Analysis of Handwritten Spirals

Alejandro C. Frery

DATA 301

1 Introduction

This project aims at starting an in-depth analysis of handwritten spirals, whose data were collected as part of a study with patients of Parkinson disease.

The initial scientific question is the following:

> Which are good descriptors that allow discriminating from control and ill patients?

In this first stage, the idea is to propose descriptors based on two examples, one control and one ill patient.

2 Data

Each patient is requested to draw a spiral between two guiding lines using a tablet. The data consists of ASCII files with one header line containing the number of subsequent lines, followed by seven columns with the time-ordered x and y coordinates, the time stamp, the on/off state of the pen, the azimuth, the altitude, and the pressure. An example of the first ten lines of an actual file is the following:

```
4183
5310 3728 1845198 1 3280 810 238
5311 3726 1845206 1 3280 810 320
5311 3726 1845213 1 3280 810 302
5312 3725 1845221 1 3280 810 260
5313 3725 1845228 1 3280 810 244
5315 3724 1845236 1 3280 810 206
5317 3724 1845243 1 3250 810 216
5319 3724 1845251 1 3250 810 260
5319 3724 1845251 1 3250 810 260
```

Fig. 1 illustrates what the azimuth and altitude of a pen measure.

Fig. 2 shows two examples of control spirals (produced by a healthy subject), and an example produced by an ill patient.

These plots were produced with the following `R` code:

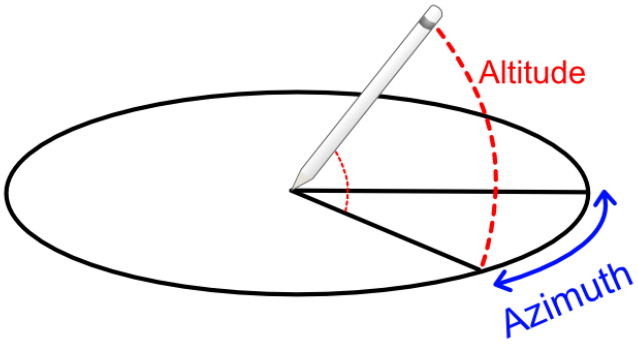

Figure 1: Azimuth and altitude of a pen (Source: https://www.raywenderlich.com/1407-apple-pencil-tutorial-getting-started).

```
require(readr)
require(ggplot2)
require(ggthemes)

theme_set(theme_clean())

u00003s00002_hw00001 <- read_table2("u00003s00002_hw00001.svc",
     col_names = FALSE, skip = 1)
names(u00003s00002_hw00001) <- c("x", "y", "Time", "On/Off",
     "Azimuth", "Altitude", "Pressure")


ggplot(u00003s00002_hw00001, aes(x=x, y=y)) +
geom_point(size=.01) +
coord_fixed()

u00005s00001_hw00001 <- read_table2("u00005s00001_hw00001.svc",
      col_names = FALSE, skip = 1)
names(u00005s00001_hw00001) <- c("x", "y", "Time", "On/Off",
      "Azimuth", "Altitude", "Pressure")


ggplot(u00005s00001_hw00001, aes(x=x, y=y)) +
geom_point(size=.01) +
coord_fixed()
```

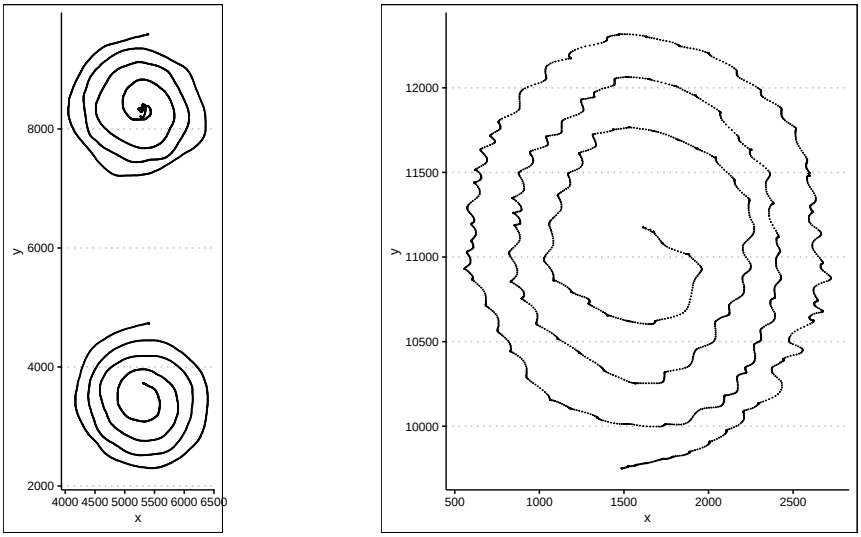

Figure 2: Control (left) and ill (right) subjects results

3 Starting point

The reports should start by describing the source of the data, and concerns about patients' privacy.

As a mere suggestion, the project may start by rectifying the data by fitting it to a spiral data model (see, for instance, Mishra 2004), then by making an EDA, followed by classical time series analysis (Cowpertwait and Metcalfe 2009), and by descriptors of statistical complexity (Zunino et al. 2017).

References

Cowpertwait, P. S. P. and Metcalfe, A. V.: 2009, *Introductory Time Series with R*, Springer.

Mishra, S. K.: 2004, An algorithm for fitting archimedean spiral to empirical data, *SSRN Electronic Journal* .
URL: *https://ssrn.com/abstract=531542*

Zunino, L., Olivares, F., Bariviera, A. F. and Rosso, O. A.: 2017, A simple and fast representation space for classifying complex time series, *Physics Letters A* **381**(11), 1021–1028.