Principal Autoparallel Analysis: Data Analysis in Weitzenböck Space

Stephen Marsland
Massey University, New Zealand
(Joint work with Carole Twining)
Motivation

- Suppose you have data in a non-Euclidean data space
- All statistics are now local
  - Curvature means the connection is path-dependent
- Three solutions:
  - Suck it up
    - Only do local stats
  - Linearise locally
    - Locally linear embedding and variants
  - Find some other properties of manifold
Riemannian Metric on Data Manifolds

- Consider infinitesimal variations at a datapoint $c$
- $c' = c(t) + e \ h(t)$
- Riemannian metric is assignment of an inner product between elements of the tangent space at $c$

This is point-to-point correspondence
**Metric & Connection**

- **Point** = data point
- **Vector(s)** = infinitesimal changes of params
- **Metric** $g$: compute dot product
  \[ (X, Y) = g_{\mu\nu} X^\mu Y^\nu \]
- Compare vectors
- Length -> distance between points
  \[ (X, X) = \|X\|^2 = g_{\mu\nu} X^\mu X^\nu \]

**Diagram Notes:**
- Points = two data points
- Relate directions at two different points A & B
- Transport $Y$ from B to A stays ‘parallel’
- Parallel transport defined by the connection
- Compare vectors at two different points = define derivative
Connection

- Mapping between elements of tangent spaces at different points by parallel transport
- There are lots of choices of the connection
- There exists a unique torsion-free connection
- Which provides a (rough) correspondence
Example: Surface of a Sphere

- Connection = constant angle relative to great circle
- After the loop, not the same vector!
- Change in vector = curvature
Summary: Metrics & Correspondence

- Need a metric to define distances between data points
  - Can’t do statistics without distances!?
- In Riemannian geometry metric defines connection
- Curvature = no consistent direction possible
- We want a consistent direction/mode(s)
- Solution: Stop using Riemannian geometry
- Question: What other options???
Alternative

● Don’t define connection based on metric
● Use a set of vector fields that form a basis of $T_cS$
  ■ Fibre over c of TS
● This is a tetrad field (zweibein, dreibein, …)

\[
h_i(x) \cdot h_j(x) = \delta_{ij}
\]

\[
\Rightarrow g_{\mu\nu} h_i^{\mu}(x) h_j^{\nu}(x) = \delta_{ij}
\]
Weitzenbock Spaces: Absolute parallelism

- Distances: Euclidean metric
- Directions: Frame defined at each/every point
- Connection: parallel-transports frames
  - Zero curvature by construction
- Shortest lines: geodesics
- ‘Straightest’, (constant) direction: autoparallels
Riemann-Cartan Space

- Non-Euclidean space with both torsion and curvature
- Geometry described by metric and affine connection
- Reduces to Riemann space if torsion is identically zero
- And Weitzenbock space (which is embedded in Euclidean space) if curvature is identically zero
Torsion & Curvature

Weitzenbock Spaces

Riemann-Cartan Spaces

Euclidean Spaces

Riemannian Spaces

$T \neq 0$

$T = 0$

$C \neq 0$

$C = 0$
Torsion

● Invented by Cartan in 1922
● Cartan tells Einstein about it in Paris 4 weeks later
● Einstein writes about teleparallelism in 1928 without referencing Cartan
● Cartan writes to complain to Einstein, especially annoyed since Weitzenböck wrote about torsion in 1923 without even mentioning Cartan
● But Weitzenböck also got his comeuppance since Vitali noted that he never actually wrote down the connection, and hence did so in 1925
VORWORT

Neue Bücher entspringen in der Regel einem Zeitbedürfnisse. Ich glaube mit diesem Buche eine Lücke in der mathematischen Literatur auszufüllen. Ein Lehrbuch der „Invariantentheorie“, das wenigstens die Hauptgegenstände dieses ausgebreiteten Teiles der Mathematik behandelt, existiert meines Wissens bisher nicht. Deshalb will ich auch gleich von vornherein betonen, daß mir die Auswahl der Gegenstände nicht leicht war und daß vieles wegbleiben mußte, was in eine umfassende Darstellung gehören würde. Erst dann, wenn man sich nur auf die algebraische Theorie beschränkt, kommt eine gewisse Abgrenzung zustande. Recht ausgebreitete Theorien, wie z. B. die der topologischen und arithmetischen Invarianten, wurden unterdrückt; ebenso die Theorie der Invarianten bei endlichen
VORWORT

Neue Bücher entspringen in der Regel einem Zeitbedürfniss. Ich glaube mit diesem Buche eine Lücke in der mathematischen Literatur auszufüllen. Ein Lehrbuch der „Invariantentheorie“, das wenigstens die Hauptgegenstände dieses ausgebreiteten Teiles der Mathematik behandelt, existiert meines Wissens bisher nicht. Deshalb will ich auch gleich von vornherein betonen, daß mir die Auswahl der Gegenstände nicht leicht war und daß vieles wegbleiben mußte, was in eine umfassende Darstellung gehören würde. Erst dann, wenn man sich nur auf die algebraische Theorie beschränkt, kommt eine gewisse Abgrenzung zustande. Recht ausgebreitete Theorien, wie z. B. die der topologischen und arithmetischen Invarianten, wurden unterdrückt; ebenso die Theorie der Invarianten bei endlichen

NIEDER MIT DEN FRANZÖSEN
Teleparallelism

- Einstein’s work that annoyed Cartan was an attempt to characterise spacetime by a curvature-free linear connection and a metric tensor field
- Torsion can also be used for non-holonomic mechanics
- And shear forces in crystals
Using for Data

- Can add a frame to each data point
- Then torsion is the mutual rotation between frames
- Make parallel transport path independent
- Then frame field is a scalar f(x)
- But by computing Lie bracket we see that parallelogram with equal and opposite sides doesn’t close
- And integral curves of the zweibein cannot be used as a coordinate grid
Moving Frame

- This is a version of Cartan’s moving frame
  - Derive a set of linearly independent vector fields
  - Use to construct derivative

- But data manifold is unknown
  - Can’t derive vector fields
Localised PCA

- Perform PCA on points within a ball
- Use first component as an important direction
- Repeat and use to define how frames rotate
- Analytic & has a well-defined limit as $N \rightarrow \infty$
Geometrically:

- Autoparallels define fibres
  - Single direction along vector field
- Choose a base space and project data point onto base space by following autoparallel
- Estimate autoparallel locally using PCA
- Algorithm:
  - Parameterise manifold by first autoparallel
  - Project that out
  - Iterate for new autoparallel
  - Until data is isotropic (noise) or other stopping criterion
Conclusions

● Replace Riemannian connection with Weitzenböck connection

● Extremely simple algorithm

● Performs well on standard datasets, but computationally expensive

● Extension to shape data (where direction, but not position is important) seems to work