

# Chapter 3 |

## Wavelet-based singularity detection for 3d signals with applications to fMRI data

### 3.1 Introduction and Motivation

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique, often used to localize brain regions associated with sensory, motor and cognitive functions. fMRI has the advantage of being noninvasive, but it is not capable of directly measuring neural activity. Instead, it measures the BOLD signal, defined as the changes in the ratio of oxygenated to deoxygenated blood, and often assumed to be a proxy measurement of neural activity. The BOLD signal is measured, as a time series, in three dimensional brain volumes, usually partitioned into volumetric elements (voxels) of equal size. Often, we are interested in measuring and analyzing the BOLD signal in response to an explicit task/stimulus. For example, the BOLD response  $x(t)$  is commonly represented by convolving a stimulus time course  $s$  with a model  $h$  for the hemodynamic response function [27] (this is referred to as the *general linear model*):

$$x(t) = \int_{-\infty}^{\infty} h(u)s(t-u)du.$$

fMRI scan images are commonly plagued with artifacts, and head motion is considered to be among the most damaging. Three independent studies [28–30], followed by a series of others [31–35], concluded that inference on the strength of signal correlation between nearby brain regions is systematically biased in the presence of head movements, even if they are very small, in the range of 0.5 mm to 1 mm. Currently, head motion artifacts

are identified by visual inspection or automated algorithms. A common practice is to employ motion estimates as nuisance covariates in the model. [36–39]. Another common approach is data scrubbing [28] – in which, informed by derivatives of head movement parameters and time courses, an algorithm removes suspect frames while pre-processing the time series. Noting that such algorithm does not always completely remove motion artifacts, [40] proposed wavelets to model and despike motion artifacts from resting state fMRI (rs-fMRI) time series, taking advantage of the localized behaviour of wavelets in both the time and frequency domains. This approach represents an adaptation to fMRI data of the *chain of maxima theory* first proposed by [41] to identify sharp changes in 1d signals. However, the adaptation by [40] does not reflect in full the approach in [41], which allows one to distinguish between sharp changes caused by artifacts and smoother changes due to real signals. We explain the chain of maxima theory and give an example in Sections 3.2.3 and 3.2.4.

In this chapter, we adopt the use of Lipschitz constants to characterize sharpness of signal changes proposed by Mallat in [41], and extend it to three dimensions. Different values of Lipschitz constants allow us to identify head motion artifacts in fMRI data while effectively distinguishing them from less sharp and smoother changes corresponding to real BOLD activity signals. We conduct a *traditional simulation study* to demonstrate the effectiveness of our method in detecting singularities in different dimensions within a general three-dimensional data set. Finally, we present two *data-based simulation settings*, built from real fMRI data sets with artificially-introduced head motion. These allow us to demonstrate that our proposed method can distinguish true BOLD activity from head motion signals.

The remainder of the chapter is organized as follow. Section 2 provides technical background on wavelet transforms and Mallat’s approach. We introduce our approach in Section 3, and provide a detailed description of our algorithm. Sections 4 and 5 contain the details and results of the traditional and data-based simulations, respectively. Finally, Section 6 provides some general discussion and remarks.

## 3.2 Technical background

### 3.2.1 Wavelet transform in one dimensional signals

In [42], Mallat introduced the theory of *multiresolution analysis* and proved that there exists a *scaling function*  $\phi(x) \in L^2(R)$ , such that if we define  $\phi_{2^j,n}(x) = \sqrt{2^{-j}}\phi(2^{-j}(x -$

$2^j n$ ), then  $\{\phi_{2^j,n}(x)\}_{n \in \mathbb{Z}}$  is an orthonormal basis of  $V_{2^j}$  – the vector space of all possible approximations at resolution  $2^j$  of functions in  $L^2(\mathbb{R})$ . The space  $V_{2^j}$  has several properties, as listed in [42]. Notably among them are:

1.  $V_{2^j} \subset V_{2^{j+1}}$ ;
2.  $\cup_{j=-\infty}^{+\infty} V_{2^j}$  is dense in  $L^2(\mathbb{R})$  and  $\cap_{j=-\infty}^{+\infty} V_{2^j} = \{0\}$ .

The second property guarantees that the approximation converges to the original function as the resolution increases. The first property means that the signal approximation at resolution  $2^{j+1}$  contains all the necessary information about the signal to be approximated at resolution  $2^j$ . The difference between the approximations at scale  $2^j$  and  $2^{j+1}$  is contained in  $O_{2^j}$ , the orthogonal complement of  $V_{2^j}$  in  $V_{2^{j+1}}$ . Let  $A_{2^j}$  and  $D_{2^j}$  be the orthogonal projections on  $V_{2^j}$  and  $O_{2^j}$ , respectively. For any  $J > 0$  and any original signal (function)  $f \in L^2(\mathbb{R})$  discretely measured at resolution  $2^0 = 1$ ,  $f$  can be represented by  $(A_{2^{-J}} f, (D_{2^j} f)_{-J \leq j \leq -1})$  [42]. This representation is called an *orthogonal wavelet representation*.

A *wavelet function*  $\psi(x)$  can be derived from the scaling function  $\phi(x)$  such that if we set  $\psi_{2^j,n}(x) = \sqrt{2^{-j}}\psi(2^{-j}(x - 2^j n))$ , then  $\{\psi_{2^j,n}(x)\}_{n \in \mathbb{Z}}$  is an orthonormal basis of  $O_{2^j}$  and  $\{\psi_{2^j,n}(x)\}_{(n,j) \in \mathbb{Z}^2}$  is an orthonormal basis of  $L^2(\mathbb{R})$  [42]. The *wavelet transform* of a signal  $f \in L^2(\mathbb{R})$  at time  $u$  and scale  $2^j$  is given by

$$\begin{aligned} Wf(u, 2^j) &= \langle f, \psi_{2^j,u} \rangle \\ &= \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j u}{2^j}\right) dt \\ &= \int_{-\infty}^{+\infty} f(t) \bar{\psi}_{2^j}(2^j u - t) dt \\ &= f * \bar{\psi}_{2^j}(2^j u) \end{aligned}$$

where  $\bar{\psi}_s(u) = \frac{1}{\sqrt{s}}\psi\left(\frac{-u}{s}\right)$ . The scaling/dilating and translating of the wavelet function allows us to reorganize the signal into a “set of details” appearing at different resolutions. This makes wavelet transforms an effective tool for detecting structures at different scales, such as local variation of the signal intensity.

### 3.2.2 Wavelet transform in two and three dimensional signals

Let  $I(x, y)$  be an image in  $L^2(\mathbb{R}^2)$  and  $V(x, y, z)$  be a volumetric image in  $L^2(\mathbb{R}^3)$ . In [42], Mallat extended the one dimensional result to two dimensions and proved that there

exist a unique scaling function  $\Phi(x, y) = \phi(x)\phi(y)$ . Then  $(2^{-j}\Phi(2^{-j}(x - 2^jn), 2^{-j}(y - 2^jm)))_{(n,m) \in \mathbb{Z}^2}$  is an orthonormal basis of  $V_{2^j}$  – the space of approximations of images in  $L^2(\mathbb{R}^2)$ . If we define the three ‘wavelets’

$$\Psi^1(x, y) = \phi(x)\psi(y) \quad \Psi^2(x, y) = \psi(x)\phi(y) \quad \Psi^3(x, y) = \psi(x)\psi(y)$$

then

$$\left(2^{-j}\Psi^k\left(\frac{x - 2^jn}{2^j}, \frac{y - 2^jm}{2^j}\right)\right)_{(k,n,m) \in \{1,2,3\} \times \mathbb{Z}^2}$$

is an orthonormal basis for  $O_{2^j}$ , and

$$\left(2^{-j}\Psi^k\left(\frac{x - 2^jn}{2^j}, \frac{y - 2^jm}{2^j}\right)\right)_{(k,j,n,m) \in \{1,2,3\} \times \mathbb{Z}^3}$$

is an orthonormal basis for  $L^2(\mathbb{R}^2)$ .

Since our application is aimed towards three dimensional volumetric data, in a similar fashion, let  $\Phi(x, y, z) = \phi(x)\phi(y)\phi(z)$ . Then  $(2^{-3j/2}\Phi(2^{-j}(x - 2^jn), 2^{-j}(y - 2^jm), 2^{-j}(z - 2^jl)))_{(n,m,l) \in \mathbb{Z}^3}$  is an orthonormal basis of  $V_{2^j}$  – the space of approximations of volumetric images in  $L^2(\mathbb{R}^3)$ . As in the case of two dimensions, in three dimensions we can define seven ‘wavelets’ as

$$\begin{aligned} \Psi^1(x, y, z) &= \phi(x)\phi(y)\psi(z) & \Psi^2(x, y, z) &= \phi(x)\psi(y)\phi(z) & \Psi^3(x, y, z) &= \phi(x)\psi(y)\psi(z) \\ \Psi^4(x, y, z) &= \psi(x)\phi(y)\psi(z) & \Psi^5(x, y, z) &= \psi(x)\psi(y)\phi(z) & \Psi^6(x, y, z) &= \psi(x)\phi(y)\phi(z) \\ \Psi^7(x, y, z) &= \psi(x)\psi(y)\psi(z) \end{aligned}$$

such that

$$\left(2^{-3j/2}\Psi^k\left(\frac{x - 2^jn}{2^j}, \frac{y - 2^jm}{2^j}, \frac{z - 2^jl}{2^j}\right)\right)_{(k,n,m,l) \in \{1,2,\dots,7\} \times \mathbb{Z}^3}$$

is an orthonormal basis for  $O_{2^j}$ , and

$$\left(2^{-3j/2}\Psi^k\left(\frac{x - 2^jn}{2^j}, \frac{y - 2^jm}{2^j}, \frac{z - 2^jl}{2^j}\right)\right)_{(k,j,n,m,l) \in \{1,2,\dots,7\} \times \mathbb{Z}^4}$$

is an orthonormal basis for  $L^2(\mathbb{R}^3)$ . Thus, a three dimensional volumetric signal can be decomposed into one approximation volume, and seven other detail volumes. The

seven wavelet functions are sometimes called filters. Since  $\phi$  is an approximation basis function and  $\psi$  is a detail basis function, the wavelet function  $\Psi^1(x, y, z)$ , for example, acts as a low pass/low pass/high pass (LLH) filter on the original signal  $f$ . At each scale (resolution), the volume resulting from applying this filter has smoothed out details in the  $x$  and  $y$  direction, and maintains the rough details in the  $z$  direction.

### 3.2.3 Singularity detection through Lipschitz exponents

The treatment in [41] characterizes singularities – or sharp variation – in one dimension through Lipschitz exponents. Based on [43] we define a Lipschitz function as follows.

**Definition 3.2.1** *A function  $f$  is pointwise Lipschitz  $\alpha \geq 0$  at  $v$ , if there exists  $K > 0$  and a polynomial  $p_v$  of degree  $m = \lfloor \alpha \rfloor$  such that*

$$\forall t \in \mathbb{R}, |f(t) - f(v)| \leq K|t - v|^\alpha. \quad (3.1)$$

*A function  $f$  is uniformly Lipschitz  $\alpha$  over  $[a, b]$  if it satisfies (3.1) for all  $v \in [a, b]$  with  $K$  independent of  $v$ .*

The use of Lipschitz exponents is directly tied to the concept of vanishing moments, as shown in the following definition.

**Definition 3.2.2** *A wavelet  $\psi(x)$  is said to have  $n$  vanishing moments if and only if  $\int_{-\infty}^{+\infty} x^k \psi(x) dx = 0$ , for all positive integer  $k < n$ .*

The detection of singularities can therefore be summarized in the following steps:

1. Suppose  $\psi(x)$  has  $n$  vanishing moments for some positive integer  $n$ . Calculate the wavelet transform  $Wf(x, s)$ .
2. For each scale  $s$ , find the local modulus maxima of  $Wf(x, s)$  along the  $x$  direction. That is, for each scale  $s_0$ , find  $(s_0, x_0)$  such that  $|Wf(s_0, x_0)|$  is maximum in a neighborhood of  $x_0$ .

Based on these modulus maxima, [41] proved the following facts:

1. **Theorem 3 from [41]** All singularities of  $f$  are contained in the closure of the wavelet transform maxima of  $f$ ; if we chain  $x$ -neighboring modulus maxima across the scales  $s$ , all singularities can be detect by tracing the chain as  $s$  goes to 0. This is the approach used in [40]. Note that, in reality, if signals are measured discretely at resolution 1, we cannot trace the chain further than  $s = 1$ .

2. **Theorem 5 from [41]** Let  $x_0 \in \mathbb{R}$ ,  $[a, b]$  be a neighborhood of  $x_0$  and  $X(s)$  be a curve in the scale space  $(s, x)$  such that  $|x_0 - X(s)| \leq Cs$ . If  $f$  has no fast oscillations, and in addition

- there exists a scale  $s_0$  such that the wavelet transform  $Wf(x, s)$  has constant sign on  $[a, b]$  for all  $s < s_0$
- there exists a constant  $B$  and an  $\epsilon > 0$  such that  $|Wf(x, s)| \leq Bs^\epsilon$  on  $[a, b]$  and for any scale  $s$
- there exists a constant  $A$  such that for any scale  $s < s_0$ , the wavelet transform satisfies  $|Wf(x, X(s))| \leq As^\gamma$ , with  $0 \leq \gamma \leq n$

then  $f$  is Lipschitz  $\alpha$  at  $x_0$ , for any  $\alpha < \gamma$

In summary, the first fact states that all singular points can be detected by tracing the modulus maxima line in the scale plane. The second fact provides a way to estimate the local Lipschitz regularity at any singular point from the decay of the wavelet coefficients along these modulus maxima line.

### 3.2.4 An example in one dimension

Here we provide an example based on [41]. At the top of Figure 3.1 is the plot of a time series with three different kinds of irregularities. The first kind, at time point 200, is a discontinuous, short-lived spike. The second kind, starting from time point 400, is a continuous transition on both sides. The third kind, starting at time point 700, is a combination of a discontinuous transition from the left, and a continuous transition to the right hand side. As shown in the bottom right plot of Figure 3.1, the highly abrupt singularity at time point 200 has an estimated Lipschitz constant  $\alpha = -0.9$ , whereas smoother signals have higher, positive estimated Lipschitz constants. As we mentioned earlier, the approach proposed in [40] stops at identifying the modulus maxima, and does not distinguish between smoother signals and sharp transitions. In many application such as fMRI, the BOLD signal resembles smooth transitions such as that at time point 400 in this example, whereas head motion artifacts will cause abrupt change such as that at time point 200.

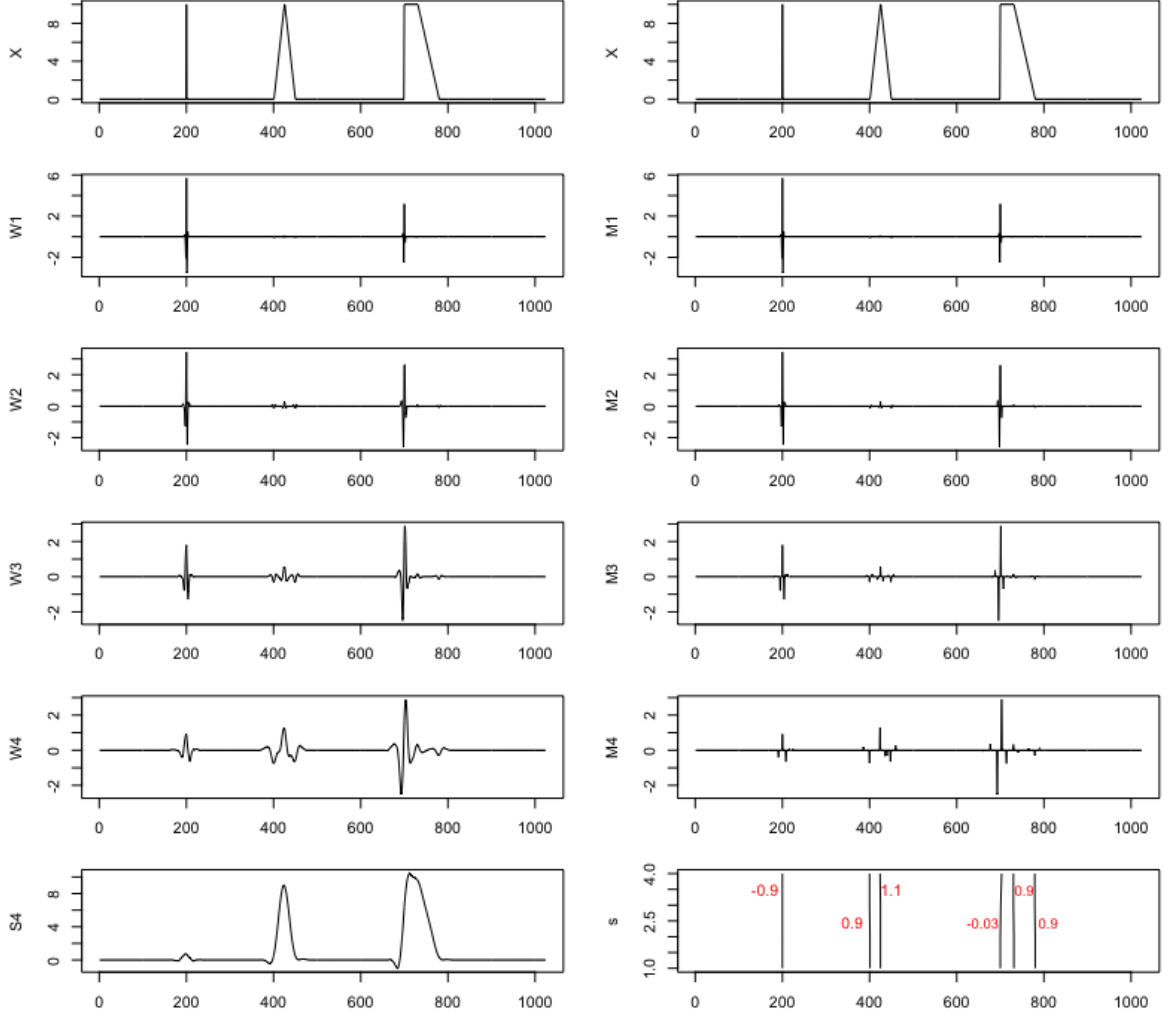


Figure 3.1: *Left column: at the top is the original time series, followed by plots of its wavelet transform coefficients at four different scales. The bottom plot shows scaling coefficients. Right column: at the top is the original time series, followed by plots of local maxima of the corresponding wavelet coefficients on the left. The bottom plot is the chaining of neighboring maxima across four different scales. The estimated decay of wavelet coefficients across scales is reported (in red) near each chain.*

### 3.3 Our approach

The structure we are interested in is a three dimensional, volumetric one. It comprises a predominant noisy signal with occasional disturbances in the form of sharp, extreme variations. This set-up can be modeled as

$$V_{xyz} \sim \beta N(f_{xyz}, \sigma^2) + (1 - \beta) N(g_{xyz}, \sigma^2) \quad (3.2)$$

where  $f_{xyz}$  is a three dimensional surface that is uniformly Lipschitz with a positive Lipschitz constant, and  $g_{xyz}$  is again a three dimensional surface, but it contains sharp variation with negative Lipschitz constant. Our goal is to separate the sharp variation from the smooth signal. Regarding the fMRI application that motivates us, we aim to identify where, in time and 2D space (a single slice of the brain), head motion occurs – under the assumption that head motion artifacts are sharp and BOLD signals are smooth.

### 3.3.1 Iterative singularities detection and smoothing

In order to identify sharp artifacts and separate them from the smooth signal, we propose the following algorithm.

- S1 (Wavelet transformation) Compute the three dimensional maximal overlap discrete wavelet transform (3D MODWT) of the volumetric data at depth  $J$  (in R, this can be done using the *waveslim* package; Whitcher, 2020). Since the depth of the wavelet decomposition is  $J$  (corresponding to scales  $s = 2^j, j = 1, \dots, J$ ), the 3D MODWT produces  $7J + 1$  3D filters

$$\left( (HHH, HLH, HHL, LHH, HLL, LLH, LHL)_j, LLL_J \right), j = 1, \dots, J$$

where  $H$  and  $L$  indicate high- and low-pass filters, respectively. These transforms are the coefficients of the level  $J$  decomposition of the original 3D data, using the basis functions in 3.2.2. There are many wavelet basis functions to choose from [43, 44]. In our simulations and applications, we use la(8) wavelet basis.

- S2 (Filter alignment) Circularly shift the wavelet transform at each scale  $2^j$  to align events between the wavelet space and the original data space. For example, in data of size  $64 \times 64 \times 64$ , a (4,5,6) shift moves the voxel at position  $x = 20, y = 2, z = 63$  to position  $x = 24, y = 7, z = 5$ . The amount of shift depends on the choice of wavelet basis and the scale. For la(8) wavelet, the shift at  $2^j$  is  $3(2^j - 1)$  for low-pass filters and  $\frac{7(2^j - 1) + 1}{2}$  for high-pass filters [44].
- S3 (Maxima search) For each filter, take a sliding window of size  $w_1$ , and define a voxel  $V_{xyz}$  to be a local modulus maxima if

$$V_{xyz} = \max_{x-w_1 \leq t \leq x+w_1, y-w_1 \leq u \leq y+w_1, z-w_1 \leq t \leq z+w_1} V_{tuv}$$



- S4 (Maxima chaining) For each filter, chain the maxima across scales if they belong to neighboring voxels. For example, if  $V_{xyz}$  and  $V_{tuv}$  are modulus maxima at scale  $2^j$  and  $2^{j+1}$ , respectively, and  $w_2$  is the size of the neighborhood, then they belong to the same chain if  $x - w_2 \leq t \leq x + w_2, y - w_2 \leq u \leq y + w_2, z - w_2 \leq t \leq z + w_2$ .
- S5 (Lipschitz constant estimation) For each maxima chain, take the maximum of modulus maxima  $M_j$  at each scale  $2^j$ , and estimate the value of  $\alpha$  such that  $M_j \leq (2^j)^\alpha$ . This can be done through a simple linear regression of  $\log(2^j)$  on  $\log(M_j)$ ,  $j = 1, \dots, J$ . Note that, instead of taking the maximum of modulus maxima, one could consider their mean or their median. However, if the signal is noisy but the singularities are large enough, the maximum would be more robust against small noise coefficients when employing a large neighborhood size  $w_2$ .
- S6 (Smoothing) For those chains whose Lipschitz constant is negative, zero out all the wavelet coefficients in that chain.
- S7 (Iteration) Take an inverse transform of the zeroed-out wavelet coefficients and repeat Step 1.

The last two steps are optional, and are employed to remove singularities through smoothing. Indeed, zeroing out wavelet coefficients has the effect of down-weighting the contribution of singularities in the decomposition of the original signal.

### 3.3.2 Directional singularity detection

The algorithm above is capable of identifying sharp transitions in a given canonical direction. For example, to identify head motion artifacts in fMRI imaging scans, we care about signal disruption in the time axis only, whereas it is quite normal for signals to have sharp transitions along the spatial axes. Suppose we are interested in identifying sharp transitions in the  $z$ -direction of the volumetric data  $V_{xyz}$ . This can be done by locating transitions with Lipschitz constant  $\alpha \leq -0.5$  in the filter  $LLH$ . The following theorem relates the smoothness of the signal in a canonical direction to the decay of wavelet coefficients in the corresponding filter. A proof can be found in the Appendix.

**Theorem 3.3.1** *Choose a wavelet function  $\psi$  with  $n$  vanishing moments and  $n$  derivatives having a fast decay. If  $f \in L^2(\mathbb{R}^3)$  is Lipschitz  $\alpha$  in the  $z$ -direction at  $\mathbf{v}$ , i.e. if there exist a polynomial  $p_{\mathbf{v}}$  such that*

$$|f(\mathbf{t}) - p_{\mathbf{v}}(\mathbf{t})| \leq K(t_1, t_2) \cdot |t_3 - v_3|^\alpha$$

and  $K(t_1, t_2) \in L^2(\mathbb{R}^2)$ , then

$$\forall (u, s) \in \mathbb{R}^3 \times \mathbb{R}^+, |W^{LLH} f(\mathbf{u}, s)| \leq A s^{\alpha+1/2} \left( 1 + \left| \frac{u_3 - v_3}{s} \right|^\alpha \right).$$

Similarly, abrupt changes in two canonical directions, say both the  $y$ - and  $z$ - direction, are expected to have Lipschitz constant  $\alpha \leq -0.5$  at the corresponding locations in the filters  $LLH$ ,  $LHL$  and  $LHH$ . The above Theorem states the necessary condition for  $f$  to be Lipschitz  $\alpha$  in one direction, but our simulations in Section 3.4 provide evidence also concerning the other direction; that is, the decay rate of wavelet coefficients in different filters informs us about the regularity of  $f$  in the corresponding (multiple) directions.

### 3.4 Traditional simulations: noisy signals with sharp changes in one, two and three directions

In the following simulations,  $V_{xyz} : 1 \leq x, y, z \leq 64$  are three dimensional volumetric data. We consider three different kinds of signal: signals that change sharply in 1 direction, 2 directions, and 3 directions (Figure 3.2). The expectation is that we can identify each type of signal by looking at the estimated Lipschitz constants of the 7 filters at the corresponding locations, as explained in Section 3.3.2. We also add independent Gaussian noise on top of the signals, and study how varying levels of noise affect the results.

For the first kind of signal, which has a sharp transition in the  $z$ -direction and is smooth in the other two directions, we let  $V_{x,y,50}$  be a scaled bivariate Gaussian density with peak at  $x = y = 30$ . The second kind of signal varies sharply in two directions and is smooth in the remaining direction. For this, we let  $V_{11 \leq x \leq 30, 20, 20}$  be an increasing sequence from 0 to 10, and  $V_{31 \leq x \leq 50, 20, 20}$  be a decreasing sequence from 10 to 0. To highlight the ability of the algorithm to differentiate between transitions in different directions, we also let  $V_{43, 43, 21 \leq z \leq 30}$  be an increasing sequence from 0 to 10, and  $V_{43, 43, 31 \leq z \leq 40}$  be a decreasing sequence from 10 to 0. For the third kind of signal, we let  $V_{10, 10, 30} = 10$  represent the sharp transition in all three directions;  $V_{xyz} = 0$  everywhere else. Adding independent Gaussian noise on top of our signals we control the signal to noise ratio (STN) defined as the ratio between the magnitude of the sharp transition and the noise level  $\sigma$ . Below we provide results with STN on the order of 100 (i.e.  $\sigma = 0.1$ )

and 10 (i.e.  $\sigma = 1$ ).

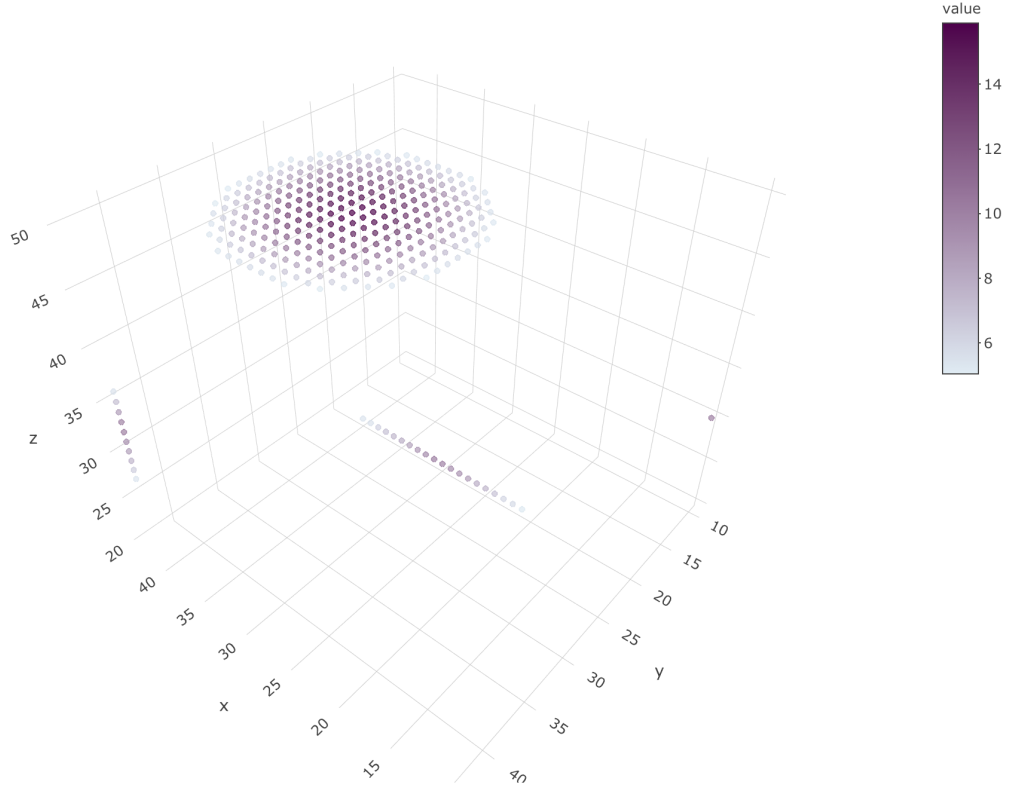


Figure 3.2: *Simulated signal with sharp transitions in one, two and three directions. The signal is shown here prior to adding noise.*

In our algorithm, we use the maximal overlap discrete wavelet transform with the Daubechies least asymmetric filter  $\mathbf{la}(\mathbf{8})$ , with  $J = 3$ . The filters at scale  $2^3$  have about 20 non-zeroes coefficient, whereas at scale  $2^4$  the number of non-zeroes coefficient is approximately 40 (Figure 3.3). Since the length of this data is 64 in any direction, at  $J = 4$  the long filter has the effect of suppressing too many fine details, which leads to identifying too few or no maxima at large scale. We use  $w_1 = 3$  and  $w_2 = 1$ . With  $w_1 = 3$  maxima are searched in each  $7 \times 7 \times 7$  block, large enough to ensure affordable computation and small enough that we do not miss the peaks that are near other peaks of larger magnitude. For more information about the  $\mathbf{la}(\mathbf{8})$  filter, see [44].

Table 3.1 shows the Lipschitz constant estimation at the locations where signals vary sharply in one, two or three directions. The STN here is 100, and the results are averages over 10 runs, with reported standard deviation in parentheses. We also indicate at the

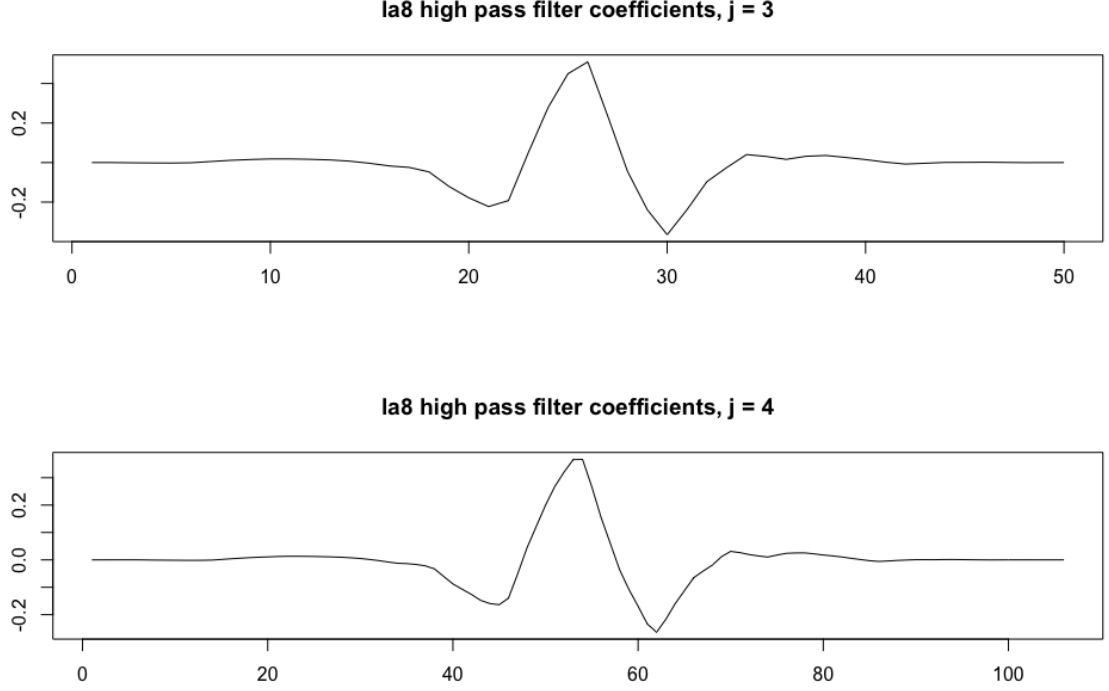


Figure 3.3: *Least asymmetric (la8) high pass filter coefficients at scales  $s = 2^{-3}$  and  $s = 2^{-4}$ .*

top the directions in which the signal changes abruptly. The results confirm that the algorithm is capable of detecting singularities in different directions. For example, at location  $(x, y, z) = (43, 43, 30)$ , the signal varies sharply in the  $x$  and  $y$  directions, and is smooth in the  $z$  direction. We expect low, negative estimated Lipschitz constant  $\hat{\alpha}$  in the HLL, LHL, and HHL filters. The filter LLH also has a negative  $\hat{\alpha} = -0.3$ , but this is much higher than the estimated  $\hat{\alpha}$  in the HLL, LHL, HHL filters. We also observe that the standard deviations of  $\hat{\alpha}$  are much lower for these three filters.

Table 3.2 shows results when the STN is lowered to 10. We observe somewhat similar outcomes compared to when STN is 100 but, as to be expected, they are now less clear cut. In particular, we observe higher standard deviation of Lipschitz estimates at all locations, and for all filters.

filter	$V_{30,30,50}$ [z]	$V_{30,20,20}$ [yz]	$V_{43,43,30}$ [xy]	$V_{10,10,30}$ [xyz]
HLL	1.21 (0.45)	-0.83 (0.31)	-1.79 (0.09)	-2.83 (0.08)
LHL	1.13 (0.33)	-1.85 (0.01)	-1.80 (0.09)	-2.84 (0.09)
LLH	-0.85 (0.00)	-1.85 (0.01)	-0.30 (0.19)	-2.85 (0.07)
HHL	-0.51 (0.32)	-0.68 (0.30)	-1.66 (0.09)	-2.67 (0.05)
HLH	1.34 (0.23)	-0.71 (0.23)	-0.17 (0.28)	-2.67 (0.04)
LHH	1.23 (0.29)	-1.70 (0.01)	-0.15 (0.15)	-2.69 (0.04)
HHH	-0.44 (0.38)	-0.62 (0.22)	-0.01 (0.27)	-2.53 (0.05)

Table 3.1: *Lipschitz constant estimation at the locations where signals vary sharply in one, two or three directions (STN = 100)*

filter	$V_{30,30,50}$ (z)	$V_{30,20,20}$ (yz)	$V_{43,43,30}$ (xy)	$V_{10,10,30}$ (xyz)
HLL	-0.27 (0.43)	-1.85 (0.45)	-1.92 (0.18)	-2.72 (0.46)
LHL	-0.42 (0.35)	-1.95 (0.14)	-1.94 (0.17)	-2.89 (0.50)
LLH	-0.87 (0.02)	-2.01 (0.13)	-1.20 (0.26)	-2.91 (0.45)
HHL	-1.52 (0.5)	-1.58 (0.32)	-1.76 (0.19)	-2.59 (0.39)
HLH	-0.29 (0.42)	-1.62 (0.49)	-1.26 (0.32)	-2.63 (0.21)
LHH	-0.41 (0.31)	-1.83 (0.09)	-1.31 (0.35)	-2.62 (0.44)
HHH	-1.23 (0.39)	-1.64 (0.39)	-1.18 (0.37)	-2.49 (0.41)

Table 3.2: *Lipschitz constant estimation at the locations where signals vary sharply in one, two or three directions (STN = 10)*

## 3.5 Data-based simulations: real fMRI data with artificial head motion

In this section, we demonstrate the effectiveness of our proposed methodology on a real fMRI data set in which we artificially introduce head motion.

### 3.5.1 Real data description

We employ data taken from a large data collection called the Amsterdam Open MRI Collection [45], made publicly available from the Openneuro data sharing platform. We focus on the fMRI study of response inhibition, conducted through a stop-signal task based on [46]. The following description is taken from [45]:

*Subjects were presented with 100 trials, during each of which an image of either a female or male face (chosen from 9 exemplars) was shown for 500 ms on a black background. Subjects had to respond whether the face was female (right index finger) or male (left index finger) as quickly and accurately as*

*possible, except when an auditory stop signal (a tone at 450 Hz for 0.5 seconds) was presented (on average 33% of the trials). The delay in presentation of the stop signal (i.e., the “stop signal delay”) was at start of the experiment 250 milliseconds, but was shortened with 50 ms if stop performance, up to that point, was better than 50% accuracy and shortened with 50 ms if it was worse. Each trial had a duration of 4000 ms and was preceded by a jitter interval (0, 500, 1000 or 1500 ms). If subjects responded too slow, or failed to respond an additional feedback trial of 2000 ms was presented. Note that due to this additional feedback trial and the fact that subjects differed in how many feedback trials they received, the fMRI runs associated with this task differ in length across subjects (i.e., the scan was manually stopped after 100 trials). Additionally 10% (on average) null trials with a duration of 4000 ms were presented randomly.*

Even though we retrieve a  $65 \times 77 \times 60 \times 231$  data array for each participant, for our simulations we manipulate a much smaller array consisting of 231 measurements in time taken on the 30th slice (dimension  $65 \times 77$ ) of one participant’s brain.

### **3.5.2 Simulation 1: Head motion is the only change in the fMRI time courses**

In this simulation, we first take the 2d slice number 30 of the brain observed at time point 30, and replicate this slice 64 times. As a next step, we induce head motion by rotating the entire original brain by  $5^\circ$  in three dimension: yaw, pitch and roll. This is done using a rotation matrix and matrix multiplication. We then extract the “motion affected” slice number 30 and place it in the middle of the otherwise 64 identical slices. The result of this procedure is a three-dimensional data set consisting of 65 “copies” of slice 30 stacked, across time, on top of one another – with the only change in signals across time due to the artificial head motion at time point 33 (Figure 3.4).

As we see in Figure 3.4, rotating the brain by even only  $5^\circ$  will severely displace the values of almost all brain voxels in the selected slice. We note that the density of the blue circles violate one of the assumptions required by our method; that is, that there is no spike in a small neighborhood of another. If there are, say, two spikes that occur next to each other, one or more of the following will happen:

- Step 3 of our algorithm will fail to identify the spike with smaller magnitude as a local maxima

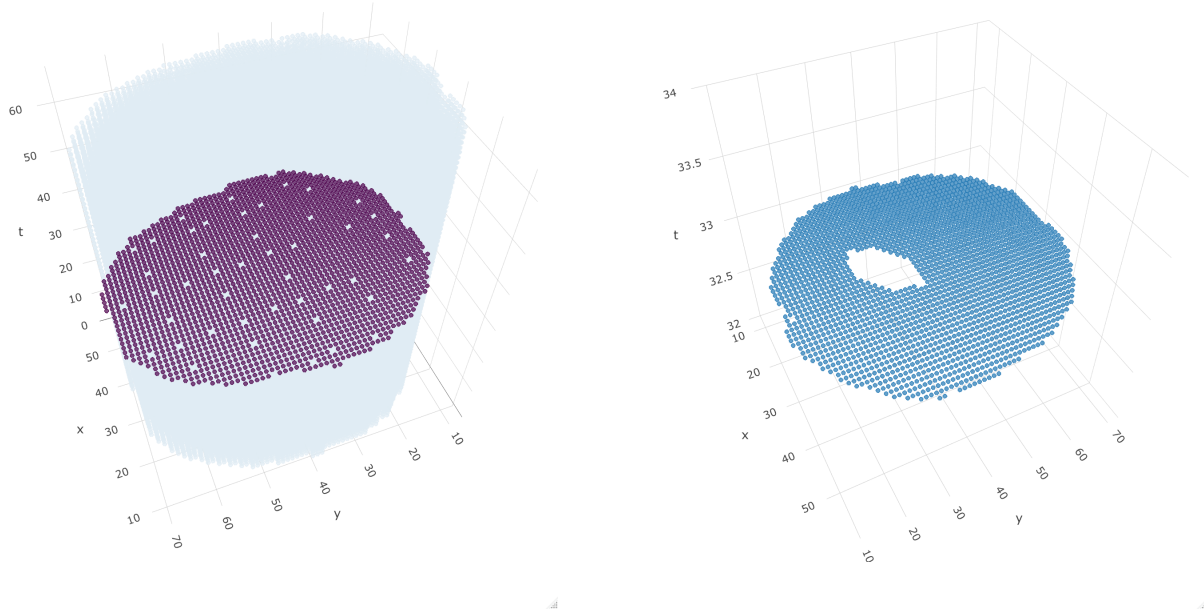


Figure 3.4: *Simulated data. Left: The motion affected slice, plotted in purple, is placed in the middle of 64 otherwise identical slices. Right: blue circles indicate the displaced voxels in the affected slice due to head motion.*

- Step 4 of our algorithm will chain two different spikes, which in turns means that in Step 5 the algorithm will give only one Lipschitz constant estimate

In the following we show that, even though our method does not identify all spikes, it still works effectively at flagging a head motion affected scan. Because we only care about abrupt changes in the time direction, we look to the *LLH* filter, as explained in Section 3.3.2. Figure 3.5 plots as blue circles the singularities ( $\alpha < -0.5$ ) identified by the algorithm. Aside from a few voxels located in neighboring time points, most of the identified singularities lie on slice 33, which is the true motion affected slice. The algorithm flagged 350 points as singularities, 304 of which are true spikes due to head motion. Here, we used  $\omega_1 = 1$  and  $\omega_2 = 1$ . For a discussion of how the values of  $\omega_1, \omega_2$  affect results, see Section 3.4.

### 3.5.3 Simulation 2: Head motion is induced in randomly selected time points in real data

In this simulation, we take a 2d-slice (slice number 38) measured at all 231 time points. We then add head motions to 23 randomly selected time points ( $\approx 5\%$  contamination).

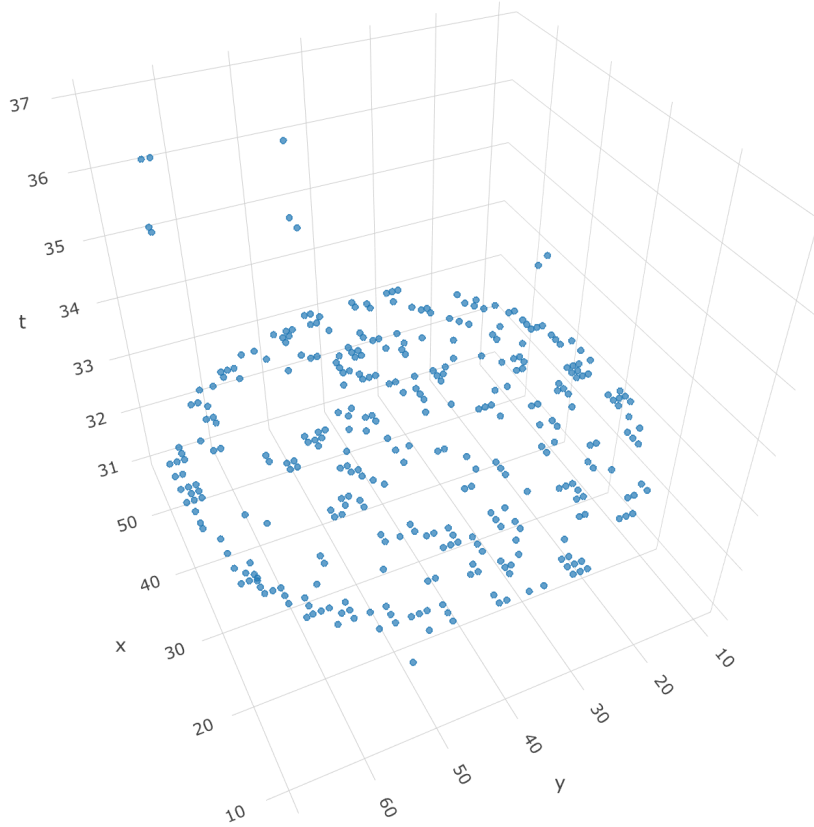


Figure 3.5: *Result of the simulation in Section 3.5.2. Blue circles indicate voxels identified as singularities in the LLH filter.*

Head motions are generated by rotating the entire 3d brain  $1^\circ$  in two dimensions (yaw and pitch) at the selected time points. Figure 3.6 shows the displaced voxels due to head motion at one affected time point (time point 33). We then perform a simple linear analysis on the original data and the data with injected head motion. The measured BOLD signal acts as the response, and the predictors are the stop/go conditions convolved with the hemodynamic response function. Figure 3.7 shows the voxels that have significant activation during the stop trials for the original data (left) and the head motion affected data (right).

Figure 3.8 shows the voxels that are flagged by our algorithm as having Lipschitz estimates  $\alpha < -1$  in the time direction. We profile the time points according to their



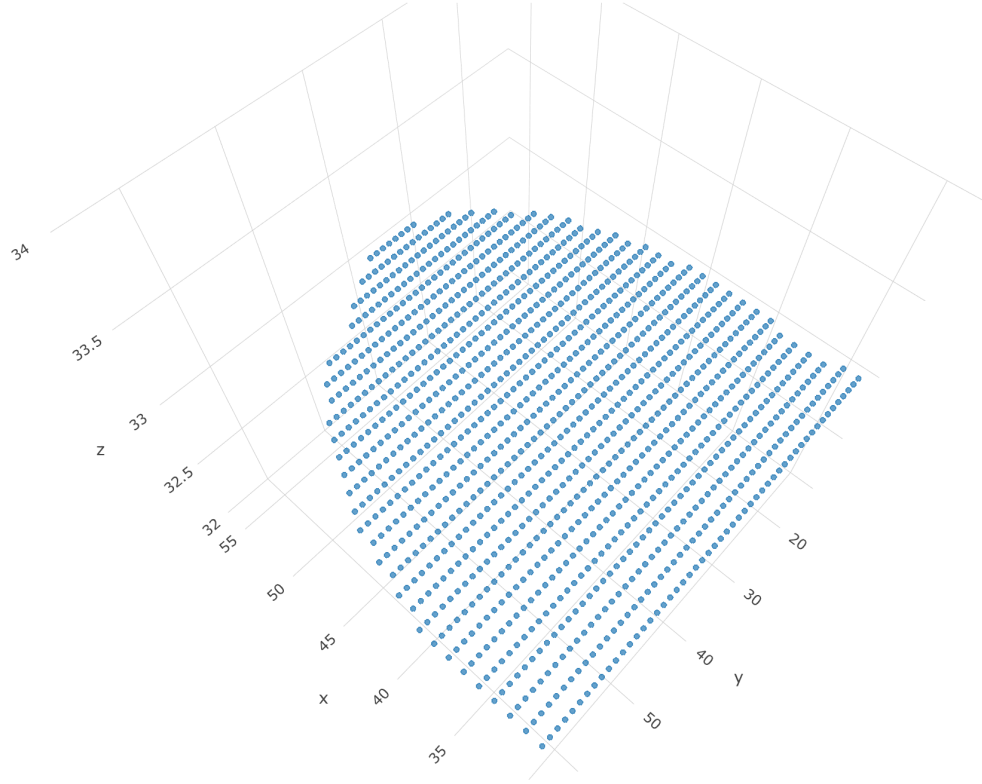


Figure 3.6: *The effect of rotating a 3d brain 1° in yaw and pitch directions. Blue circles indicate the displaced voxels due to head motion at slice number 38.*

number of flagged voxels  $N_{FV}$ , and plot the  $N_{FV}$  in a histogram. Based on this histogram, a natural cutoff value would be 95, and using this cutoff, we achieve 78% (18 out of 23 affected time points) true positive rate. This rises to 87% (20 out of 23 affected time points) if we bring the cutoff value down to 60. It is worth noting that all of the false positives are time points immediately neighboring true positives, so even the false positives in our analysis may be used to help pinpoint head motions.

### 3.6 Discussion

We described a wavelet-based algorithm that is capable of detecting directional singularities in three-dimensional data, based on a procedure proposed in [41] for one-dimensional signal. These directional singularities are characterized by a number, called the Lipschitz constant, which can be estimated by exploiting its relationship with the decay rate

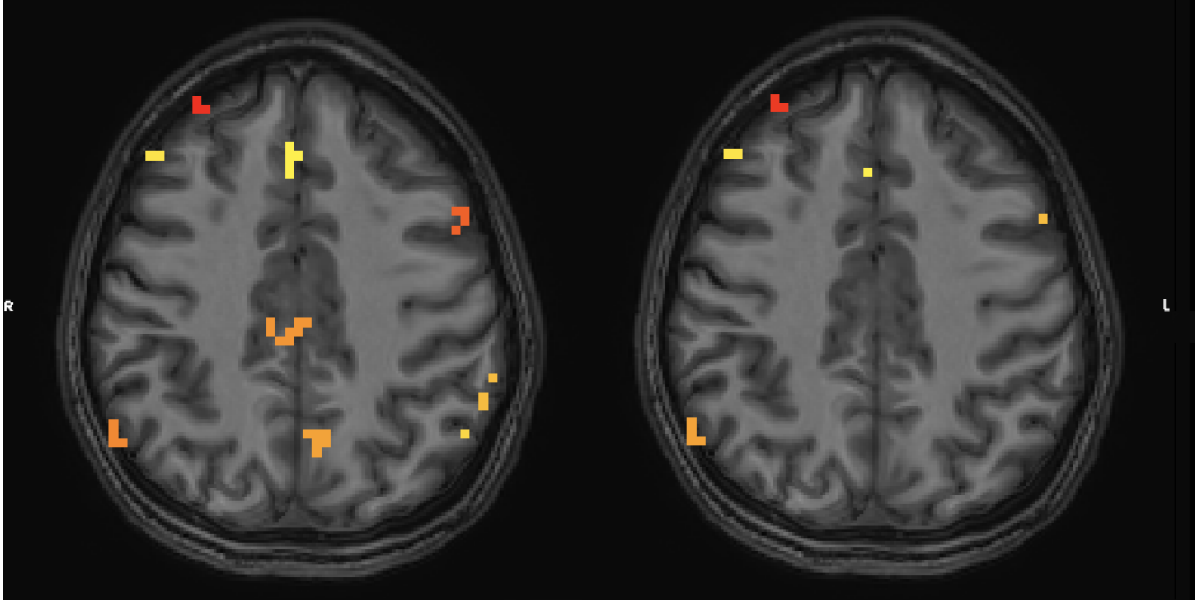


Figure 3.7: *Result of simulation 3.5.3: Colored pixels indicate voxels with significant activation during the stop trials for the original data (left) and head motion data (right)*

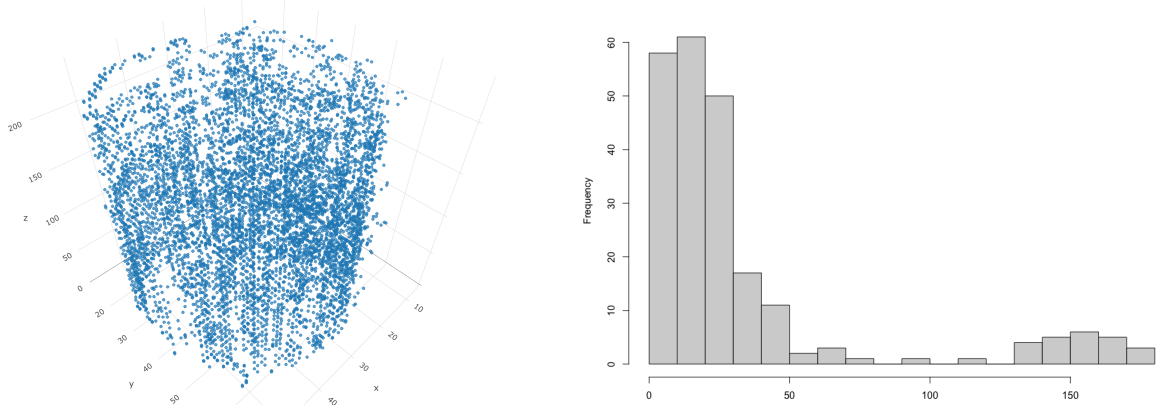


Figure 3.8: *Result of simulation 3.5.3: Left: blue circles indicate voxels identified as singularities in the LLH filter, right: histogram of  $N_{FV}$*

of wavelet coefficients in corresponding filters. We studied aspects of this relationship in Theorem 3.3.1 and in multiple simulations – including some where our algorithm is applied to a real fMRI data set, into which we “spiked” head motions at random points in time. Head motion is a common artifact that has severe consequences on the validity of statistical inferences. Our algorithm can detect the motion-affected time points, but performance is sensitive to a cutoff value – which is currently determined *ad-hoc*. In

the future, a more systematic, data-driven approach (e.g., a statistical test) may further improve our approach. Below, we discuss two additional aspects that also affect the performance of our algorithm; namely, denoising and tuning parameters.

### 3.6.1 Effect of Denoising

Most real life data contain some level of noise, which translates into noisy wavelet coefficients when we take the wavelet transform step. Noisy wavelet coefficients have the effect of creating almost everywhere singularities that need to be taken care of. We exploit the fact that the wavelet transform of a Gaussian white noise is also Gaussian [41], and incorporate a denoising step as follow:

- D1 For each filter and at each scale  $j$ , obtain a robust estimate of the standard deviation  $s$  of wavelet coefficients  $\mathbf{w}$  as:

$$s = \sqrt{\text{med}([w_i - \text{med}(\mathbf{w})]^2)}$$

The underlying assumption here is that most wavelet coefficients are noise, and the actual signal is sparsely represented by a few large coefficients.

- D2 Set to zero those wavelet coefficients that lie three standard deviation away from the mean. This reasoning comes from a classical result that almost 99% of observation generated from a Gaussian distribution lies within three standard deviation from the mean.

The above denoising procedure is not a hard rule. The assumption that noise terms are distributed as independent Gaussian random variables might not hold. In fact, there has been evidence that noise terms in fMRI data exhibit long-range dependence [47]. There are many wavelet-based approaches in existing literature that handle denoising of signals, but which lie beyond the scope of this Chapter. For more information on denoising methods, see for example [43, 48–51]. For our purposes, the above procedure is sufficiently good at denoising the signal. As a point of comparison, Table 3.3 shows the same results as Table 3.1, except that no denoising was performed.

### 3.6.2 Effect of tuning parameters

**Choice of  $\omega_1$ .**  $\omega_1$  is the size of the moving cube used to search for local maxima. Typically, the choice of  $w_1$  is driven by how close to each other the singularities are

filter	$V_{30,30,50}$ [z]	$V_{30,20,20}$ [yz]	$V_{43,43,30}$ [xy]	$V_{10,10,30}$ [xyz]
HLL	0.52 (0.19)	-1.26(0.25)	-1.79 (0.09)	-2.76 (0.09)
LHL	0.38 (0.13)	-1.85 (0.01)	-1.79 (0.09)	-2.77 (0.09)
LLH	-0.85 (0.00)	-1.85 (0.02)	-0.55 (0.14)	-2.76 (0.08)
HHL	-1.19 (0.25)	-1.11 (0.27)	-1.66 (0.09)	-2.61 (0.06)
HLH	0.76 (0.19)	-1.11 (0.17)	-0.44 (0.26)	-2.59 (0.06)
LHH	0.60 (0.17)	-1.70 (0.01)	-0.39 (0.20)	-2.62 (0.06)
HHH	-1.04 (0.21)	-1.09 (0.24)	-0.21 (0.27)	-2.42 (0.14)

Table 3.3: *Without denoising: Lipschitz constant estimation at the locations where signals vary sharply in one, two or three directions (STN = 10).*

expected to be, and by computational burden considerations. Note that computation aside, too small an  $\omega_1$  can also cause issues, since the maxima search described in Section 3.3.1 is a search for local maxima. If  $w_1 = 0$ , every voxel becomes a maximum, and a maxima chain will be formed at every voxel/location, which is not helpful for identifying singularities.

**Choice of  $\omega_2$ .**  $\omega_2$  is the size of the neighborhood used to chain neighboring maxima across scales. The choice of  $\omega_2$  is also driven by how close the singularities are expected to be. Too large a  $w_2$  will bind together maxima that belong to different singularities, and distort Lipschitz constant estimates. If  $\omega_1$  is small, more voxels become maxima, which means  $\omega_2$  also needs to be reasonably small so that different singularities can still be distinguished from each other.

**The maximum scale  $J$ .** The maximum scale  $J$  should be chosen to accommodate the dimension of the data set. If the data dimension is small, at large scale  $J$ , the wavelet transform becomes closer to an approximation/smoothing mechanism. This means a lot of maxima will disappear at large scale, and no chain will be formed.

# Bibliography

- [1] BULLMORE, E., J. FADILI, M. BREAKSPEAR, R. SALVADOR, J. SUCKLING, and M. BRAMMER (2003) “Wavelets and statistical analysis of functional magnetic resonance images of the human brain,” *Statistical Methods in Medical Research*, **12**, pp. 375–399.
- [2] RAMSAY, J. O. and B. SILVERMAN (2007) *Applied functional data analysis: methods and case studies*, Springer.
- [3] KOKOSZKA, P. and M. REIMHERR (2017) *Introduction to Functional Data Analysis*, CRC Press.
- [4] YAO, F. and T. C. M. LEE (2006) “Penalized spline models for functional principal component analysis,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(1), pp. 3–25.
- [5] XIAO, L., V. ZIPUNNIKOV, D. RUPPERT, and C. CRAINICEANU (2016) “Fast covariance estimation for high-dimensional functional data,” *Statistics and Computing*, **26**(1), pp. 409–421.
- [6] XIAO, L., C. LI, W. CHECKLEY, and C. CRAINICEANU (2018) “Fast covariance estimation for sparse functional data,” *Statistics and Computing*, **28**(3), pp. 511–522.
- [7] GOLDSMITH, J., J. BOBB, C. M. CRAINICEANU, B. CAFFO, and D. REICH (2011) “Penalized functional regression,” *Journal of Computational and Graphical Statistics*, **20**(4), pp. 830–851.
- [8] WOOD, S. (2006) “Low-rank scale-invariant tensor product smooths for generalized additive mixed models,” *Biometrics*, **62**, pp. 1025–1036.
- [9] NASON, G. P. (2008) *Wavelet Methods in Statistics with R*, Springer.
- [10] DESCARY, M. H. and V. M. PANARETOS (2019) “Functional data analysis by matrix completion,” *The Annals of Statistics*, **47**(1), pp. 1–38.
- [11] DEMPSTER, A. P., N. M. LAIRD, and D. B. RUBIN (1977) “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **39**, pp. 1–38.

- [12] LIU, C. and D. B. RUBIN (1994) “The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence,” *Biometrika*, **81**, pp. 633–648.
- [13] LOUIS, T. A. (1982) “Finding the observed information matrix when using the EM algorithm,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **44**, pp. 226–233.
- [14] MEILIJSON, I. (1989) “A fast improvement to the EM algorithm on its own terms,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **51**, pp. 127–138.
- [15] MENG, X. and D. B. RUBIN (1993) “Maximum likelihood estimation via the ECM algorithm: A general framework,” *Biometrika*, **80**, pp. 267–278.
- [16] CLAESKENS, G., T. KRIVOBOKOVA, and J. D. OPSOMER (2009) “Asymptotic properties of penalized spline estimators,” *Biometrika*, **96**(3), pp. 529–544.
- [17] EILERS, P. and B. MARX (1996) “Flexible smoothing with B-splines and penalties (with Discussion),” *Statistical Science*, **11**, pp. 89–121.
- [18] EILERS, P., B. MARX, and M. DURBAN (2015) “Twenty years of p-splines,” *SORT*, **39**(2), pp. 149–186.
- [19] HALL, P. and J. D. OPSOMER (2015) “Theory for penalised spline regression,” *Biometrika*, **92**(1), pp. 105–118.
- [20] KAUERMANN, G., T. KRIVOBOKOVA, and L. FAHRMEIR (2009) “Some asymptotic results on generalized penalized spline smoothing,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**(2), pp. 487–503.
- [21] WANG, X., J. SHEN, and D. RUPPERT (2011) “On the asymptotics of penalized spline smoothing,” *Electronic Journal of Statistics*, **5**, pp. 1–17.
- [22] XIAO, L. (2019) “Asymptotic theory of penalized splines,” *Electronic Journal of Statistics*, **13**, pp. 747–794.
- [23] WU, C., C. YANG, H. ZHAO, and J. ZHU (2017) “On the Convergence of the EM Algorithm: A Data-Adaptive Analysis,” ArXiv:1611.00519v2.
- [24] BALAKRISHNAN, S., M. J. WAINWRIGHT, and B. YU (2017) “Statistical guarantees for the EM algorithm: from population to sample-based analysis,” *The Annals of Statistics*, **45**(1), pp. 77–120.
- [25] DE BOOR, C. (1978) *A practical guide to splines*, Springer.
- [26] O’SULLIVAN, F. (1986) “A statistical perspective on ill-posed inverse problems,” *Statistical Science*, **1**(4), pp. 502–518.
- [27] LAZAR, N. (2008) *The statistical analysis of functional MRI data*, Springer.

- [28] POWER, J., K. BARNES, A. SNYDER, B. SCHLAGGAR, and S. PETERSEN (2012) “Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion,” *NeuroImage*, **59**(3), pp. 2142–2154.
- [29] SATTERTHWAITE, T., D. WOLF, J. LOUGHEAD, K. RUPAREL, M. ELLIOTT, H. HAKONARSON, R. GUR, and R. GUR (2012) “Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth,” *NeuroImage*, **60**(1), pp. 623–632.
- [30] VAN DIJK, K., M. SABUNCU, and R. BUCKNER (2012) “The influence of head motion on intrinsic functional connectivity MRI,” *NeuroImage*, **59**(1), pp. 431–438.
- [31] MOWINCKEL, A., T. ESPESETH, and L. WESTLYE (2012) “Network-specific effects of age and in-scanner subject motion: a resting-state fMRI study of 238 healthy adults,” *NeuroImage*, **63**(3), pp. 1364–1373.
- [32] BRIGHT, M. and K. MURPHY (2013) “Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data,” *NeuroImage*, **64**, pp. 526–537.
- [33] SATTERTHWAITE, T., M. ELLIOTT, R. GERRATY, K. RUPAREL, J. LOUGHEAD, M. CALKINS, S. EICKHOFF, H. HAKONARSON, R. GUR, R. GUR, and D. WOLF (2013) “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data,” *NeuroImage*, **64**, pp. 240–256.
- [34] YAN, C., B. CHEUNG, C. KELLY, S. COLCOMBE, R. CRADDOCK, A. DI, Q. LI, X. ZUO, F. CASTELLANOS, and M. MILHAM (2013) “A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics,” *NeuroImage*, **76**, pp. 183–201.
- [35] TYSZKA, J., D. KENNEDY, L. PAUL, and R. ADOLPHS (2014) “Largely typical patterns of resting state functional connectivity in high-functioning adults with autism,” *Cerebral Cortex*, **24**(7), pp. 1894–1905.
- [36] FOX, M., A. SNYDER, J. VINCENT, M. CORBETTA, D. VAN ESSEN, and M. RAICHLE (2005) “The human brain is intrinsically organized into dynamic, anticorrelated functional networks,” *Proceedings of the National Academy of Sciences - PNAS*, **102**(27), pp. 9673–9678.
- [37] FOX, M., M. CORBETTA, A. SNYDER, J. VINCENT, and M. RAICHLE (2006) “Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems,” *Proceedings of the National Academy of Sciences - PNAS*, **103**(26), pp. 10046–10051.
- [38] FOX, M., D. ZHANG, A. SNYDER, and M. RAICHLE (2009) “The Global Signal and Observed Anticorrelated Resting State Brain Networks,” *Journal of Neurophysiology*, **101**(6), pp. 3270–3283.

- [39] WEISSENBACHER, A., C. KASESS, F. GERSTL, R. LANZENBERGER, E. MOSER, and C. WINDISCHBERGER (2009) “Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies,” *NeuroImage*, **47**(4), pp. 1408–1416.
- [40] PATEL, A., P. KUNDU, M. RUBINOV, P. JONES, P. VÉRTES, K. ERSCHKE, J. SUCKLING, and E. BULLMORE (2014) “A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series,” *NeuroImage*, **95**, pp. 287–304.
- [41] MALLAT, S. and W. HWANG (1992) “Singularity detection and processing with wavelets,” *IEEE transactions on information theory*, **38**(2), pp. 617–643.
- [42] MALLAT, S. (1989) “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,” *IEEE transactions on pattern analysis and machine intelligence*, **11**(7), pp. 674–693.
- [43] ——— (2009) *A wavelet tour of signal processing: the sparse way*, Elsevier / Academic Press.
- [44] PERCIVAL, D. (2000) *Wavelet methods for time series analysis*, Cambridge University Press.
- [45] SNOEK, L., M. VAN DER MIESON, T. BEEMSTERBOER, A. VAN DER LEIJ, A. EIGENHUIS, and H. SCHOLTE (2021) “The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses,” *Scientific Data*, **8**(1), p. 85.
- [46] JAHFARI, S., L. WALDORP, K. R. RIDDERINKHOF, and H. S. SCHOLTE (2015) “Visual information shapes the dynamics of corticobasal ganglia pathways during response selection and inhibition,” *Journal of cognitive neuroscience*, **27**(7), pp. 1344–1359.
- [47] FADILI, M. and E. BULLMORE (2002) “Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors,” *NeuroImage*, **15**, pp. 217–232.
- [48] DONOHO, D. (1995) “De-noising by soft-thresholding,” *IEEE transactions on information theory*, **41**(3), pp. 613–627.
- [49] DONOHO, D. and I. JOHNSTONE (1995) “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association*, **90**(432), pp. 1200–1224.
- [50] PIZURICA, A. and W. PHILIPS (2003) “A versatile wavelet domain noise filtration technique for medical imaging,” *IEEE transactions on medical imaging*, **22**(3), pp. 323–331.



- [51] JOHNSTONE, I. and S. B.W. (1997) “Wavelet threshold estimators for data with correlated noise,” *Journal of the Royal Statistical Society. Series B, Statistical methodology*, **59**(2), pp. 319–351.
- [52] ANDERSON, C., S. LOWEN, and P. RENSHAW (2006) “Emotional task-dependent low-frequency fluctuations and methylphenidate: Wavelet scaling analysis of 1/f-type fluctuations in fMRI of the cerebellar vermis,” *Journal of neuroscience methods*, **151**(1), pp. 52–61.
- [53] CIUCIU, P., P. ABRY, and B. HE (2014) “Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks,” *NeuroImage*, **95**, pp. 248–263.
- [54] RISK, B., D. MATTESON, R. SPRENG, and D. RUPPERT (2016) “Spatiotemporal mixed modeling of multi-subject task fMRI via method of moments,” *NeuroImage*, **142**, pp. 280–292.
- [55] COSTAFREDA, S., G. BARKER, and M. BRAMMER (2009) “Bayesian wavelet-based analysis of functional magnetic resonance time series,” *Magnetic Resonance Imaging*, **27**, pp. 460–469.
- [56] VARIN, C., N. REID, and D. FIRTH (2011) “An overview of composite likelihood methods,” *Statistical Sinica*, **21**, pp. 5–42.
- [57] TAK, H., K. YOU, S. GHOSH, B. SU, and J. KELLY (2020) “Data transforming augmentation for heteroscedastic models,” *Journal of Computational and Graphical Statistics*, **29**(3), pp. 659–667.
- [58] BERTSEKAS, D. P. (1995) *Nonlinear Programming*, Athena Scientific.
- [59] POLLARD, D. (1984) *Convergence of stochastic processes*, Springer-Verlag.
- [60] VERSHYNIN, R. (2018) *High-dimensional probability*, Cambridge University Press.