Neuronal Connectivity based on Cell Morphologies

Project Course in Scientific Computing

Shriya Bhatija BHATIJA.SHRIYA@GMAIL.COM
Graduate student at the Technical University of Munich, Germany
Exchange student at the Royal Institute of Technology, Sweden

Abstract. In this project, connectivity structures of neuronal networks that arise from varying cell morphologies were studied. For this intent, suitable representations of neuronal shapes were constructed via advanced multivariate probability distributions such as the skew-elliptical and Sinh-Arcsinh distributions. Network connectivity was represented as matrices which were able to reveal how the connectivity is altered with varied neuronal morphologies.

1. Introduction

Understanding the functionality and mechanisms of the human brain has been an important objective in medicine that has inspired various scientific areas to investigate towards. The brain may be understood as a neuronal network, a circuit composed of biological neurons that are able form connections amongst each other. The formation of such connections is dependent on various factors such as proximity, cell morphologies, spatial density and many more. Since the dynamics and information processing properties of the brain are heavily influenced by neuronal connectivity, it is vital to understand how a networks connectivity structure arises. Numerous efforts have already been made in this research area, see for example Jiang et al. (2015), Kalisman et al. (2003), Peng et al. (2021), Udvary et al. (2022), van Pelt and van Ooyen (2013).

The aim of this project is to estimate connectivity based on distinct neuronal morphologies. Since the connectivity of neuronal networks determines their computational repertoire, the results of this project may indicate how certain cell morphologies can alternate a network's functionality. Here, the emphasis however, lies on modelling those neuronal shapes in a probabilistic sense enabling us to consider distances to other neurons. More specifically, a connection should be more likely within close proximity of a given neuron depending on its shape. This will give a new (probabilistic) representation of cells with different morphologies. For this intent, the multivariate normal distribution will be expanded to a broader family of distributions that is able to capture skewness and tailweights, see section 4. For visualisation purposes, it was set to only conduct bivariate modelling as adjusting the respective model parameters may yield appropriate 3D representations. Next, it will be explained in more detail how cells can exhibit different shapes and what this entails for their ability to form appositions.

2. Neuronal Morphologies

Neurons are comprised of three main components: dendrites, an axon and a cell body (soma). For a simplified visualisation, see the the figure below:

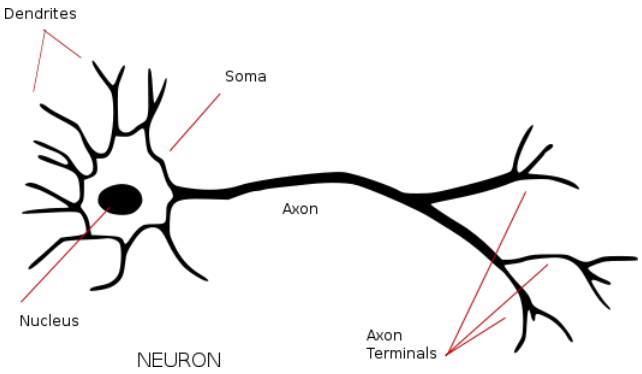


Figure 1: Sketch of a neuron with its three main components: dendrites, an axon and a soma. Source: ClipartMax

Dendrites are branch-like structures that receive synaptic input from surrounding neurons and forward them to the respective soma for further processing. The axon extends from the cell body and generates electrical signals that are transmitted to connecting neurons. An apposition between two neurons arises when a synapse is formed between an axon and a dendrite enabling a transfer of information from one neuron to another. For a more detailed explanation on neurons and the formation of synapses, Sevush (2016) is recommended.

As previously stated, the occurence of synapses depends on various factors, two of which are proximity and neuronal morphology. The neuronal morphology is determined by the axonal and dendritic arbors which appear in various shapes and sizes and are indeed the most important determinants for the connectivity between neurons as explained in Jiang et al. (2015).

The shapes and sizes of the axonal and dendritic arbors may differ significantly depending on the neurons functionality, purpose and/or brain region. Some more common morphologies are displayed in Figure 2 and Figure 3. In order to investigate, to what extend a networks connectivity is influenced by distinct neuronal morphologies, it is essential to represent these shapes in an appropriate manner within a suitable model. The probabilistic representation of dendrites and axons will be the focal point of this report. In the following section, the model choice and project outline will be discussed.

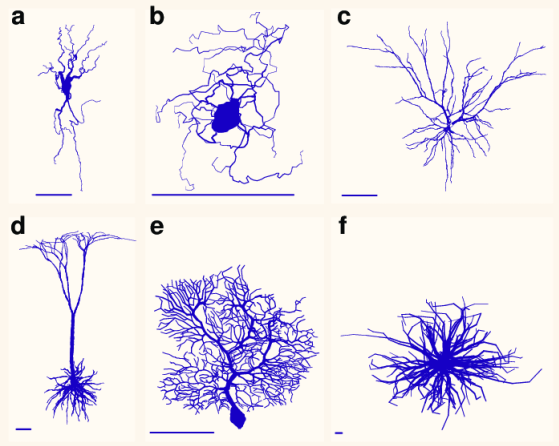


Figure 2: Several morphologies of dendritic arbors. Source: Koch and Segev (2000)

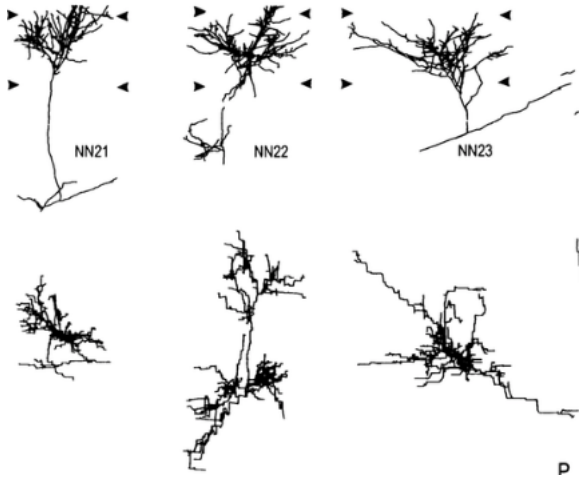


Figure 3: Several morphologies of axonal arbors. Croppoed image from source: Antonini and Stryker (1998)

3. Project Pipeline

The main focus of this project lies in the exploration of new multivariate distributions for the stochastic modelling of different neuronal morphologies. For further computations and connectivity estimation, it was set to represent the various neuronal shapes as density matrices since those are compact and yet flexible mean of storing 2D spacial information. This will be carried out in the following way:
Given new multivariate probability distributions, shapes will be generated as bivariate random variables in a realm of size $[0, 100]^2 \subset \mathbb{R}^2$. As an initial consideration, the space needs to be discretized by dividing it into 100^2 voxels, each of size 1×1. After sampling a collection of points

$$\mathcal{T} = \{(x_k, y_k)\}_{k=1}^{n}$$

from the newly defined random variables, it is now possible to assign each point to its corresponding voxel. At this point, density matrices can easily be generated from the point clouds. Such a matrix A shall preserve the spacial information about the point density within each voxel, hence it is obtained via the following calculation:

$$A_{ij} = v_{ij}/n,$$

where v_{ij} is the number of samples in the voxel at row i and column j. This gives a new representation

of neuronal morphologies as density matrices which was desired. Since the dendritic and axonal arbors extend from the cell body, it is important to consider the location of the somata within the matrices. They shall be placed at the center of the density matrices.

In order to estimate network connectivity with varying neuronal shapes, a model network X needs to be initialized where one may insert the density matrices and calculate the prospect of appositions between any two pairs of neurons. For this purpose, 100 instances were randomly sampled in a space of size $([1, 200] \cap \mathbb{N})^2$. Each of these samples shall represent the cell body of a neuron, hence a 2D space of 100 neurons is obtained. The network X will be stored as a matrix in $\{0, 1\}^{200,200}$ where

$$X_{ij} = \begin{cases} 1 & \text{if } (i, j) \text{ was sampled,} \\ 0 & \text{else.} \end{cases}$$

For further computations, X needs to be enlarged via symmetric zero-padding such that the dimensions 400×400 are obtained. This concludes the construction of the model. It is of course based on an extremely simple network architecture, however it suffices for the purpose of connectivity analysis based on stochastic modelling of varying morphologies.

Having designed both the network as well as the neuronal shapes, it is now possible to generate the so called *connectivity matrices* which are a mean of storing the appositions between any two neurons in a given network. Here, a connectivity matrix $C \in \{0, 1\}^{100,100}$ is given by

$$C_{ij} = \begin{cases} 1 & \text{if the dendrite of neuron } i \text{ connects to} \\ & \text{the axon of neuron } j, \\ 0 & \text{else.} \end{cases}$$

In neuronal networks information flows from axon to dendrite, hence a connection between two neurons does not yield the ability to freely interchange signals. This is a distinct characteristic of neuronal networks and hence it is vital to capture such asymmetry within C.

Since the prospect of connectivity between two neurons is heavily influenced by their morphologies, the structure of the connectivity matrices is intricately tied to the dendritic and axonal shapes. In a condensed brain region it is common for all surrounding neurons to have consistent morphology within the set of dendrites and axons, respectively. For this reason, it is reasonable to fix consistent dendrite and axon morphologies for all neurons in the network X prior to estimating connectivity. Particularly, each pair of shapes will give rise to a new connectivity matrix. To what degree these matrices differ, is yet to be studied and will be part of the later analysis.

The set of neurons will be denoted

$$\mathcal{N} = \{N_i = (N_i^A, N_i^D)\}_{i=1}^{100}$$

where N_i^A (N_i^D) representes the axonal (dendritic) morphology of neuron i. The likelihood of connectivity between any two neurons will be measured via the overlap of dendritic arbors of one with the axonal tree of the other. For this purpose, pairs of neurons in X will iteratively be fixed and equipped with prechosen axonal and dendritic morphologies. The above mentioned overlap can then be measured via convolution. Note that in general the convolution between two matrices $A, B \in \mathbb{R}^{n,m}$ is defined as the sum over the elementwise product of all matrix entries:

$$\text{conv}(A, B) := \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij}.$$

Convolution provides a score for the chance of appositions between two neurons. A higher value indicates a more extensive overlap and therefore a higher plausability of connectivity. The convolution value needs to be calculated twice for every pair of neuron due to the asymmetry of C, i.e. one may have $\text{conv}(N_i^A, N_j^D) \neq \text{conv}(N_i^D, N_j^A)$ for $i \neq j$.

After carrying out all necessary computations, the convolutions need to be aligned with the binary set of events consisting of the presence and absence of connectivity. For this purpose, a threshold value is introduced as a new hyperparameter. More specifically, a previously fixed threshold value *thr* will determine connectivity as follows: if $\text{conv}(N_i^A, N_j^D)$ surpasses *thr*, it will interpreted as the overlap between dendrite of neuron i and the axon of neuron j being sufficiently high for the formation of appositions. With this reasoning, the connectivity matrix C is obtained as follows:

$$C_{ij} = \begin{cases} 1 & \text{if } \text{conv}(N_i^D, N_j^A) \geq thr, \\ 0 & \text{else.} \end{cases}$$

The threshold value determines the extend of the networks connectivity, i.e. the lower it is set the more

connections are attained. Therefore, the connectivity matrices will vary not only with changing pairs of dendritic and axonal morphologies but also with different threshold values. Since the incentive is to analyse changes in connectivity that arise exclusively from distinct morphlogies, the threshold needs to be fixed. The value 20^{-6} was chosen due to the manageable number of appositions it yields.

After having generated connectivity matrices for different combinations of dendritic and axonal shapes, they shall be analysed in order to investigate upon their differences and similarities. This will be done utilizing different metrics, ideally yielding a consensus about how neuronal morphologies can alter a networks connectivity structure.

All computations for this project were carried out using Python code, relying heavily on the NumPy, SciPy and TensorFlow libraries.

4. Mathematical Framework

In this section, the mathematical foundations needed for generating the different neuronal morphologies will be established. As already outlined in section 3, shapes are going be presented as bivariate random variables which will subsequently undergo sampling in order to generate density matrices. For the purpose of constructing random variables that are able to capture complex geometries, appropriate bivariate distributions are needed. Therefore, several multivariate probability distributions will now be reviewed, roughly following Babić et al. (2019) (unless otherwise stated) for explanations as well as visualizations. Starting with the multivariate normal distribution, the scope will extent to the so called *skew-elliptical* and *SAS* distributions which allow for significantly more flexible modelling. It is assumed that the reader has a basic understanding of notions related to probability theory in high dimensions.

For completeness, all distributions will be defined for the d-dimensional case and fixing $d = 2$ reduces to the bivariate setting. In the upcoming subsections, $=_d$ shall express equality in distribution between random quantities.

4.1. Normal Distribution

A well known and in practice often utilized probability distribution is the so called *multivariate normal distribution*. Its density clusters around a mean value and the variance spreads with radial symmetry around the mean. Precisely, it is defined as follows, also see Brémaud (2020):

A d-dimensional random vector $X = (X_1, ..., X_d)^{\mathrm{T}}$ is called a *standard normal random vector* iff all of its components X_i are independent and identically distributed following $X_i \sim \mathcal{N}(0, 1)$.

A d-dimensional random vector $X = (X_1, ..., X_d)^{\mathrm{T}}$ has *multivariate normal distribution* iff there exist $\mu \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times k}$ and a k-dimensional standard normal random vector Z such that

$$X =_d AZ + \mu. \tag{1}$$

It can be shown that μ is equal to the mean vector $\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d))^{\mathrm{T}}$ which can also be interpreted as a location parameter. Furthermore, the covariance matrix $\Sigma \in \mathbb{R}^{d,d}$ given by $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$ is attained via $\Sigma = AA^{\mathrm{T}}$. It captures the relation between the components of X and may be considered as a scatter parameter. Important properties of the covariance matrix include symmetry and positive semi-definiteness which result in the density being radially symmetric around the mean. The multivariate normal distribution is therefore solely specified by the parameters μ and Σ. Commonly, the notation $X \sim \mathcal{N}(\mu, \Sigma)$ is used to denote that X is multivariate normal with mean μ and covariance matrix Σ.

While it is powerful for many applications, the multivariate normal distribution is sometimes not flexible enough for modelling purposes which is partially due to the fact that it does not account for heavy tails and asymmetry. The next sections address and circumvent this limitation by introducing more adaptable distributions.

4.2. Elliptically Symmetric Distribution

The family of elliptically symmetric distributions, originally introduced by Kelker (1970), forms a natural extension of the normal distribution since it allows for both lighter-than-normal and heavier-than-normal tails while adhering to elliptical geometry. Formally, it is defined as follows:

A d-dimensional random vector $X = (X_1, ..., X_d)^{\mathrm{T}}$ is said to be *elliptically distributed* iff it has the stochastic representation

$$X =_d \mathcal{R}\Lambda U + \mu, \tag{2}$$

where $\mu \in \mathbb{R}^d$, $\Lambda \in \mathbb{R}^{d\times k}$ has maximal rank $k \leq d$, U is a k-dimensional random vector uniformly distributed on the unit hypersphere, and $\mathcal{R}$ is a non-negative random variable independent of U. The scaling matrix Λ accounts for the ellipticity while $\mathcal{R}$ regulates tail thickness. The covariance matrix Σ is obtained through $\Sigma = \Lambda\Lambda^{\mathrm{T}}$ and μ is again equal to the mean vector.

If the distribution of $\mathcal{R}$ is absolutely continuous, then there exists a corresponding density function. In this case, the density function of X exists as well and is given by

$$x \mapsto c_{d,g}|\Sigma|^{-1/2} g((x-\mu)^T\Sigma^{-1}(x-\mu)), \tag{3}$$

where $g : \mathbb{R}_0^+ \to \mathbb{R}$ is the radial function related to the distribution of $\mathcal{R}$ and $c_{d,g}$ is a normalizing constant. The notation $X \sim E(\mu, \Sigma, g)$ is used to express that X is multivariate elliptical with mean μ, covariance matrix Σ and radial density g.

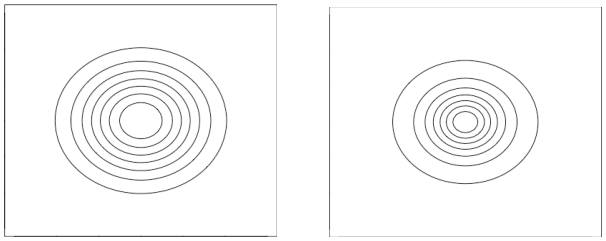
Example 1: The elliptically symmetric distribution generalizes the normal distribution by introducing a single tail weight parameter that is captured in $\mathcal{R}$. Hence, the normal distribution falls within the family of elliptical distributions and moreover has density

$$x \mapsto \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\Big(-\frac{1}{2}(x-\mu^T)\Sigma^{-1}(x-\mu)\Big). \tag{4}$$

Example 2: Another well known elliptically symmetric distribution is the so called multivariate *t-distribution*. Its density is given by

$$x \mapsto \frac{\Gamma[(\nu+d)/2]}{\Gamma(v/2)(\nu\pi)^{d/2}|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]^{-(\nu+d)/2}, \tag{5}$$

where $\nu > 0$ is the parameter (also called *degree of freedom*) that regulates tail thickness. The lower the value of ν, the heavier the tails.



(a) standard normal (b) t-distribution

Figure 4: Contour plots of bivariate elliptically symmetric distributions with $\mu = 0$, $\Sigma = I_2$ and for plot (b) additionally $\nu = 1$.

The family of elliptically symmetric distributions consists of a vast variety of many more distributions, including the so called *symmetric multivariate stable distribution*, *multivariate power-exponential distribution* and the *symmetric multivariate Laplace distribution*. It indicates that there is a plethora of distributions that allow altering tail thickness while being elliptically centered around the mean.

Controlling tail thickness is already a significant improvement in terms of modelling flexibility as compared to the multivariate normal case. However, the family of elliptically symmetric distributions is exposed to two prime limitations. Firstly, they are necessarily (elliptically) symmetric by construction and hence unable to model asymmetric/skewed data. Secondly, the tail weight can only be regulated with a single parameter and therefore these distributions cannot model varying tail weight in distinct directions. Next, the so called *skew-elliptical* distribution will be introduced to circumvent these limitations.

4.3. Skew-Elliptical Distribution

Skew-elliptical distributions can be attained by modulating the symmetry in elliptical distributions via multiplication with a skewing function. Precisely, the d-dimensional skew-elliptical distribution is defined as follows:

Let $\Pi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ be the skewing function such that is satisfies $\Pi(-z,\delta) + \Pi(z,\delta) = 1$ for $z, \delta \in \mathbb{R}^d$ and $\Pi(z,0) = 1/2$ for all $z \in \mathbb{R}^d$. Further, let Y be an elliptically symmetric d-dimensional random variable with density (3). U shall denote a random variable which is uniformly distributed on (0,1) and independent of Y. Now, for $\delta \in \mathbb{R}^d$ the

random vector

$$X =_d \begin{cases} Y & \text{if } U \leq \Pi(\Sigma^{-1/2}(Y-\mu),\delta), \\ -Y & \text{if } U > \Pi(\Sigma^{-1/2}(Y-\mu),\delta) \end{cases}$$

follows a skew-elliptical distribution and its density is given by

$$\begin{aligned} x \mapsto \; & 2c_{d,g}|\Sigma|^{-1/2} g((x-\mu)^T\Sigma^{-1}(x-\mu)) \\ & \Pi(\Sigma^{-1/2}(x-\mu),\delta). \end{aligned} \tag{6}$$

Here, the quantities $c_{d,g}, g, \mu, \Sigma$ are exactly the ones from subsection 4.2 and arise from the density function of Y. Partularly, μ again represents the mean vector whereas Σ denotes the covariance matrix. The parameter δ controls the skewness such that setting $\delta \neq 0$ skews the data in the direction of δ. Choosing $\delta = 0$ on the other hand, means setting no skewness and an elliptically symmetric distribution arises. However, such a distribution, that is attained via symmetry modulation of an elliptically symmetric distribution, is unable to model varying tail thickness due to the absence of a tail weight parameter. Nevertheless, the class of skew-elliptical distributions includes probability distributions that indeed allow the modelling of varying tail thickness along with skewness. The so called multivariate *skew-t* is a popular example of such an instance. It is, as the name suggests, a skewed version of the multivariate t-distribution and constructed as follows:

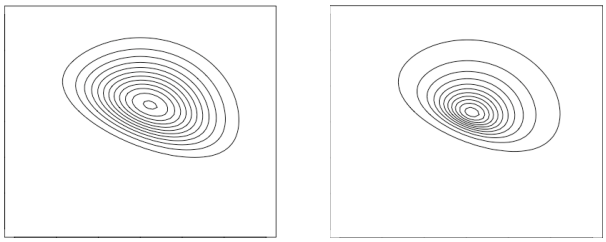
Firstly, the so called multivariate *skew-normal* is needed as one of the building blocks. It is obtained from (6) by choosing g as the radial density of a multivariate normal and setting $\Pi(z,\delta) = \Phi(z^T\delta)$ where Φ is the cumulative distribution function (cdf) of a univariate standard normal. Then the *skew-t* distribution is obtained via the relation

$$X =_d \mu + V^{-1/2}Y, \tag{7}$$

where Y is multivariate skew-normal and $V \sim \chi^2_\nu/\nu$ is a chi-square random variable with $\nu > 0$ and independent of Y. The density of X is given by

$$\begin{aligned} x \mapsto \; & 2t_d(\Sigma^{-1/2}(x-\mu);\nu) \\ & T_1\left(\delta^T\Sigma^{-1/2}(x-\mu)\left(\frac{\nu+d}{\mathcal{Q}_x+\nu}\right)^{1/2}; \nu+d\right) \end{aligned} \tag{8}$$

where $\mathcal{Q}_x = (x-\mu)^T\Sigma^{-1}(x-\mu)$, $t_d(x;\nu)$ the density of a d-dimensional t variate with v degrees of freedom and $T_1(x,\nu+d)$ is the skewing function which also represents the cdf of the scalar t-distribution with $v+d$ degrees of freedom. Here, tail thickness can be regulated with ν.



(a) skew-normal (b) skew-t

Figure 5: Contour plots of bivariate skew-elliptical distributions with $\mu = 0$, $\Sigma = I_2$, $\delta = (1,2)^T$ and for plot (b) additionally $\nu = 1$.

Hence, the family of skew-elliptical distributions enables the regulation of tail thickness as well as asymmetry which allows for significantly more flexibility when modelling complex geometries. As previously outlined, such distributions are obtained by utilizing the multivariate normal as a base and intricately introducing more and more parameters for a more general family of distributions.

4.4. SAS - Transformation

Another method for obtaining complex distributions from relatively simple ones goes via the so called *transformation approach*. The idea is as follows:

Given a "basic" random vector Y, e.g. multivariate normal in d-dimensions, it is transformed via a diffeomorphism $H : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$X =_d H^{-1}(Y). \tag{9}$$

If f is the density of Y, then the density function of X is given by

$$x \mapsto f(H(x))|DH(x)|, \tag{10}$$

where $DH(x)$ shall denote the determinant of the Hessian matrix corresponding to the transformation H.

The transformation approach is a powerful tool for generating complex geometries in high dimensions since there are no restrictions in regards to the choice of H. One of the more common transformations is obtained via the sinh-arcsinh function which

ultimately results in the so called *Sinh-Arcsinh (SAS)* distribution, first introduced by Jones and Pewsey (2009). It includes parameters for regulating tail thickness as well as asymmmetry. Formally, it is defined as follows:

Starting with the univariate case, a standard normal random variable $Z \sim \mathcal{N}(0,1)$ is transformed by the inverse of the sinh-arcsinh function

$$z \mapsto S_{\delta,\nu}(z) := \sinh(\nu \sinh^{-1}(z) - \delta), \tag{11}$$

where $\delta \in \mathbb{R}$ is a skewness parameter and $\nu > 0$ is a tail weight parameter. After introducing a location parameter μ and a scale parameter $\sigma > 0$, the resulting *SAS-normal* random variable is defined as

$$X =_d \mu + \sigma S_{\delta,\nu}(Z) \tag{12}$$

and has density

$$x \mapsto \frac{\nu/\sigma}{\sqrt{2\pi(1+(x-\mu)^2/\sigma^2)}} \left(1 + S^2_{\delta,\nu}\left(\frac{x-\mu}{\sigma}\right)\right)^{1/2} \exp\left(-\frac{S^2_{\delta,\nu}\left(\frac{x-\mu}{\sigma}\right)}{2}\right). \tag{13}$$

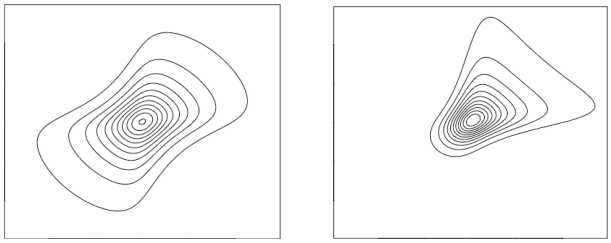
The SAS distribution has numerous appealing properties, one of which being parameter interpretability: Choosing $\nu < 1$ yields heavy-tailed distributions while $\nu > 1$ results in light-tailed distributions. Furthermore, the more positive δ is set the more right-skewed X will be, while a negative value of δ results in a left-skewed distribution. Note that for $\nu = 1$ and $\delta = 0$ a normal distribution $X \sim \mathcal{N}(\mu, \sigma)$ is obtained. Having constructed the univariate SAS distribution, it is now possible to extend it to the multivariate case as was done by Jones and Pewsey (2009):
Let $Y \sim \mathcal{N}(0, \Sigma)$ be a d-dimensional random vector, then the random vector $X = (X_1, \ldots, X_d)$ defined by

$$X_i =_d S_{\delta_i,\nu_i}(Y_i) \tag{14}$$

for some $\delta = (\delta_1, ..., \delta_d)$ and $\nu = (\nu_1, ..., \nu_d)$, represents the multivariate case of the SAS distribution. Its density is given by

$$x \mapsto \{(2\pi)^d |\Sigma|\}^{-1/2} \exp\{-S_{\delta,\nu}(x)^T \Sigma^{-1} S_{\delta,\nu}/2\} \prod_{i=1}^{d} \left\{(1+x_i^2)^{-1/2} \nu_i \left(1 + S^2_{\delta_i,\nu_i}(x_i)\right)^{1/2}\right\}. \tag{15}$$

Similarly to the univariate case, the parameters ν and δ influence the tail thickness and asymmetry, respectively. However, in this case these parameters are d-dimensional vectors instead of scalar values which makes it possible to control these attributes distinctly in different directions. The figure below displays the contour plots of bivariate SAS distributions, including a symmetric as well as skewed example.



(a) symmetric (b) asymmetric

Figure 6: Contour plots of SAS distributions.

In summary, this chapter discussed how the multivariate normal distribution can be altered to incorporate further attributes such as tail thickness and asymmetry for the purpose of generating more complex geometries. The family of skew-elliptical distributions was introduced along with the SAS distribution as an example for the transformation approach. Babić et al. (2019) present many more examples, including mixture models as well as various other distributions that arise from the transformation approach. The efforts of constructing such complex probability distributions trace back to several different origins which shows the relevance of this topic and that flexible modelling is desired for many purposes. For the intent of this project, these probability distributions will be utilized to construct different shapes mimicking some neuronal morphologies. The next chapter presents some bivariate geometries that can be generated from the above discussed probability distributions.

5. Probabilistic Modelling

In this section, new bivariate shapes will be constructed by utilizing the probability distributions introduced in the last chapter. The resulting geometries shall represent axonal and dendritic morphologies and are therefore inspired by the illustrations presented in Figure 2 and Figure 3. Recall from section 3 that these shapes are going to be represented

in a space of size $[0, 100]^2 \subset \mathbb{R}^2$ with the cell body being fixed at position (50,50). Considering these model choices, the following shapes have been created:

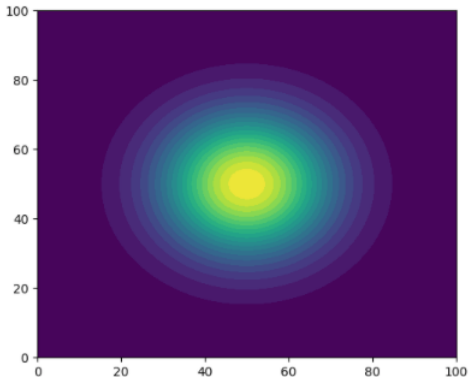


Figure 7: Shape S_1. Bivariate normal distribution.

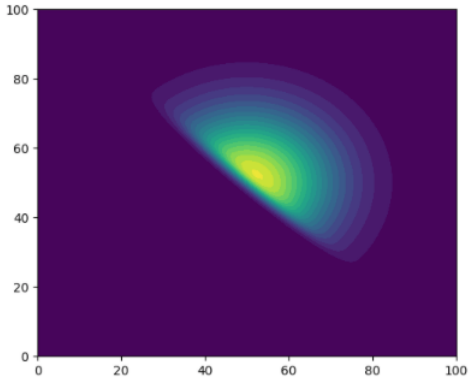


Figure 8: Shape S_2. Skew-normal distribution.

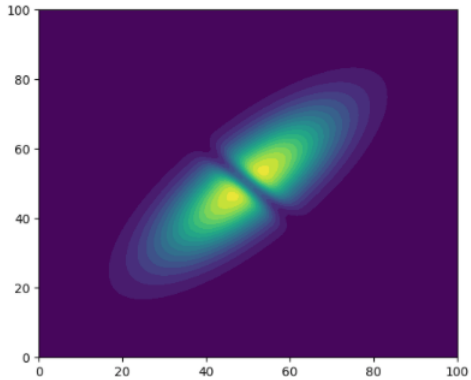


Figure 9: Shape S_3. Sum of two skew-normal distributions.

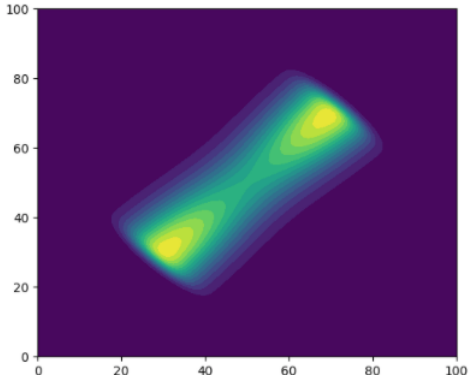


Figure 10: Shape S_4. Sum of two skew-normal distributions.

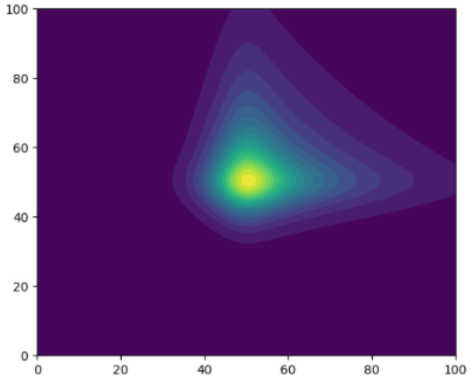


Figure 11: Shape S_5. SAS distribution.

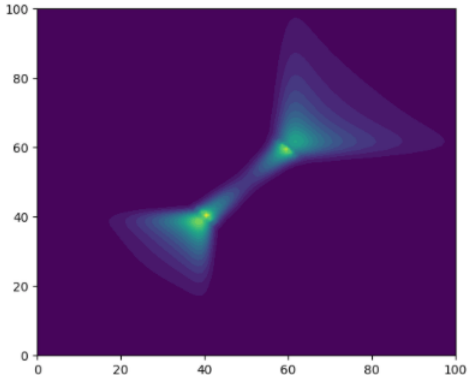


Figure 12: Shape S_6. Concatenation of two SAS and two skew-normal distributions.

Starting with the 'base' distribution, Figure 7 shows the density plot of a bivariate normal random

variable with parameters $\mu = (50, 50)$ and $\Sigma = 200I_2$. This is equivalent to a skew-normal distribution with the additional restraint of setting the skewness parameter $\delta \in \mathbb{R}^2$ to the zero vector. As expected, the mass is spherically concentrated around the mean.

Next, the previous bivariate normal distribution will be made asymmetrical by choosing a non-zero value for the skewness parameter. The parameters $\mu = (50, 50)$ and $\Sigma = 200I_2$ remain the same as before while $\delta = (0, 0)$ will be replaced by $\delta = (0.5, 0.5)$. The resulting shape is right-skewed in both directions which Figure 8 clearly shows.

In order to model more complex morphologies, it is possible to consider the sum of multiple random variables which have different underlying distributions. In this case, Figure 9 presents the geometry that results when two bivariate skew-normal distributions are summed up. The bottom-left part of this shape arises from a skew-normal with parameters $\mu_1 = (49, 49)$, $\Sigma_1 = [[180, 135], [180.135]]$ and $\delta_1 = (-0.5, -0.5)$ while the top-right part results from choosing the parameters $\mu_2 = (51, 51)$, $\Sigma_2 = [[180, 135], [180.135]]$ and $\delta_2 = (0.5, 0.5)$.

Similarly to the the last instance, Figure 10 displays the density of the sum of two bivariate skew-normal random variables. Here, the parameters corresponding to the skew-normal in the bottom-left part are $\mu_1 = (28, 28)$, $\Sigma_1 = [[280, 180], [180, 280]]$ and $\delta_1 = (0.5, 0.5)$ while the ones associated with the top-right part are $\mu_2 = (72, 72)$, $\Sigma_2 = [[280, 180], [180, 280]]$ and $\delta_2 = (0.5, 0.5)$. Note that even though this shape is again generated via a sum of skew-normals, altering the parameters yields entirely different morphologies.

Figure 11 displays the contour plot of an instance of the bivariate SAS distribition. The underlying parameters are $\mu = (50, 50)$, $\Sigma = 15I_2$, $\delta = (0.5, 0.5)$ for the skewness and $\nu = (1.5, 1.5)$ for the tailweights. Note that unlike the previous cases, there is no trace of elliptical geometry resulting in an entirely new class of distributions that can be obtained via the SAS transformation.

Lastly, the most complex shape of this project is constructed via the concatenation of densities of four random variables, two of which are bivariate skew-normal and two of which are of the bivariate SAS distribution. Figure 21 shows the contour plot of such a shape, where the parameters are chosen as follows: The triangular shape at the bottom-left arises from the SAS distribution with parameters $\mu_1 = (40, 40)$, $\Sigma_1 = 4I_2$, $\delta_1 = (-1, -1)$ and $\nu_1 = (1, 1)$ whereas the shape at the top-right has underlying parameters $\mu_2 = (60, 60)$, $\Sigma_2 = 7I_2$, $\delta_2 = (1, 1)$ and $\nu_2 = (1.2, 1.2)$. The piece connecting those two shapes is given by the sum of two skew-normal random variables. The one to the bottom-left has parameters $\mu_3 = (40, 40)$, $\Sigma_3 = [[30, 28], [28, 30]]$ and $\delta_3 = (5, 5)$ and for the other one $\mu_4 = (60, 60)$, $\Sigma_4 = [[30, 28], [28, 30]]$ and $\delta_4 = (-5, -5)$ were chosen.
Note that the sum of a SAS random variable and skew-normal one does not necessarily result in a new probability distribution which is why the contour plot is not smooth in regions where the distributions overlap. However, since the neuronal morphologies will be presented as density matrices of sampled instances, this is not of any concern here.

After having generated various candidate shapes for neuronal morphologies, they will be converted to density matrices as explained in section 3. These density matrices will be implemented into the model and for each pair or axon and dendrite shape, connectivity matrices are going to be computed. In the next chapter, the results will be presented along with some analysis of the connectivity structure.

6. Results and Analysis

One of the project's main incentives was to study connectivity structures of neuronal networks that arise from varying cell morphologies. Moreover, a strong emphasis was placed on modelling neuronal shapes to which the previous two chapters were extensively dedicated. As explained in section 3, connectivity matrices are given for each pair of axon and dendrite shape. Therefore, 36 connectivity matrices were computed. Examining each of those separately will be inconvenient and not very fruitful, which is why the analysis will focus on three pairs of shapes. For each pair the pseudocolor plot of the associated connectivity matrix is displayed below.

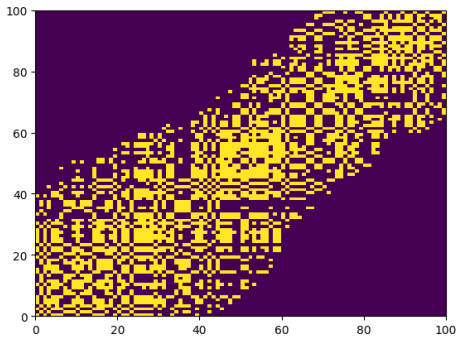


Figure 13: Matrix C_1. Connectivity matrix corresponding to choosing S_1 for the axon and S_6 for the dendrite morphology.

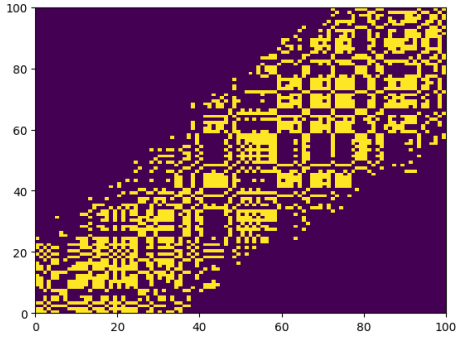


Figure 14: Matrix C_2. Connectivity matrix corresponding to choosing S_2 for the axon and S_4 for the dendrite morphology.

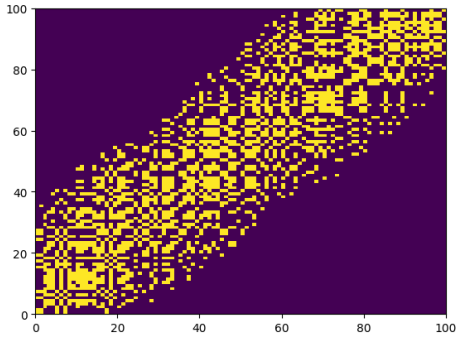


Figure 15: Matrix C_3. Connectivity matrix corresponding to choosing S_3 for the axon and S_5 for the dendrite morphology.

When examining the displayed connectivity matrices, the most prominent feature appears to be the diagonal spread of connections. It is important to note that this is not a selection bias but indeed noticeable for all 36 connectivity matrices. However, the diagonal structure arises trivially because of the ordering of neurons $N_1, \dots, N_{100} \in \mathbb{N}^2$. They were randomly sampled in the space $([1, 200] \cap \mathbb{N})^2$ in order to initiate the network X, see again section 3 for more details. The neurons were automatically sorted with increasing value in the first component which is why this diagonal shape is to be expected. Since the ordering has no meaning for the connectivity structure, this feature is to be neglected for the analysis.

Further note that these matrices are by design asymmetrical when the axon morphology is different from the dendrite morphology. This is the case for all three instances here. The asymmetry arises locally for single pairs of neurons as well as on a global scale which is especially evident for the matrices C_2 and C_3. There, the connections are not spread with symmetrical outreach from the diagonal line. For C_2 it appears that on average more signals are sent than received. The opposite can be noticed for C_3. Because all other hyperparameters were fixed, it suggests that this global asymmetry arises due to the varying neuronal morphologies.

Another eminent attribute that these matrices seem to adhere to is a grid-like structure. While for C_3 the connections seem more sporadic (at least for this ordering of neurons), the block structure is significantly more present for the matrices C_1 and C_2 where it is visible that there are groups of neurons having high/low intra-connectivity. The block structure changes with varied neuronal shapes which can be noticed by comparing C_1 and C_2. This means that the connectivity structure and therefore the flow of information within a network is intricately tied to neuronal morphologies. However, this is the extend of what these matrices can express since the grid structure is also dependent on the ordering of neurons. Therefore, it is not possible to say which exact groupings there exist within the network.

After having analysed some general properties of these connectivity matrices, it is desired to quantify the extend of the networks connectivity. For each neuron the number of axons connecting to it will be counted. This will indicate how many

signals each neuron receives from surrounding ones. The plots are shown below:

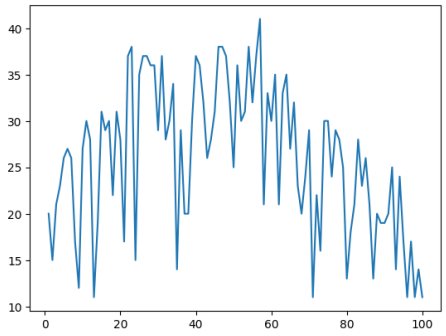


Figure 16: Number of axonal connections plotted for each neuron, corresponding to C_1.

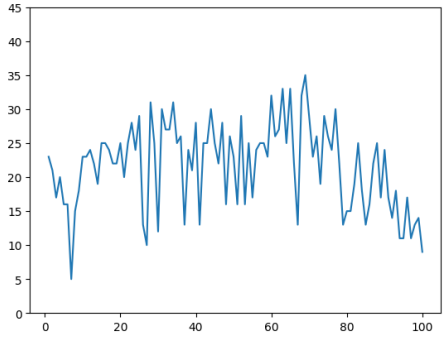


Figure 17: Number of axonal connections plotted for each neuron, corresponding to C_2.

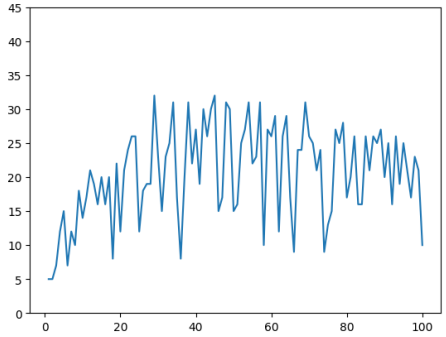


Figure 18: Number of axonal connections plotted for each neuron, corresponding to C_3.

When studying the plots for the number of axonal connections, their volatile behaviour immediately raises attention. However, the development over x-values has no meaning since it depends on the enumeration and ordering of neurons. Hence, it is only feasible to globally compare the plots between different connectivity matrices instead of performing an extensive individual analysis.
The most striking difference between these figures is the variation in the number of axonal connections each neuron can have. While for the matrix C_1 the number of connections per neuron varies from 10 to 40, there is relatively less disparity for C_2 and C_3 where it varies mostly between 10 and 30. This may be due to the fact that the shape S_1 covers a larger area in space which enables more connections than if operating with other shapes.
Another interesting metric would be to compare the number of neurons whose number of axonal connections surpasses a threshold value. For example, it was calculated that for C_1 38 neurons have > 30 axonal connections while for C_2 and C_3 only 10 and 11 neurons attain this many connections, respectively. It indicates that more connections are formed with the pair (S_1, S_6) than with (S_2, S_4) or (S_3, S_5). This conjunction can also be confirmed by simply counting the number of appositions formed: C_1 stores a total of 2600 connections while C_2 and C_3 store 2170 and 2055, respectively. Therefore, it is evident that varying neuronal morphologies can significantly alter the extend to which a network is connected.

For the last part of the analysis, the spectra of the connectivity matrices will be examined.

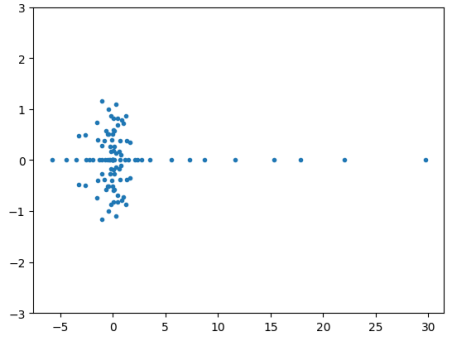


Figure 19: Spectrum of matrix C_1 consisting of complex values.

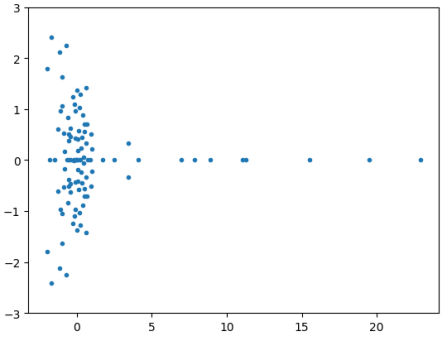


Figure 20: Spectrum of matrix C_2 consisting of complex values.

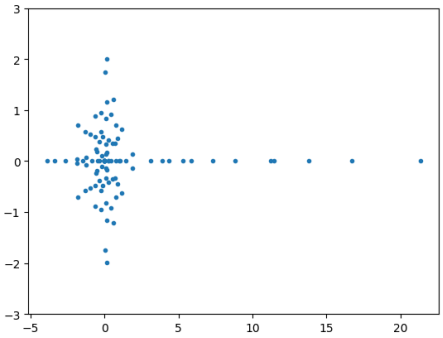


Figure 21: Spectrum of matrix C_3 consisting of complex values.

Note that the spectrum of a matrix is simply the set of its eigenvalues, which can either be real or complex valued. Eigenvalues and eigenvectors provide valuable information about a given matrix, including solubility of underlying systems of linear equations. Hence, studying the spectrum can be useful for a more extensive analysis of the connectivity structures.
Here, the most prominent difference between the spectra is the value of the imaginary parts of the eigenvalues. For C_1 their absolute value hardly exceeds 1 whereas for C_2 and C_3 they are more diluted. Similar to the previous comparison of axonal connections, the connectivity matrix C_1 stands out as compared to C_2 and C_3. It might be worthwhile to investigate whether the number of connections and the spectrum have any correlation to each other. However, this exceeds the scope of the project since it may involve spectral theory, but nonetheless could be interesting for future projects.

In this chapter it became evident that varying neuronal morphologies can have noticeable impact on a networks connectivity. In the first part of the analysis it was shown that different morphologies raise distinct block like structures in the connectivity matrix as well as significantly influence the number of connections a network is able to form. Due to the model limitations of only considering the bivariate setting and the extremely simple network architecture, it is difficult to make any conclusive statements for certain pairs of shapes. However, the multivariate probability distributions are still a powerful tool for modelling neuronal morphologies.

7. Conclusion and Future Work

In summary, this project developed a pipeline for estimating neuronal connectivity based on cell morphologies. For this intent, advanced multivariate probability distributions were reviewed in order to generate shapes that emulate axons and dendrites in 2D space. These distributions allowed significantly more flexibility due to the ability of regulating asymmetry as well as tail thickness. Especially the transformation approach was shown to be very powerful for modelling purposes since it allows to easily introduce new parameters due to the fact that the transformation itself can be chosen freely. For future work, it might be interesting to utilize these distributions in order to generate 3D morphologies based on empirical data. Within a suitable model, network connectivity could again be estimated and analysed.

References

Antonella Antonini and Michael Stryker. Effect of sensory disuse on geniculate afferents to cat visual cortex. *Visual neuroscience*, 15, 03 1998.

Slaana Babić, Christophe Ley, and David Veredas. Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data. *Symmetry*, 11(10), 2019.

Pierre Brémaud. *Probability Theory: A Comprehensive Course*. Springer Cham, 04 2020. ISBN 978-3-030-40182-5.

ClipartMax. 527px-neuron - annotated - svg - structure of biological neuron. URL https://www.clipartmax.com/middle/m2i8N4H7i8d3b1m2_527px-neuron-annotated-svg-structure-of-\biological-neuron/. [Online; accessed January 12, 2023].

Xiaolong Jiang, Shan Shen, Cathryn R Cadwell, Philipp Berens, Fabian Sinz, Alexander S Ecker, Saumil Patel, and Andreas S Tolias. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, 350(6264), 2015.

M. C. Jones and Arthur Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96(4), 10 2009.

Nir Kalisman, Gilad Silberberg, and Henry Markram. Deriving physical connectivity from neuronal morphology. *Biological Cybernetics*, 88(3), 2003.

Douglas Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 32(4), 1970.

Christof Koch and Idan Segev. The role of single neurons in information processing. *Nature neuroscience*, 3 Suppl, 12 2000.

Yangfan Peng, Federico J. Barreda Tomas, Paul Pfeiffer, Moritz Drangmeister, Susanne Schreiber, Imre Vida, and Jörg R.P. Geiger. Spatially structured inhibition defined by polarized parvalbumin interneuron axons promotes head direction tuning. *Science Advances*, 7(25), 2021.

Steven Sevush. *The Single-Neuron Theory*. Palgrave Macmillan Cham, 01 2016. ISBN 978-3-319-33707-4.

Daniel Udvary, Philipp Harth, Jakob H. Macke, Hans-Christian Hege, Christiaan P.J. de Kock, Bert Sakmann, and Marcel Oberlaender. The impact of neuron morphology on cortical network architecture. *Cell Reports*, 39(2), 2022.

Jaap van Pelt and Arjen van Ooyen. Estimating neuronal connectivity from axonal and dendritic density fields. *Frontiers in Computational Neuroscience*, 7, 2013.