# Application of Multivariate functional principal component analysis on high-dimensional gait data of children with cerebral palsy.

Sajal Kaur Minhas*, Michelle Carey*, Morgan Sangeux** and Julia Polak***
*University College Dublin, ** Murdoch Childrens Research Institute, Australia and *** University of Melbourne, Australia
Email: sajal.minhas@ucdconnect.ie, Twitter: https://twitter.com/SajalUCD , LinkedIn: https://www.linkedin.com/in/sajalkaurminhas/ and GitHub: https://github.com/Sajal010/MFPCA_gaitanalysis.

## Background

The Royal Childrens Hospital, Australia recorded the movement on nine kinematic variables collected from 811 children, 776 children of which had cerebral palsy. For each child, measurements on nine kinematic variables on each leg representing the movement of :

- the pelvis in left-right direction, forward-backward direction and upward-downward direction,
- the hip in left-right direction, forward-backward direction and upward-downward direction,
- the knee in left-right direction,
- the ankle in left-right direction,
- and the footprogression in left-right direction

were recorded over the entire gait cycle ranging over 101 time points. The gait cycle is measured from heel-strike to heel-strike (i.e. when one-foot contacts the ground to when that same foot again contacts the ground) as shown in Figure(1). The data has considerably different gait patterns over the different kinematic variables as well as large variation among subjects.

### Aim of this work:

- To calculate the amount by which a subject's gait deviates from an average normal profile and to represent this deviation as a single number.
- Such a measure can quantify the overall severity of a condition affecting walking, monitor progress, or evaluate the outcome of an intervention prescribed to improve the gait pattern.

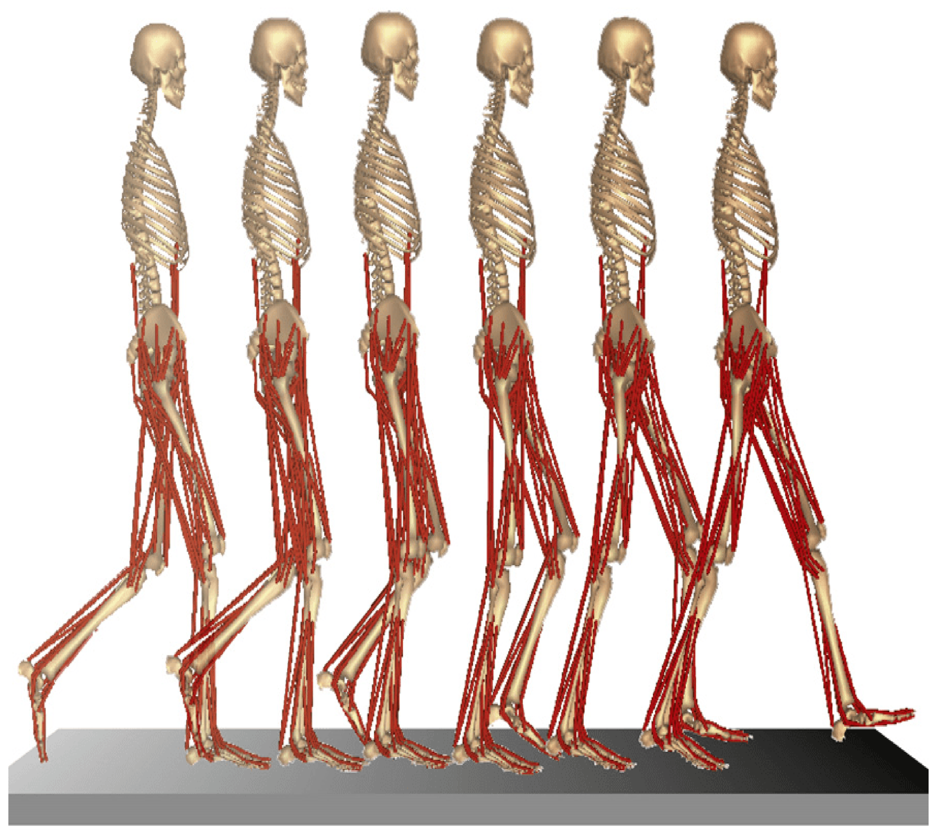

Figure 1: Gait pattern of a person.

## 1. Existing methods

The Gait Deviation Index (GDI) [1] and Gait Profile Score (GPS) [2] are the standard indices for measuring gait abnormality such as cerebral palsy, rheumatoid arthritis and Parkinson's disease.

### GDI:

- is easy to interpret and is normally distributed allowing for standard parametric statistical testing.

### GPS:

- has the ability to decompose scores by individual joints/planes and altered indices without the need for a large control database but it is not normally distributed.

### Both the GDI and GPS do not account for:

- The potential co-variation between the kinematic variables for any individual subject, i.e. the motions of one joint affect the motions of adjacent joints.
- The position of a joint at one time affects the positions at a later instant.

## 4. Advantages of MFPCA

1. A stable quantification of the overall severity if any child is added or removed the ranking does not deviate widely.
2. It has the advantages of both the approaches such as normality assumption (which GPS doesn't have) as well as analysis of separate joints (which GDI doesn't have).
3. Classification accuracy of Cases: New GDI Index (97.6%), GPS (58.3%) whereas GDI for left leg (84.8%) and GDI for right leg (83.6%).

## 5. References

[1] Michael H Schwartz and Adam Rozumalski. The gait deviation index: a new comprehensive index of gait pathology. *Gait & posture*, 28(3):351–357, 2008.
[2] Richard Baker, Jennifer L McGinley, Michael H Schwartz, Sarah Beynon, Adam Rozumalski, H Kerr Graham, and Oren Tirosh. The gait profile score and movement analysis profile. *Gait & posture*, 30(3):265–269, 2009.
[3] Clara Happ and Sonja Greven. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659, 2018.

## 6. Acknowledgements

This work has emanated from research conducted with financial support of Science foundation of Ireland under Grant No. 18/CRT/6049. The opinions, findings and conclusions or recommendations expressed in this material are those of author(s) and do not necessarily reflect the views of the Science Foundation Ireland.

## 2. Multivariate functional principal component analysis (MFPCA)

Multivariate functional data typically comprises of several recorded time course measurements for a sample of subjects. The term functional refers to the intrinsic structure of the data, i.e., the belief that for each subject the movement in each joint/direction, $j = 1, \dots, Q$ (where $Q$ is the number of movements recorded at each direction for each joint), which are being generated by some underlying function $X_j$ and the discrete measurements collected $X_1(t_1), \dots, X_Q(t_N)$ are a snapshot of that function at various points in time $t_1, \dots, t_N$.
MFPCA attributable to [3] to determine the main modes of variation in the overall gait movement. MFPCA produces a multivariate functional Karhunen-Loéve representation of the data,

$$\hat{X}_j(t) \approx \hat{\mu}_j(t) + \sum_{w=1}^{W} \hat{\rho}_w \hat{\psi}_w^{(j)}(t), \tag{1}$$

where $\hat{\mu}_j(t)$ is an estimate of the mean function for the $j^{th}$ joint/direction across all subjects, $\hat{\rho}_w$ is the $w^{th}$ principal component score and $\hat{\psi}_w^{(j)}(t)$ is the $w^{th}$ principal component function for the $j^{th}$ joint/direction. The number of principal components $W$ is determined by the percentage of variance explained.

### MFPCA steps:

1. Calculate a univariate FPCA for each joint/direction $j = 1, \dots, Q$. This results in principal component functions $\hat{\phi}_1^j, \dots, \hat{\phi}_{K_j}^j$ and principal component scores $\hat{\xi}_1^j, \dots, \hat{\xi}_{K_j}^j$, for each subject $i = 1, \dots, M$ with suitably chosen truncation lags $K_j$.
2. Combine all coefficients into one big matrix $\Xi \in R^{M \times K_+}$ with $K_+ = K_1, \dots, K_Q$, having rows $\Xi_i = (\hat{\xi}_1^{i,1}, \dots, \hat{\xi}_{K_1}^{i,1}, \dots, \hat{\xi}_1^{i,Q}, \dots, \hat{\xi}_{K_Q}^{i,Q})$, and estimate the joint covariance matrix $\hat{Z} = \frac{1}{M-1}\Xi^T \Xi$.
3. Find eigenvectors $\hat{c}_w$ and eigenvalues $\hat{\nu}_w$ of $\hat{Z}$ for $w = 1, \dots, W$, for some truncation lag $W < K_+$.
4. Calculate the estimated multivariate principal component functions $\hat{\psi}_w$ and scores $\hat{\rho}_{i,w}$ based on the results from steps 1 and 3:
$$\hat{\psi}_w^{(j)} = \sum_{k=1}^{K_j} [\hat{c}_w]_k^{(j)} \hat{\phi}_k^{(j)}, \quad \text{and} \quad \hat{\rho}_{i,w} = \sum_{j=1}^{Q} \sum_{k=1}^{K_j} [\hat{c}_w]_k^{(j)} \hat{\xi}_k^{i,j}, \quad \text{for} \quad j = 1, \dots, Q.$$
5. Evaluates the Euclidean distance between the principal component scores of a subject with gait abnormalities $\hat{\rho}_{i,w}$ relative to the average principal component scores of all subjects without gait abnormalities $\bar{\hat{\rho}}_w^{(Control)}$, which we denote by $d_{i,w} = \|(\hat{\rho}_{i,w} - \bar{\hat{\rho}}_w^{(Control)}\|$.
6. For scalability, we divide the integral of the distances $d_{i,w}$ by its maximum $\max(\mathrm{d}_w)$, i.e. $scd_{i,w} = \frac{d_{i,w}}{\max(\mathrm{d}_w)}$. The proposed functional gait deviation index for subject $i$ is $\mathrm{FGDI}_i = log\,(d_{i,w})$. The functional GDI index is bounded between 0 and 1, where 0 indicates no deviation from the average normal profile and 1 represents a subject exhibiting the largest deviation from the average normal profile for all components $w = 1, \dots, W$ and all kinematic variables $j = 1, \dots, Q$.

## 3. Results

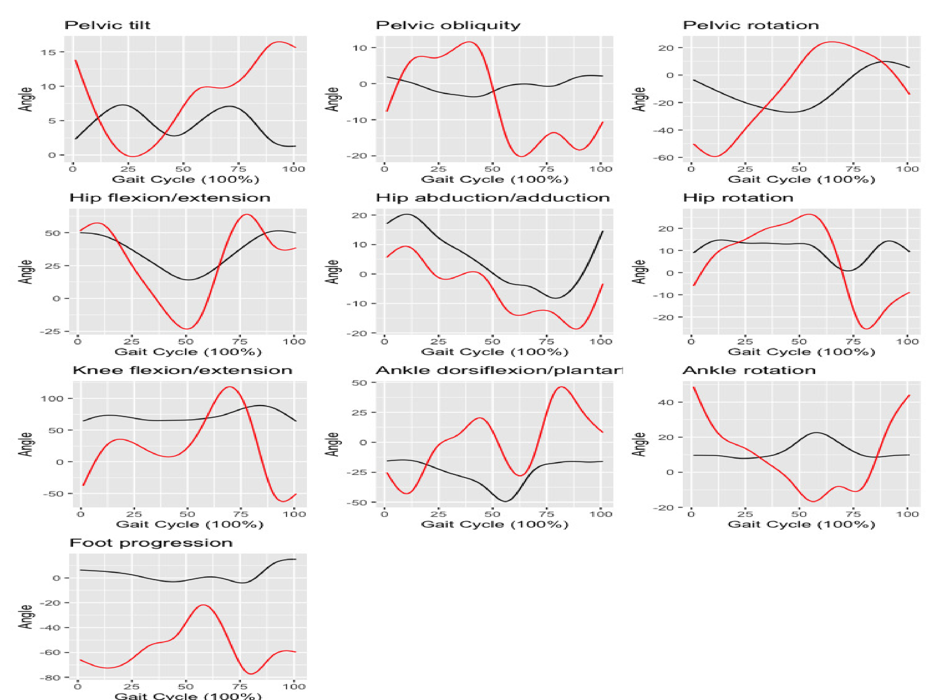


Figure 2: Movement of most abnormal behaviour and average healthy children across all joints.

In Figure(2), the movement on the L.H.S. across the gait cycle for the subject with the highest gait abnormalities (red in color) as indicated by the index relative to the movement of the average principal component scores of all subjects without gait abnormalities (black in color) for each of the joints.

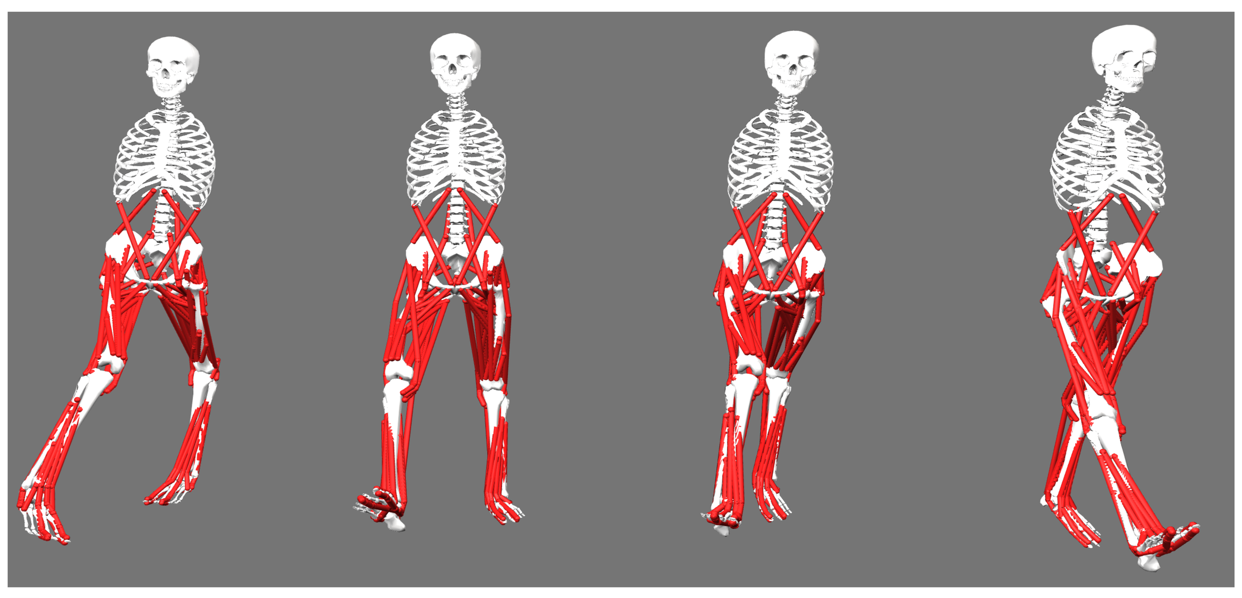

Figure 3: The movement of different level of severity in children.

In Figure(3), the movement of different level of severity using new GDI index in children is depicted where left corner image shows the most abnormal behaviour given by this new index, second person from left shows the 3rd quantile of new index abnormality in children, third image from left gives the children with median i.e., 50 quantile abnormal behaviour given by the new index and lastly, right corner image shows the movement of walking stride of normal children given by the proposed index. This plot helps us to distinguish between different pattern of walking among different level of abnormality in children.