Stats, Math and AI

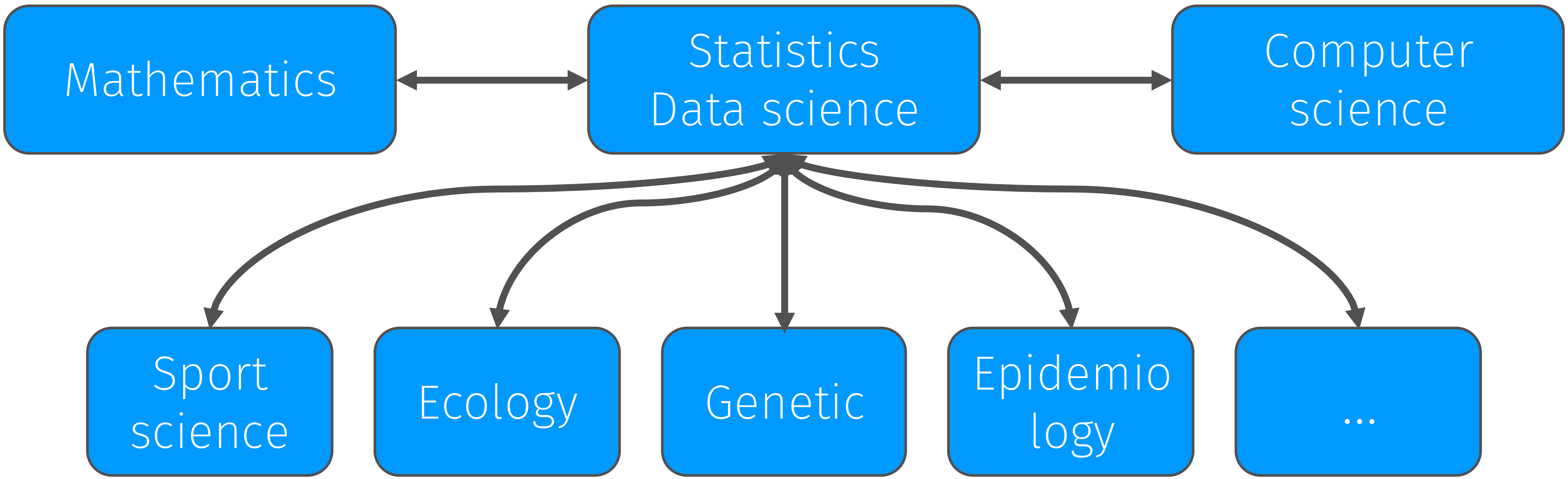
An introduction to object data analysis

Steven Golovkine

October 6th, 2025

UNIVERSITÉ LAVAL
Faculté des sciences et de génie
Département de mathématiques et de statistique

Generated using ChatGPT.



The best thing about being a statistician is that you get to play in everyone's backyard. — John Tukey[1]

[1]see www.princeton.edu/pr/news/00/q3/0727-tukey.htm or The New York Times, July 28, 2000.

AI is the science of making machines capable of performing tasks that would require intelligence if done by humans. — Marvin Minsky[1]

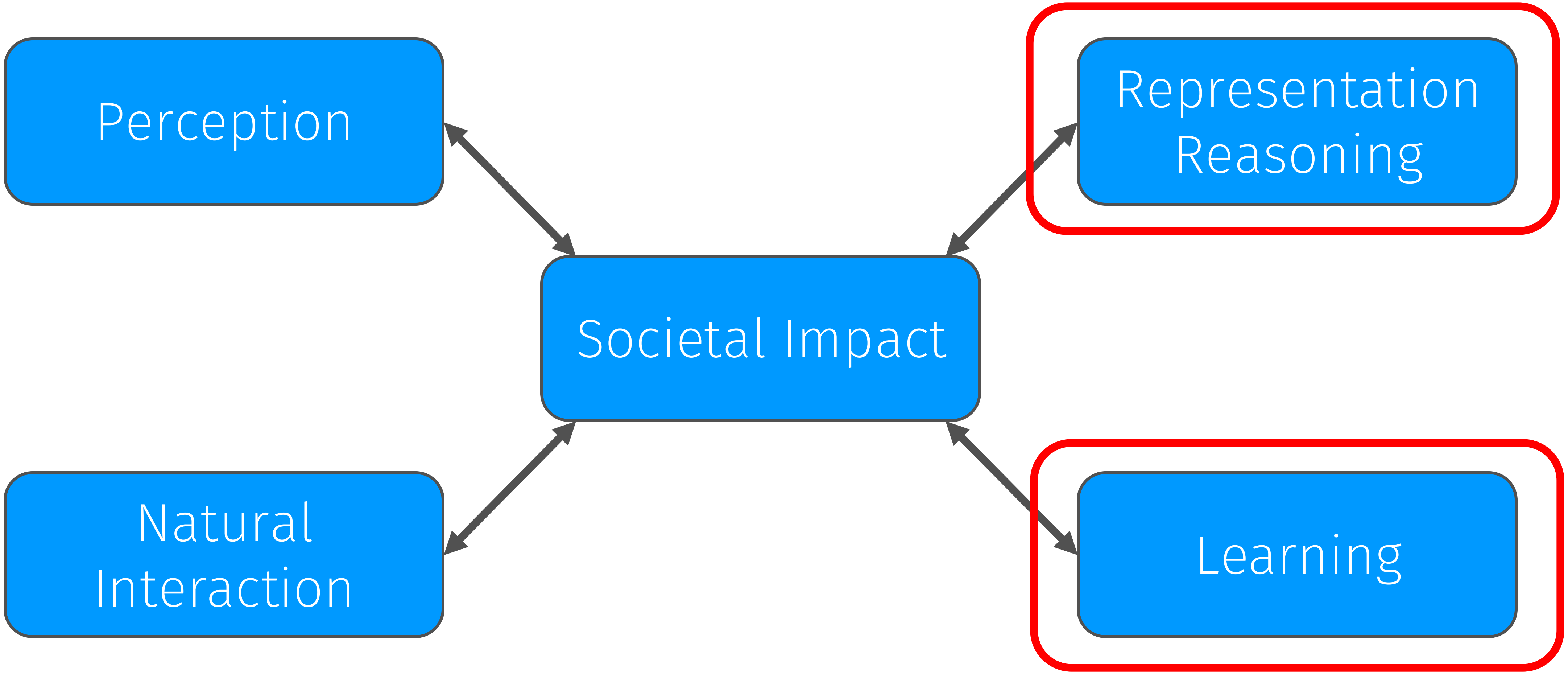
AI is the science and engineering of making machines do tasks they have never seen and have not been prepared for beforehand. — José Hernández-Orallo[2]

The human mind is not merely a collection of special-purpose programs hard-coded by evolution. The mind is not a single, general-purpose “blank slate” system capable of learning anything from experience. — François Chollet

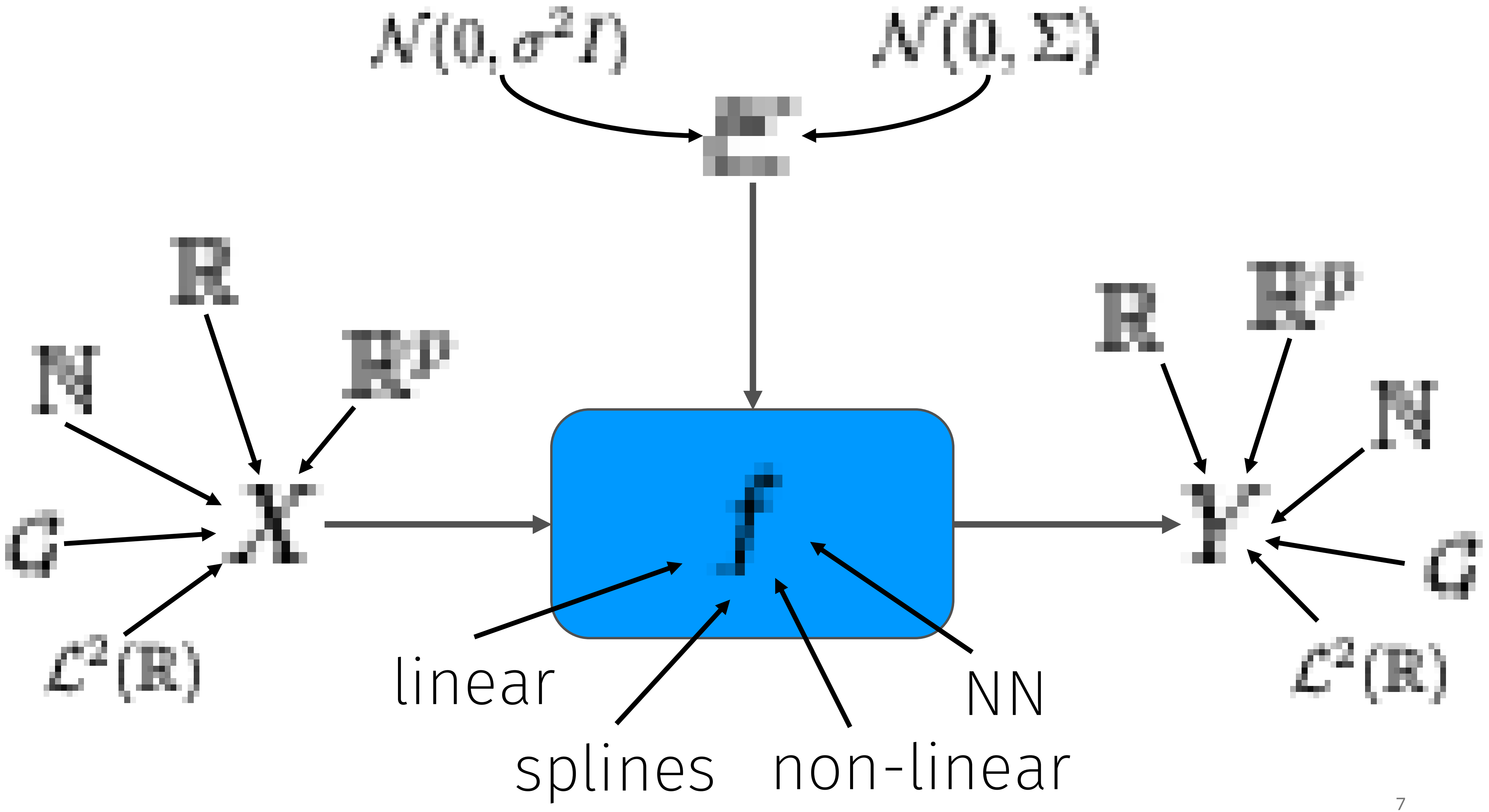
[1]Minsky, M. L. Semantic information processing. Cambridge, MIT Press (2003).

[2]Hernández-Orallo, J. Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. Artificial Intelligence Review (2017).

[3] Chollet, F. On the measure on intelligence (2019).



[1]Minsky, M. L. Semantic information processing. Cambridge, MIT Press (2003).


$\mathcal{N}(0, \sigma^2 I)$
$\mathcal{N}(0, \Sigma)$
$\mathbf{E}$
$\mathbb{R}$
$\mathbb{N}$
$\mathbb{R}^p$
$\mathbb{C}$
$\mathcal{L}^2(\mathbb{R})$
$\mathbf{X}$
f
$\mathbf{Y}$
$\mathbb{R}$
$\mathbb{R}^p$
$\mathbb{N}$
$\mathbb{C}$
$\mathcal{L}^2(\mathbb{R})$
linear
NN
splines
non-linear

$$Y = f(X) + \varepsilon$$

Loss function:

$$\mathcal{L}(Y, f(X)) = (Y - f(X))^2$$

Solution:

$$f(X) = \mathbb{E}(Y \mid X = x)$$

K-nearest neighbors

$$Y = f(X) + \varepsilon$$

$$f(X) = \frac{1}{k} \sum_{i \in N_k(X)} Y_i$$

Linear regression

$$Y = f(X) + \varepsilon$$

$$f(X) = X\beta$$

Student's test

$$Y = f(X) + \varepsilon$$

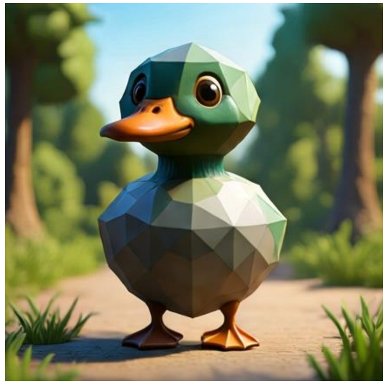
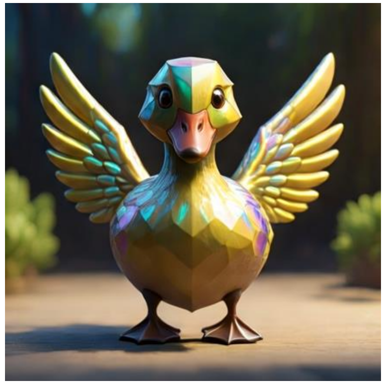
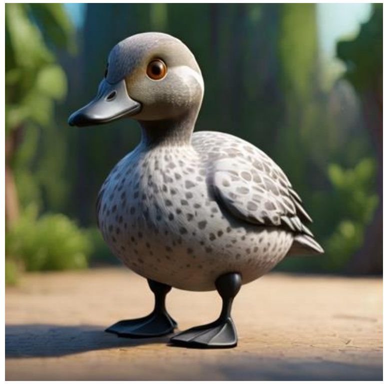
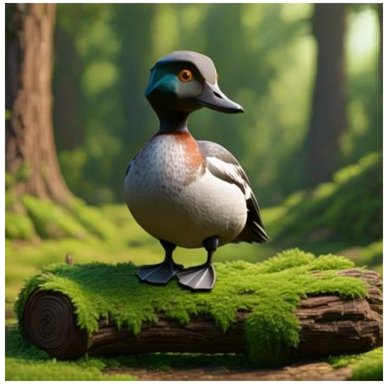
$$f(X) = \beta_0 + \beta_1 X, \quad X \in \{0, 1\}$$

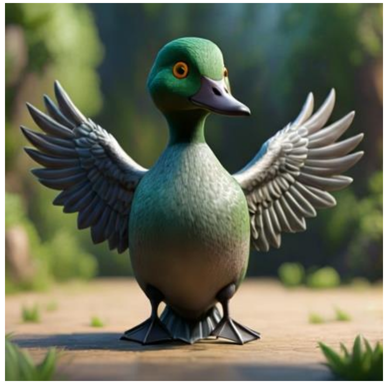
Deep neural network

$$Y = f(X) + \varepsilon$$

$$f(X) = W^{(L)}\sigma(W^{(L-1)}\sigma(\cdots\sigma(W^{(1)}X + b^{(1)})\cdots) + b^{(L-1)}) + b^{(L)}$$

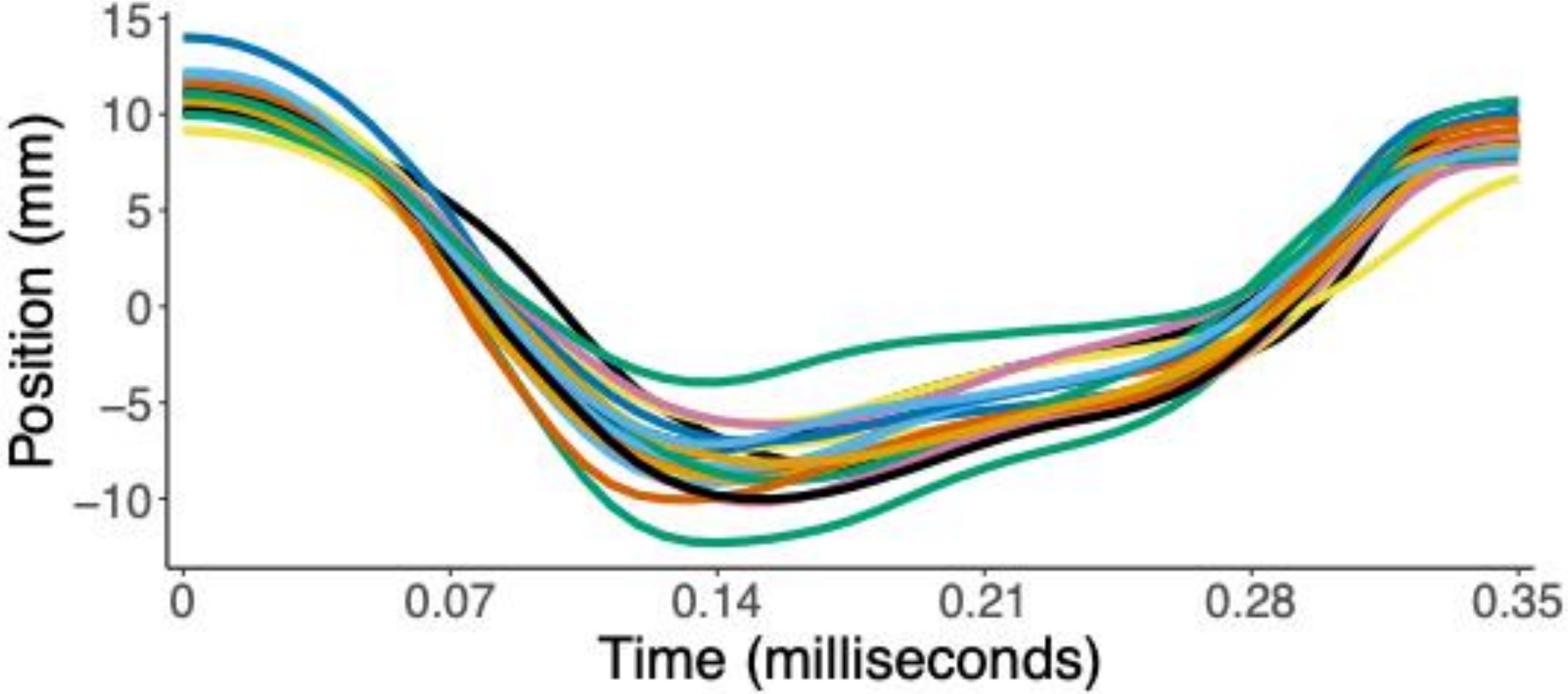
~~$\mathcal{X} = \{duck\}$~~



Ceci n'est pas un canard.

Generated using Apple Intelligence.



Ramsay, J. and Silverman B., Functional Data Analysis, Springer (2005).

Karhunen-Loève decomposition

$$X(t) = \sum_k a_k \phi_k(t)$$

Solution:

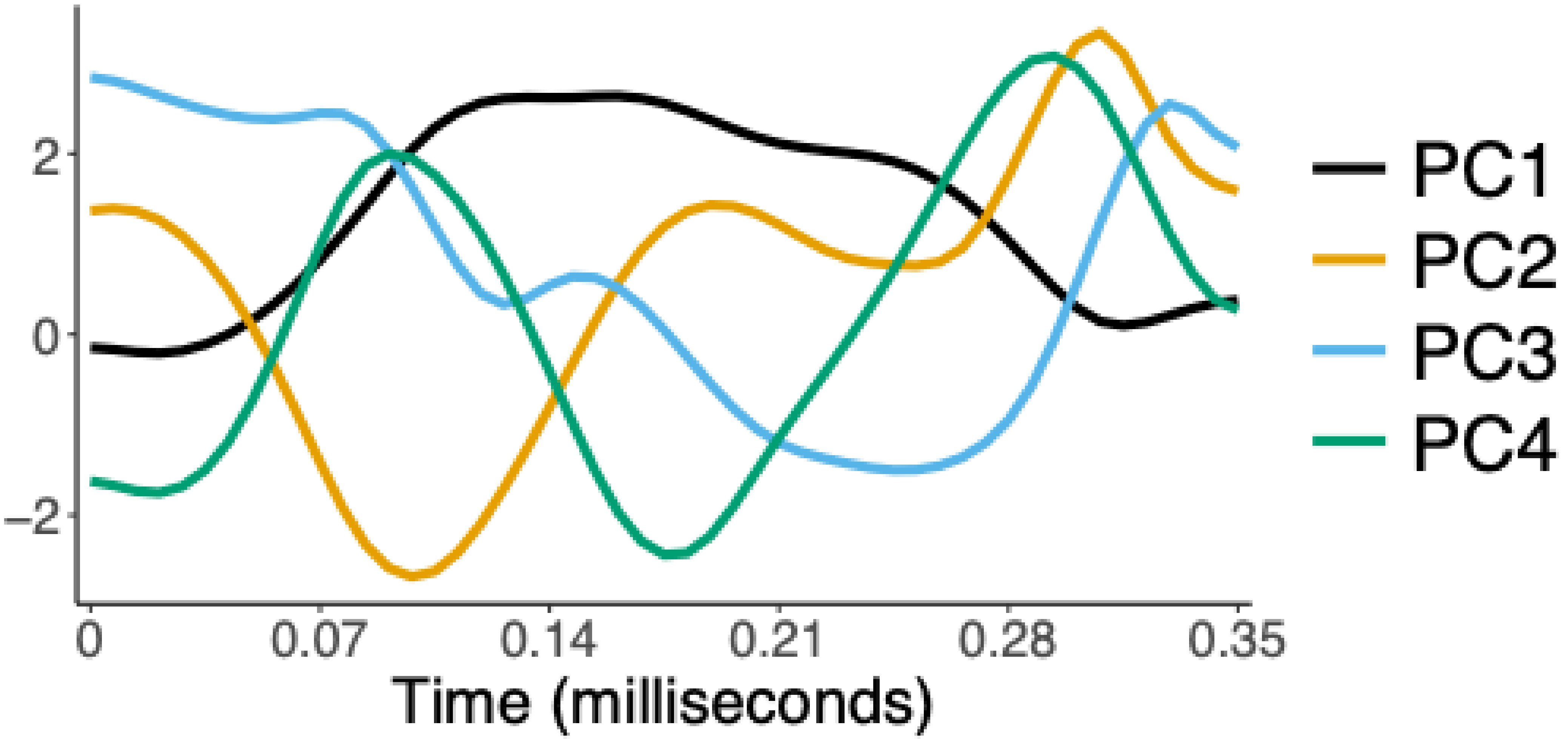
ϕ_k eigenfunctions of the covariance function C

$a_k = \langle X, \phi_k \rangle$

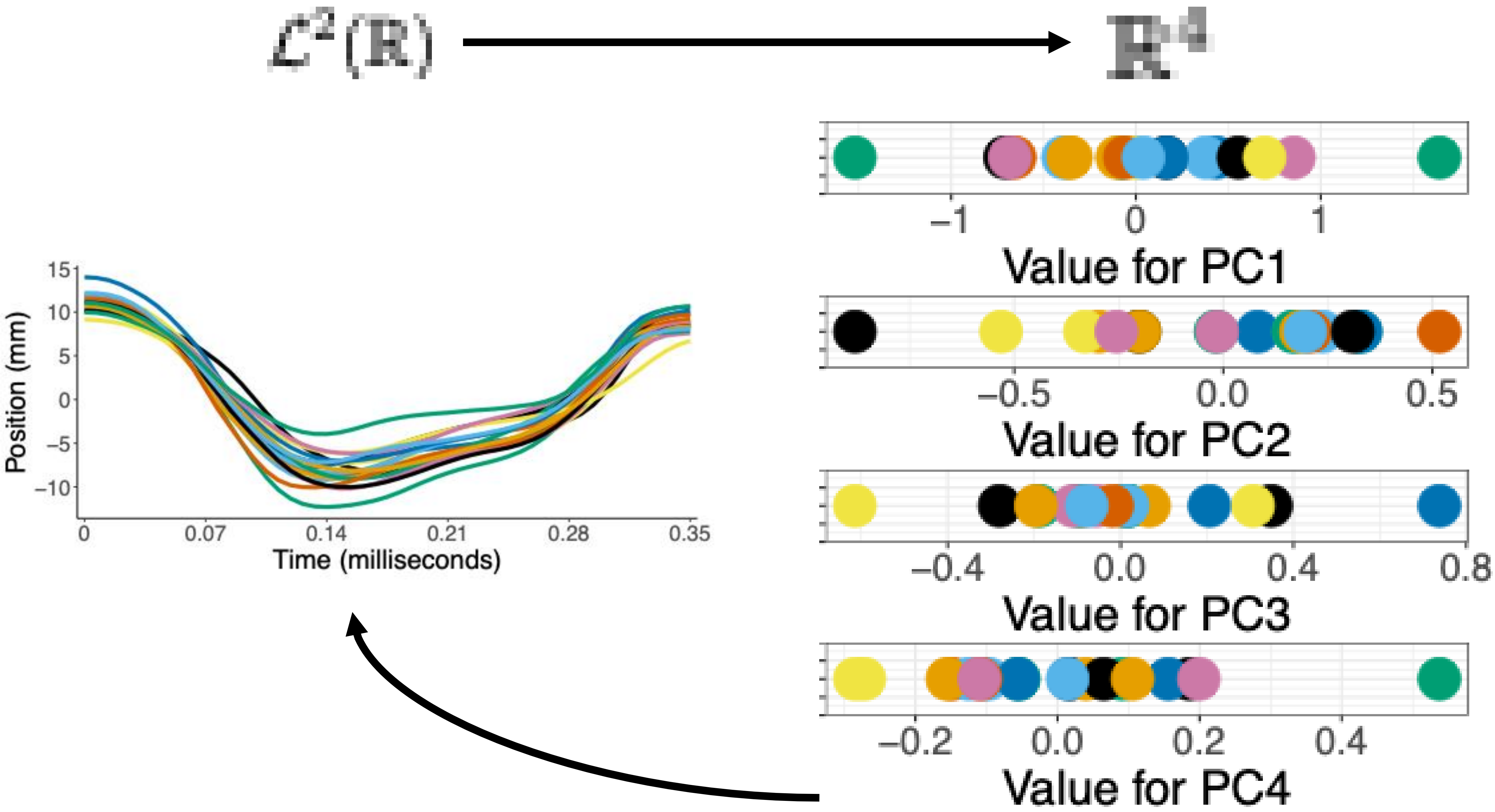
Proof:

Study of the operator $\Gamma f = \int C(s, \cdot) f(s) \mathrm{d}s$

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).



2
0
−2
0
0.07
0.14
0.21
0.28
0.35
Time (milliseconds)
PC1
PC2
PC3
PC4


$\mathcal{L}^2(\mathbb{R})$
$\mathbb{R}^4$
Position (mm)
Time (milliseconds)
15
10
5
0
−5
−10
0
0.07
0.14
0.21
0.28
0.35
−1
0
1
Value for PC1
−0.5
0.0
0.5
Value for PC2
−0.4
0.0
0.4
0.8
Value for PC3
−0.2
0.0
0.2
0.4
Value for PC4

$$Y = f(X) + \varepsilon$$

$$Y = \langle X, \phi \rangle$$

$$Y = \langle X, \sum_k a_k X_k \rangle = \sum_k a_k \langle X, X_k \rangle$$

Golovkine, S. et al., On the use of the Gram matrix for multivariate functional principal components analysis (2025).

Principal differential analysis

$$L = \beta_0 I + \beta_1 D + D^2$$

$$\frac{d^2}{dt^2} X(t) = -\beta_1(t) \frac{d}{dt} X(t) - \beta_0(t) X(t)$$

Objective:

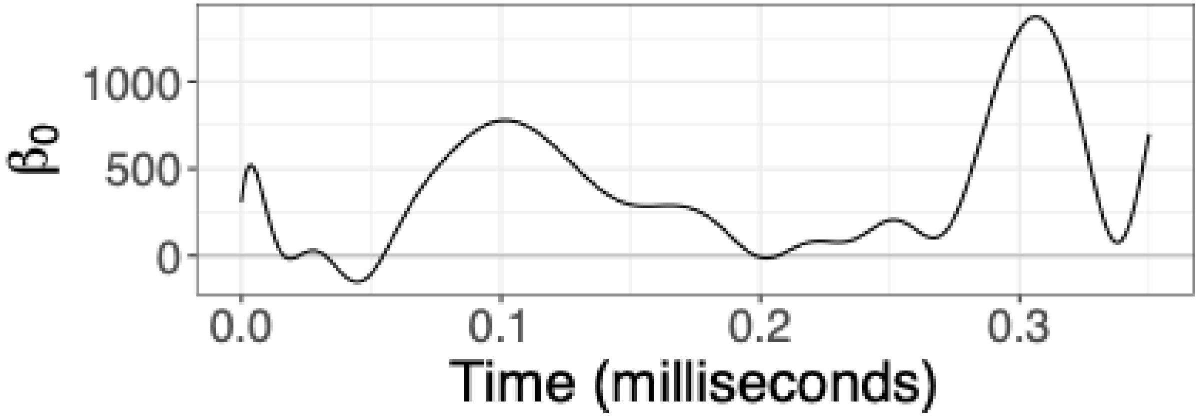
Find L such that $LX \approx 0$

Find ϕ such that $L\phi = 0$

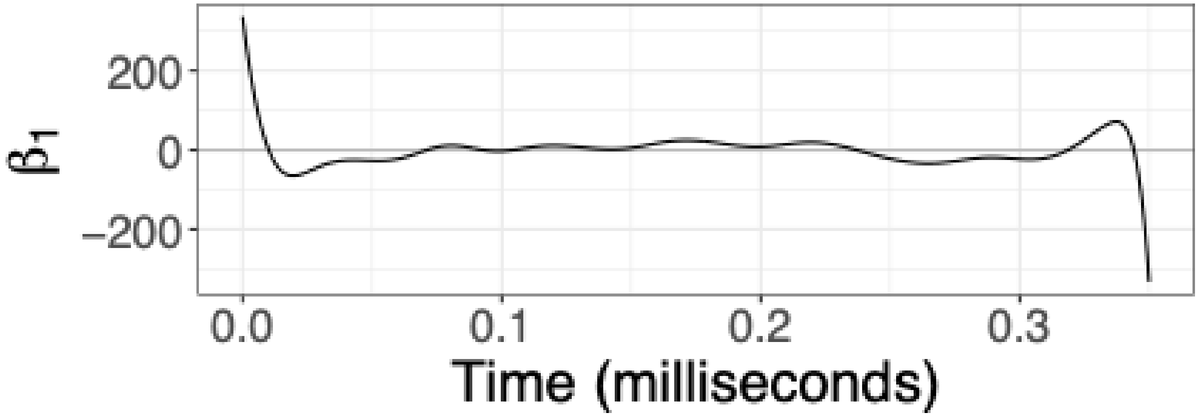
Solution:

Estimation by minimization of least-squares.

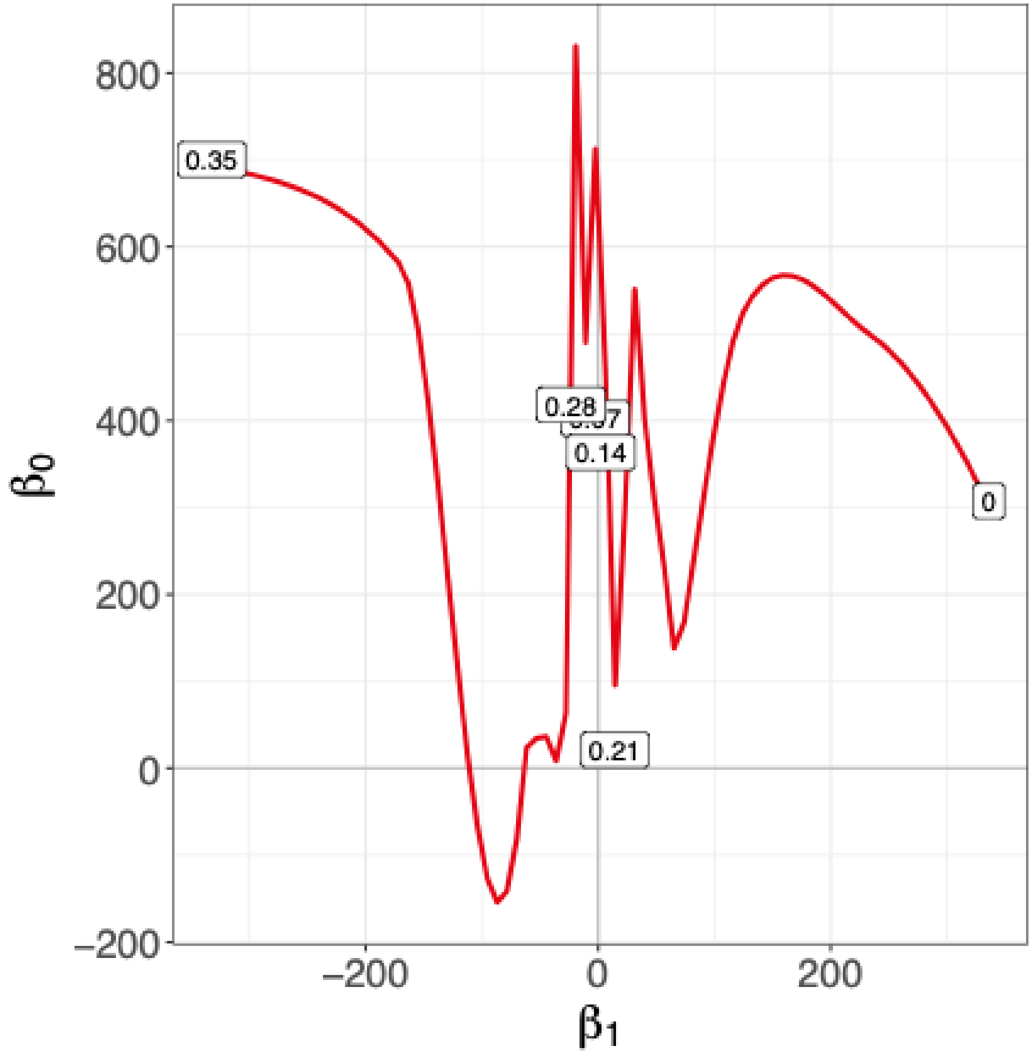
Ramsay, J. O., Principal Differential Analysis: Data Reduction by Differential Operators (1996).



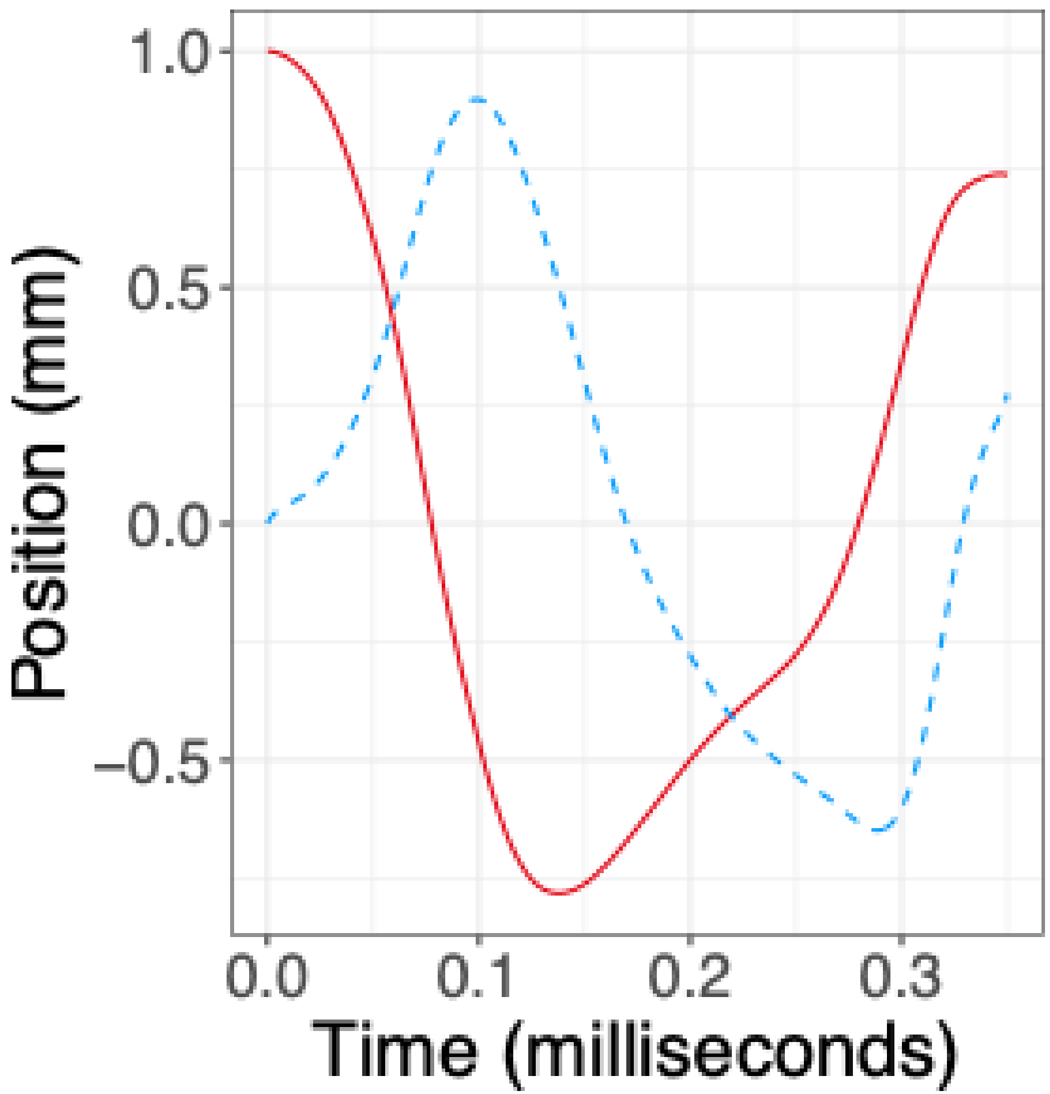
β0
1000
500
0
0.0
0.1
0.2
0.3
Time (milliseconds)



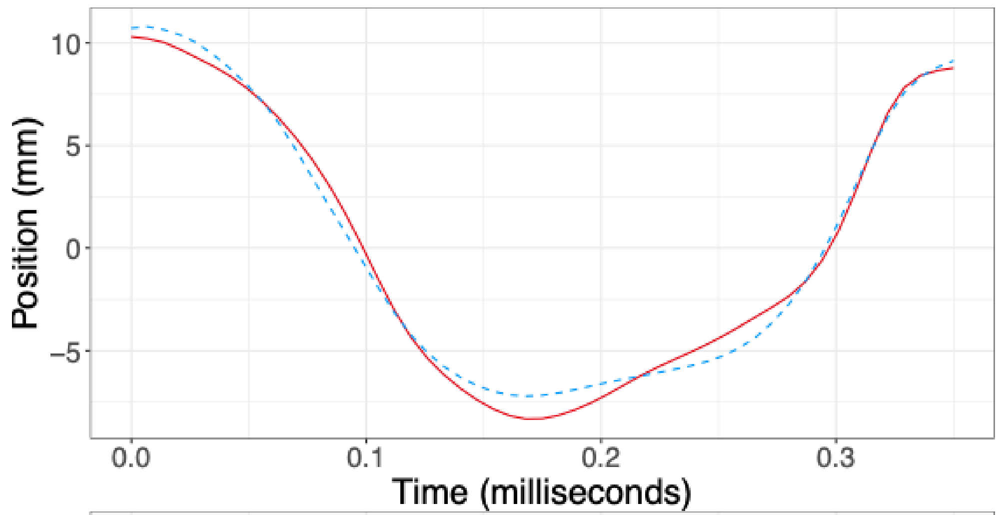
β1
200
0
−200
0.0
0.1
0.2
0.3
Time (milliseconds)



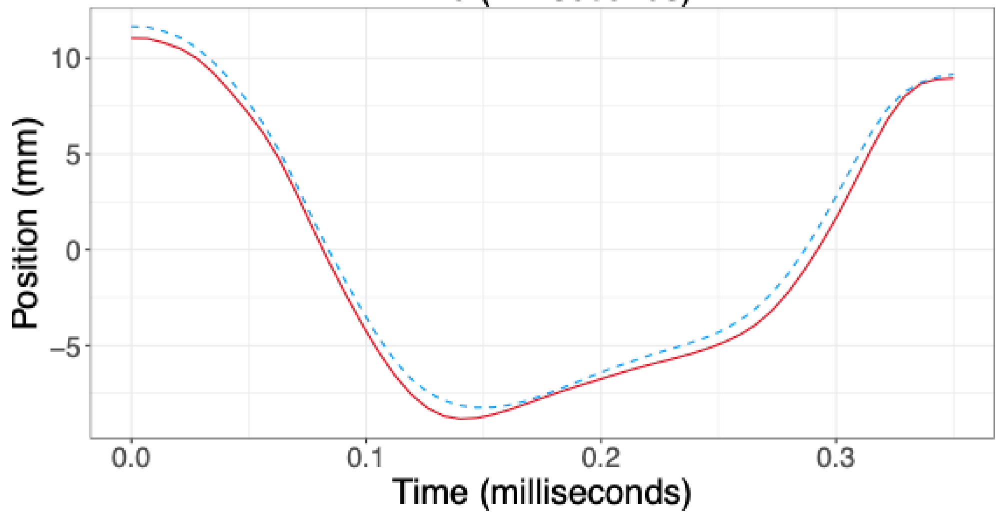
β0
800
600
400
200
0
−200
0.35
0.28
0.14
0.21
0
−200
0
200
β1



1.0
0.5
0.0
−0.5
Position (mm)
0.0
0.1
0.2
0.3
Time (milliseconds)



10
5
0
−5
Position (mm)
0.0
0.1
0.2
0.3
Time (milliseconds)



10
5
0
−5
Position (mm)
0.0
0.1
0.2
0.3
Time (milliseconds)

Lots of stuff to do in ODA

and we need math development for that…

Stop being humble: you’re doing AI!

$\mathbb{R}$ $\mathbb{R}^p$

$\mathbb{N}$ $\mathbb{C}$

$\mathcal{L}^2(\mathbb{R})$

Duck space