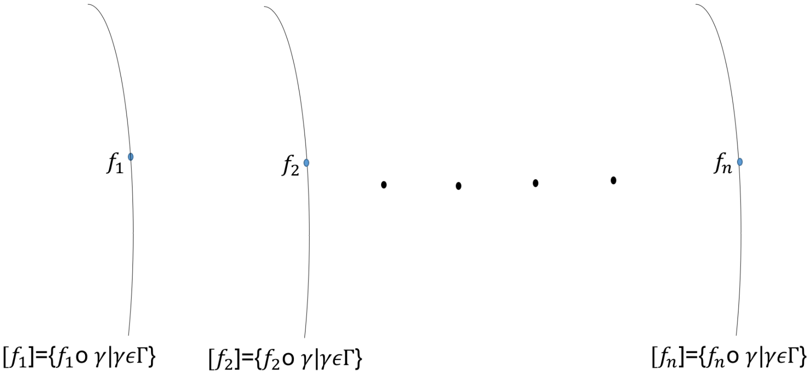
$$f_i \circ \gamma_i(t) + a_i, \quad 0 \leq t \leq 1, \quad i = 1, \ldots, n,$$

where $\gamma_i \in \Gamma := \{\gamma : [0,1] \to [0,1],$ increasing, smooth and $\gamma(0) = 0, \gamma(1) = 1\}$. (Group with composition)

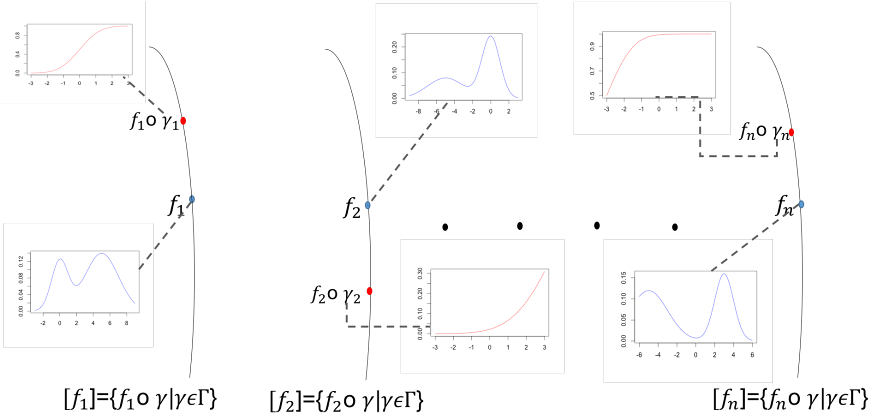
Need to consider equivalence classes:

$$[f_i] := \{f_i \circ \gamma | \gamma \in \Gamma\}, \quad i = 1, \ldots, n.$$

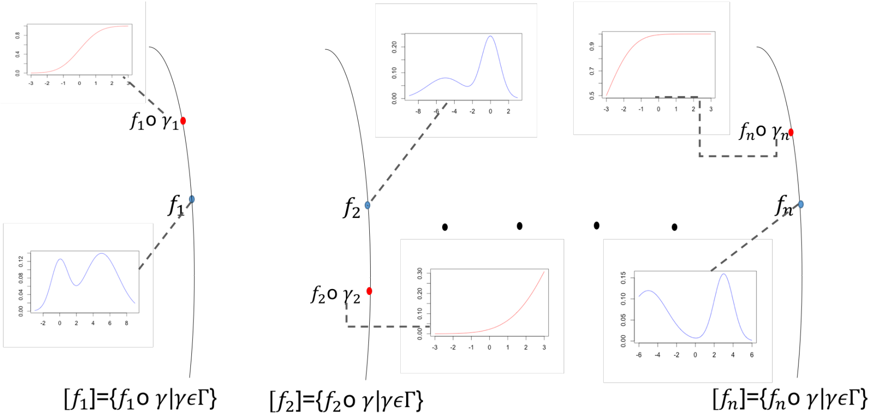
Geometry of phase-amplitude separation



Geometry of phase-amplitude separation



Geometry of phase-amplitude separation



$$\|f_1 \circ \gamma - f_2 \circ \gamma\| \neq \|f_1 - f_2\|$$

Geometry of phase-amplitude separation under Fisher–Rao metric [1]

- **Amplitude**:
  - Transform: $f \mapsto q := \text{sign}(f')\sqrt{|f'|}$ with $(q, \gamma) = q \circ \gamma \sqrt{\gamma'}$.

[1] *Registration of Functional Data Using Fisher-Rao Metric*. A. Srivastava, W. Wu, S. Kurtek, E. Klassen and J. S. Marron. arXiv: 1103.3817

Geometry of phase-amplitude separation under Fisher–Rao metric [1]

- **Amplitude**:
  - Transform: $f \mapsto q := \text{sign}(f')\sqrt{|f'|}$ with $(q, \gamma) = q \circ \gamma \sqrt{\gamma'}$.
  - $\|(q_1, \gamma) - (q_2, \gamma)\| = \|q_1 - q_2\|$

[1] *Registration of Functional Data Using Fisher-Rao Metric*. A. Srivastava, W. Wu, S. Kurtek, E. Klassen and J. S. Marron. arXiv: 1103.3817

Geometry of phase-amplitude separation under Fisher–Rao metric [1]

- **Amplitude**:
  - Transform: $f \mapsto q := \text{sign}(f')\sqrt{|f'|}$ with $(q, \gamma) = q \circ \gamma \sqrt{\gamma'}$.
  - $\|(q_1, \gamma) - (q_2, \gamma)\| = \|q_1 - q_2\|$
  - $D_{\text{amp}}([f_1], [f_2]) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2, \gamma)\|$.

[1] *Registration of Functional Data Using Fisher-Rao Metric*. A. Srivastava, W. Wu, S. Kurtek, E. Klassen and J. S. Marron. arXiv: 1103.3817

Geometry of phase-amplitude separation under Fisher–Rao metric [1]

- **Amplitude**:
  - Transform: $f \mapsto q := \text{sign}(f')\sqrt{|f'|}$ with $(q, \gamma) = q \circ \gamma \sqrt{\gamma'}$.
  - $\|(q_1, \gamma) - (q_2, \gamma)\| = \|q_1 - q_2\|$
  - $D_{\text{amp}}([f_1], [f_2]) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2, \gamma)\|$.
- **Phase Space**:
  - Transform: $\gamma \mapsto \psi := \sqrt{\gamma'} \in S^\infty$ since $\int_{[0,1]} \psi^2(t) dt = 1$.

---

[1] *Registration of Functional Data Using Fisher-Rao Metric*. A. Srivastava, W. Wu, S. Kurtek, E. Klassen and J. S. Marron. arXiv: 1103.3817

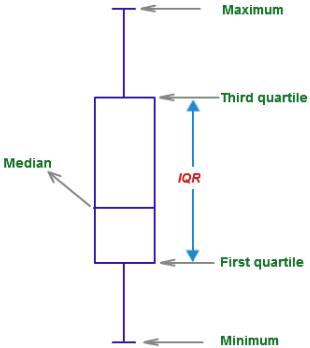
Geometry of phase-amplitude separation under Fisher–Rao metric [1]

- **Amplitude**:
  - Transform: $f \mapsto q := \text{sign}(f')\sqrt{|f'|}$ with $(q, \gamma) = q \circ \gamma \sqrt{\gamma'}$.
  - $\|(q_1, \gamma) - (q_2, \gamma)\| = \|q_1 - q_2\|$
  - $D_{\text{amp}}([f_1], [f_2]) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2, \gamma)\|$.
- **Phase Space**:
  - Transform: $\gamma \mapsto \psi := \sqrt{\gamma'} \in S^\infty$ since $\int_{[0,1]} \psi^2(t) dt = 1$.
  - S^∞ is nonlinear manifold but well-understood.
  - $D_{\text{phase}}(\gamma_1, \gamma_2) = \cos^{-1}(\langle \psi_1, \psi_2 \rangle)$.

[1] *Registration of Functional Data Using Fisher-Rao Metric*. A. Srivastava, W. Wu, S. Kurtek, E. Klassen and J. S. Marron. arXiv: 1103.3817

Boxplot: ingredients

- Median. (Fréchet median on $\mathcal{F}/\Gamma$ and S^∞)
- Upper and Lower quartiles. (Ordering based on distance and direction from Fréchet median).
- IQR
- Outlier cutoffs.
- Extremes.



Amplitude boxplot

- Compute $q_i = \text{sign}(f')\sqrt{|f_i'|}$.

Amplitude boxplot

- Compute $q_i = \text{sign}(f')\sqrt{|f_i'|}$.
- Compute median (equivalence class):

$$[f_m] := \underset{[f]\in\mathcal{F}/\Gamma}{\text{argmin}} \sum_{i=1}^{n} D_{\text{amp}}([f_i],[f]) \quad \text{(Gradient-based algorithm with } q_i\text{)}.$$

- Pick one element f_m from $[f_m]$ and q_m—-this is the median.

Amplitude boxplot

- Compute $q_i = \text{sign}(f')\sqrt{|f_i'|}$.
- Compute median (equivalence class):

$$[f_m] := \underset{[f]\in\mathcal{F}/\Gamma}{\text{argmin}} \sum_{i=1}^{n} D_{\text{amp}}([f_i],[f]) \quad \text{(Gradient-based algorithm with } q_i\text{)}.$$

- Pick one element f_m from $[f_m]$ and q_m—-this is the median.
- Obtain $\{q_{(1)}, q_{(2)}, \ldots, q_{(\lceil n/2 \rceil)}\}$ based on distance from q_m.

Amplitude boxplot

Upper and lower amplitude ‘quartiles’:

Choose two elements from $\{q_{(1)}, q_{(2)}, \ldots, q_{(\lceil n/2 \rceil)}\}$ such that:

- Sum of their distances from the median is the largest.
- They lie on opposite ‘sides’ of the median.

Amplitude boxplot

- Upper and lower amplitude 'quartiles':

$$(q_{Q_1}, q_{Q_3}) := \text{argmax of } (a, b) \text{ over } \{q_{(1)}, q_{(2)}, \ldots, q_{(\lceil n/2 \rceil)}\}$$

$$(1-\lambda)\left[\frac{\|a - q_m\|}{\max_i \|q_{(i)} - q_m\|} + \frac{\|b - q_m\|}{\max_i \|q_{(i)} - q_m\|}\right]$$

$$-\lambda\left[\left\langle \frac{\|a - q_m\|}{\|a - q_m\|}, \frac{\|b - q_m\|}{\|b - q_m\|} \right\rangle + 1\right]$$

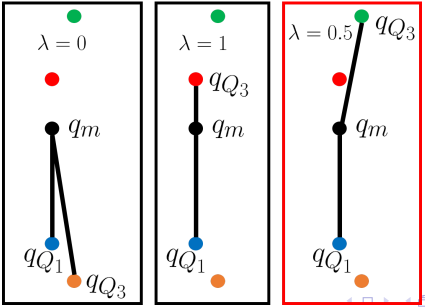
Amplitude boxplot

- Upper and lower amplitude 'quartiles':

$$(q_{Q_1}, q_{Q_3}) := \text{argmax of } (a, b) \text{ over } \{q_{(1)}, q_{(2)}, \ldots, q_{(\lceil n/2 \rceil)}\}$$

$$(1-\lambda)\left[\frac{\|a-q_m\|}{\max_i \|q_{(i)}-q_m\|} + \frac{\|b-q_m\|}{\max_i \|q_{(i)}-q_m\|}\right]$$

$$-\lambda\left[\left\langle \frac{\|a-q_m\|}{\|a-q_m\|}, \frac{\|b-q_m\|}{\|b-q_m\|}\right\rangle + 1\right]$$



Amplitude boxplot

- Inter-quartile range: $IQR = \|q_{Q_1} - q_m\| + \|q_{Q_3} - q_m\|$.

- Amplitude outlier cut-offs:

$$C1 := q_{Q_1} + \kappa \times IQR \times \frac{q_{Q_1} - q_m}{\|q_{Q_1} - q_m\|}$$

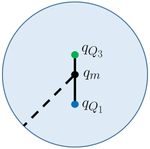
$$C2 := q_{Q_3} + \kappa \times IQR \times \frac{q_{Q_3} - q_m}{\|q_{Q_3} - q_m\|}$$

Amplitude boxplot

- Inter-quartile range: $IQR = \|q_{Q_1} - q_m\| + \|q_{Q_3} - q_m\|$.
- Amplitude outlier cut-offs:

$$C1 := q_{Q_1} + \kappa \times IQR \times \frac{q_{Q_1} - q_m}{\|q_{Q_1} - q_m\|}$$
$$C2 := q_{Q_3} + \kappa \times IQR \times \frac{q_{Q_3} - q_m}{\|q_{Q_3} - q_m\|}$$

f_{new} is an outlier if $\|q_{new} - q_m\| > \max\{\|C_1 - q_m\|, \|C_2 - q_m\|\}$.



Amplitude boxplot: Simulation example

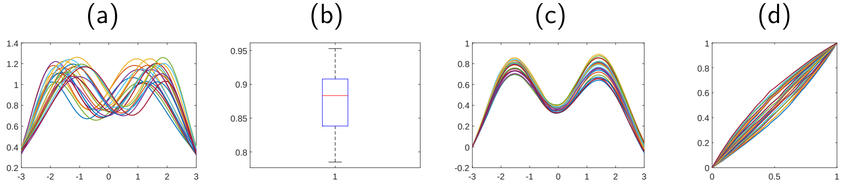


Figure: (a) Original functions. (b) Translation. (c) Amplitude. (d) Phase.

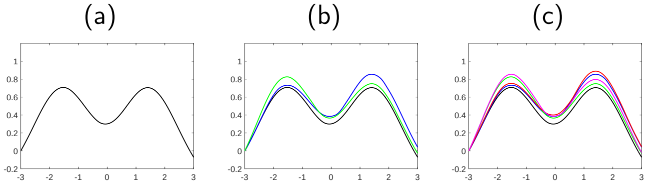
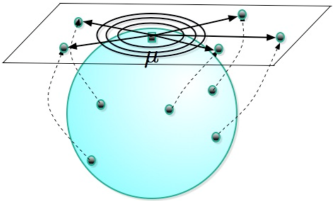


Figure: (a) Amplitude median. (b) Amplitude median with quartiles. (c) Amplitude median with quartiles, and extremes.

Phase boxplot

- $\Gamma \ni \gamma \mapsto \psi := \sqrt{\gamma'} \in S^\infty$.
- Fréchet median computed using $D_{\text{phase}}(\gamma_1, \gamma_2) = \cos^{-1}(\langle \psi_1, \psi_2 \rangle)$.
- Exponential and log-map available analytically.
- Quartiles, cutoffs computed on tangent space of median, and mapped down.



Phase boxplot: Simulation example

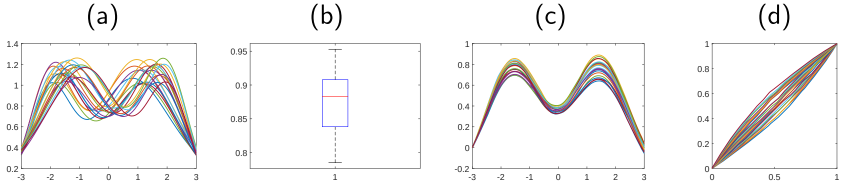


Figure: (a) Original functions. (b) Translation. (c) Amplitude. (d) Phase.

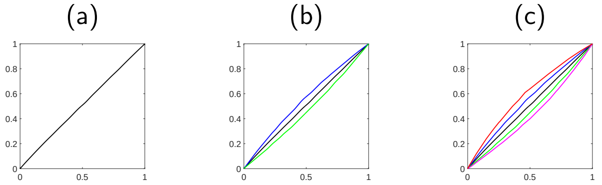
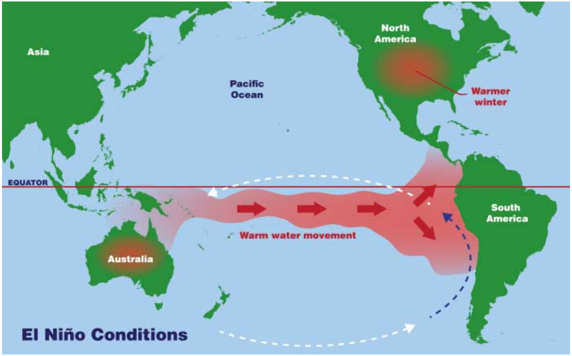


Figure: (a) Phase median. (b) Phase median with quartiles (c) Phase median with quartiles and extremes.

Sea surface temperature: El Niño effect



- Sea surface temperature data from 1950-2014 from Climate Prediction Centre[2].
- Band of warm water in central and east-central equatorial pacific region.

[2] http://www.cpc.ncep.noaa.gov/data/indices/ersst3b.nino.mth.81-10.ascii

Sea surface temperature: Boxplots

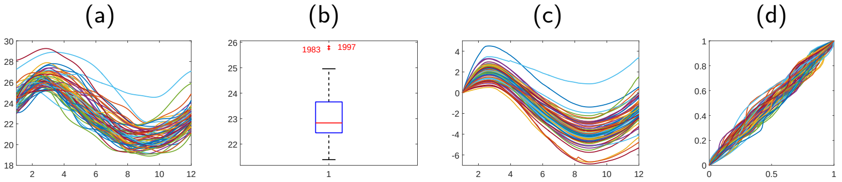


Figure: (a) Original functions. (b) Translation. (c) Amplitude. (d) Phase.

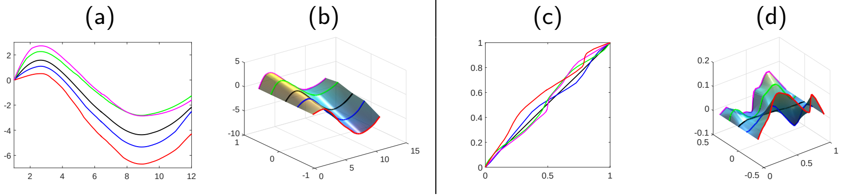


Figure: (a)&(b) Amplitude boxplot and its surface display. (c)&(d) Phase boxplot and its surface display.

Sea surface temperature: Boxplots

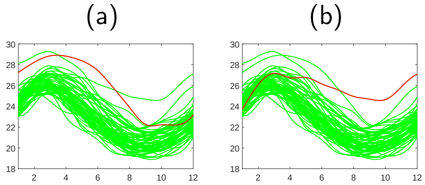


Figure: (a) Translation outlier: 1983. (b) Translation outlier: 1997.

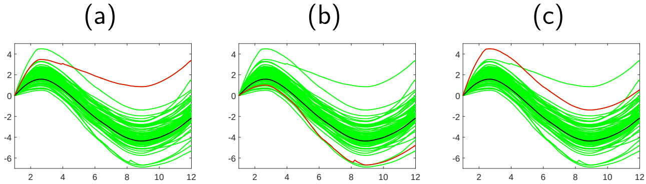


Figure: (a) Amplitude outlier: 1997. (b) Mild amplitude outlier: 2007. (c) Mild amplitude outlier: 1957.