# Formatting instructions for NeurIPS 2018

**Alec J. Hoyland**
Center for Systems Neuroscience
Boston University
Boston, MA 02215
ahoyland@bu.edu

## Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

## 1 Introduction

Symbolic processing is an important feature of general intelligence, crucial for understanding underlying rules and latent features, to develop models of the world that can be used for prediction. In animals, simple prediction tasks such as object tracking and navigation, are necessary for survival. These tasks fall under the umbrella of symbolic processing. One of the most straightfoward symbolic processing tasks which can be abstracted to a computational model is trajectory prediction.

This task involves observing a system with a state that evolves in time for a few epochs, before making predictions of the system's future state based on those observations. Real-world applications involve pedestrian and vehicle tracking, and robot navigation and agility. In essence, it is a type of supervised learning problem, in which the prediction model is being updated at each time step.

The goal of this paper is to put forward several theories on how this can be practically done, with an emphasis on discovering fundamental mathematical relationships which could be performed by ensembles of neurons in the brain.

### 1.1 What is a trajectory?

For the purpose of this section, a *trajectory* is an ordered set of states which are evolved according to one or more rules. The purpose of symbolic trajectory prediction is, if given a few states of the trajectory, to intuit the latent rules and predict future states. In a real-world scenario, the agent predicts simultaneously to the evolution of the trajectory in time, meaning that more data are available to the agent each time step.

Mathematically, a trajectory is formally represented as a sequence of values $\left[f^k(x)\right]_{k\in\mathbb{N}}$, calcuilated by the iterated application of a mapping $f$ to an element $x$ of its source. It is analogous to the time-ordered set of states in a dynamical system (*e.g.* a Poincaré map) (CITE) such as the function of position produced by integrating the equations of motion in Newtonian mechanics.

### 1.2 What is a feature?

A feature is any important defining characteristic of a snapshot of a trajectory. For example, if the trajectory represents the motion of a ball in flight across a 2-dimensional coordinate grid, the crucial features are the coordinates $(x, y)$ parametrixed by the time-coordinate $t$. If the color of the ball mattered (*i.e.* isn't time-invariant), then it would also be a feature. For the purpose of this paper,

Preprint. Work in progress.

the agent is assumed to be able to discern which features are important and which are not. From a biological perspective, we will take for granted the miraculous specificity of the visual system in mammals. In artificial intelligence, we will assume a sufficiently-advanced computer vision algorithm and peripherals. If, for example, three polygonal shapes were drawn on a blackboard and the agent were asked to determine the pattern, it is reasonable to neglect the minute deviations in the lines caused by the hand of an unskilled artist. Similarly, imperfections in printed symbols could be ignored as well. The best heuristic for determining what is a feature and what is not relies on comparing adjacent states of the trajectory. Major differences are feature; minor differences can be neglected.

## 2 Representing the trajectory through vector transformations

One way to imagine a trajectory is to consider a series of points in the plane formed by the time-coordinate $t$ and a function $f(t)$ which produces a vector output $\vec{x}$, where each element in $\vec{x}$ is the numerical value of a feature at time $t$. In this situation, the trajectory is a vector parametrized by the time-coordinate. To compare adjacent states, the first time-derivative can be taken. Since time is discrete, the derivative operator is best represented by the finite difference.

A finite difference is a mathematical expression of the form $f(x+b) - f(x-a)$. If the finite difference is divided by $b-a$, the different quotient is defined. The first forward finite difference is defined by

$$\mathrm{D}x_n = x_n - x_{n-1} \tag{1}$$

for variable $x$ at discrete indices $n$ and $n-1$. Higher-order finite differences can also be defined, such as the second-order forward finite difference:

$$\mathrm{D}^2x_n = x_n - 2x_{n-1} + x_{n-2} \tag{2}$$

These finite differences are first-order correct in accuracy with uniform grid-spacing $t_n$. More accurate finite differences can be computed using more points. At minimum, approximating a numerical derivative of order $k$ with the fewest points requires $k+1$ points.

If two states of the trajectory $x_n$ and $x_{n-1}$ are known, the next point can be approximated as:

$$x_{n+1} \approx \frac{1}{\Delta}\mathrm{D}x_n + x_n = \frac{x_n - x_{n-1}}{\Delta} + x_n \tag{3}$$

where $\Delta = t_n - t_{n-1}$ is the difference in the time-coordinate between adjacent indices. If higher-order finite differences are used, the estimate will be more accurate, up to the derivative order equal to the order of the polynomial representation of the function. This requires knowing more past states.

Since any well-behaved function can be approximated by its (truncated) Taylor series, the discrete Taylor series generalization, given by Einar Hille, can approximate a well-behaved function using finite differences (CITE). For $t > 0$

$$f(a+t) = \lim_{\Delta \to 0^+} \sum_{n=0}^{\infty} \frac{t^n}{n!\Delta^n}\mathrm{D}^n f(a) \tag{4}$$

This equation describes how if a function is known to be evaluatable at $a$: $f(a)$, and finite differences can be taken, then the function can be approximated at any future point $a+t$.

The vector transformation formulation is satisfactory for predicting trajectories based on linear rules. It is mathematically equivalent to least-squares fitting, and therefore can be represented in matrix form (CITE). This results in a linear transformation called the predictor:

$$\mathrm{P}(a) = \left[\frac{1}{n!}\mathrm{D}^n f(a)\right]_{n\in\mathbb{N}} \tag{5}$$

The dot product of the predictor can be taken with $\left[\frac{t^n}{\Delta^n}\right]_{n\in\mathbb{N}}$ to find the predicted trajectory $f(a+t)$ for $t > 0$.

### 2.1 Example: one-dimensional trajectory

Here we consider a trajectory produced by a linear combination in a polynomial basis. The goal is to predict the one-dimensional trajectory

$$f(t) = -40 + 100t - 10t^2 \tag{6}$$

For $\Delta = 0.01$, the first and second finite differences are computed at $f(2\Delta)$. By the Hille series (4), we can compute

$$f(2\Delta + t) = f(0) + \frac{t}{\Delta}\mathrm{D}f(2\Delta) + \frac{t^2}{2\Delta^2}\mathrm{D}f(2\Delta) \tag{7}$$

which is equivalent to the Taylor series

$$f(t) = x_0 + v_0 + \frac{1}{2}a_0 t^2 \tag{8}$$

where $x_0$, $v_0$, and $a_0$ are parameters fit by the model. Any extra parameters, for example, for a 3rd order polynomial fit are zero, and thus omitted.

ADD CODE & FIGURE

### 2.2 Example: multi-dimensional trajectory

Here we consider a multi-dimensional example. The latent rule is at time $t_n \in \mathbb{N}$, we draw a square, where the top-left vertex is at the origin, with a side length of $s = t_n + 1$.

Treating the observation mechanism (*e.g.* the visual cortex) as an oracle, we identify eight features: the four coordinate-pairs which note the vertices of the square. One vertex (top left) never changes position – it is fixed at the origin. The top right vertex proceeds according to the rule $f_{tr}(t_n) = (0, t_n + 1)$. The bottom left vertex proceeds according to $f_{bl}(t_n) = (-1 - t, 0)$. Corresponding, the bottom right vertex evolves according to $f_{br}(t_n) = (-1 - t, -1 - t)$.

Thus, each coordinate pair evolves either linearly or doesn't change (is constant). Finite difference approximations can be taken for each coordinate independently, and the rule can be determined after two steps.

ADD CODE & FIGURE

## 3 Representing the trajectory through function transformations

The idea of mapping vectors across a time-coordinate can be expanded to a larger space of problems. In this generalization, the latent rule involves defining a series of functions ordered in time: $[f_n(\mathcal{S})]_{n\in\mathbb{N}}$.

Each function acts on a surface $\mathcal{S}$, the specifics of which should not matter for the general formalism.

In the field of functional calculus, arithmetic can be defined on functions. The "sum" of two functions is the convolution $*$, and the "difference" of two functions is the cross-correlation $\star$.

If the rule can be modeled as a first-order transformation in function space, the evolution rule is:

$$f_{n+1}(\cdot) = (f_n(\cdot) \star f_{n-1}(\cdot)) * f_n(\cdot) \tag{9}$$

This formalism is an extension of the vector representation formalism (2), since convolution and cross-correlation are defined for all well-behaved functions.

# 4 Submission of papers to NeurIPS 2018

NeurIPS requires electronic submissions. The electronic submission site is

`https://cmt.research.microsoft.com/NeurIPS2018/`

Please read the instructions below carefully and follow them faithfully.

## 4.1 Style

Papers to be submitted to NeurIPS 2018 must be prepared according to the instructions presented here. Papers may only be up to eight pages long, including figures. Additional pages *containing only acknowledgments and/or cited references* are allowed. Papers that exceed eight pages of content (ignoring references) will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2018 are the same as since 2007, which allow for $\sim$15% more words in the paper compared to earlier years.

Authors are required to use the NeurIPS LaTeX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

## 4.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the World Wide Web at

`http://www.neurips.cc/`

The file `neurips_2018.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

The only supported style file for NeurIPS 2018 is `neurips_2018.sty`, rewritten for LaTeX $2_{\varepsilon}$. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

The LaTeX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

**New preprint option for 2018** If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text "Preprint. Work in progress." in the footer. This version may be distributed as you see fit. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS.

At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `neurips_2018.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in Sections 5, 6, and 7 below.

# 5 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section 7 regarding figures, tables, acknowledgments, and references.

# 6 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

## 6.1 Headings: second level

Second-level headings should be in 10-point type.

### 6.1.1 Headings: third level

Third-level headings should be in 10-point type.

**Paragraphs** There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

# 7 Citations, figures, tables, references

These instructions apply to everyone.

## 7.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2018` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2018}
```

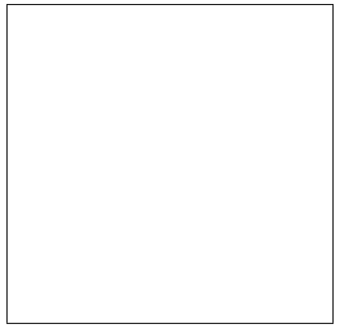

Figure 1: Sample figure caption.

As submission is double blind, refer to your own published work in the third person. That is, use “In the previous work of Jones et al. [4],” not “In our previous work [4].” If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form “A. Anonymous.”

### 7.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

### 7.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

### 7.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table 1.

## 8 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

[1] Sample of the first footnote.
[2] As in this example.

Table 1: Sample table title

| Part | | |
|---|---|---|
| Name | Description | Size ($\mu$m) |
| Dendrite | Input terminal | $\sim$100 |
| Axon | Output terminal | $\sim$10 |
| Soma | Cell body | up to $10^6$ |

# 9 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

  ```
  \newcommand{\RR}{I\!\!R} %real numbers
  \newcommand{\Nat}{I\!\!N} %natural numbers
  \newcommand{\CC}{I\!\!\!\!C} %complex numbers
  ```

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

## 9.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.


### Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

## References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**


[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.