Localizing filamentary structures in 2D point clouds using beamlet-based approximation networks

Rafael Reisenhofer[1], Gitta Kutyniok[2], Jakob Lemvig[3]

[1]Universität Osnabrück, [2]TU Berlin, [3] Technical University of Denmark

UNIVERSITÄT OSNABRÜCK

Fluorescence Photoactivation Localization Microscopy (FPALM) [4]

In general, due to diffraction and intereference, the best resolution achievable by methods of light microscopy is bounded by the so called Abbe-limit at roughly 200 nm. This is a major issue in fields of research where other methods of microscopy are not applicable, like the research of living cells. In fluorescence photoactivation localization microscopy (FPALM) this limit is surpassed by the successive activation of single fluorophores (fluorescent molecules that are bonded to macromolecules), so the emissions of each fluorophore can be recorded without interference from others. Using this technique, the location of each fluorophore can be determined by computing the center of its corresponding recorded emission (point spread function).

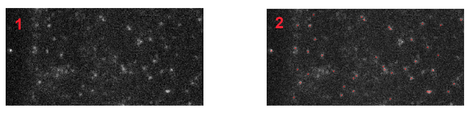

Figure 1: The recordings of single point spread functions in (1) are used to localize the corresponding molecules. The computed locations are represented by red dots in (2).

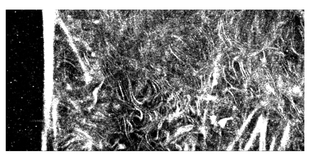

Figure 2: FPALM recording of the actin cytoskeleton of a HeLa cell. This picture visualizes the locations of single molecules which were determined by the analysis of 15000 single recordings.

Despite a localization precision of up to 30 nm, FPALM recordings of living cells suffer from two drawbacks:

- In order to create a complete FPALM recording, a living cell has to be observed for a certain period of time. Due to the ongoing dynamics within a living cell, this causes the gain of spatial information to be traded against an uncertainty regarding time.
- If the time of observation is limited, the maximum density of localized molecules is severly bounded.

This begs the question:

Can filamentary structures be reconstructed from a sparse set of localized positions by means of a mathematical analysis?

The detection of Hölder-functions in 2D point clouds with uniformly distributed random noise [1, 2]

Model

Let $X_1, \dots, X_n \in [0,1]^2$ be a point cloud and $f \in \mathcal{H}(\alpha, \beta)$ a function within the Hölder-class $\mathcal{H}(\alpha, \beta)$, that is f fulfills both of the following conditions:

$$\left|f^{(\lfloor\alpha\rfloor)}(x)\right| \leq \beta \qquad \forall x \in [0,1] \tag{1}$$

$$\left|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)\right| \leq \beta\,|x-y|^{\alpha-\lfloor\alpha\rfloor} \qquad \forall x, y \in [0,1] \tag{2}$$

Under the assumption of uniformly distributed random noise, the question whether a subset of the point cloud $X_1, \dots, X_n \in [0,1]^2$ is caused by a function $f \in \mathcal{H}(\alpha, \beta)$ can be answered by the the analysis of the following hypotheses:

$$H_0 : X_1, \dots, X_n \sim^{\text{u.i.v.}} \mathcal{U}([0,1]^2) \tag{3}$$

$$H_1 : X_1, \dots, X_n \sim^{\text{u.i.v.}} (1-\epsilon)\cdot\mathcal{U}([0,1]^2) + \epsilon\cdot\mathcal{U}(\text{graph}_\eta(f)) \tag{4}$$

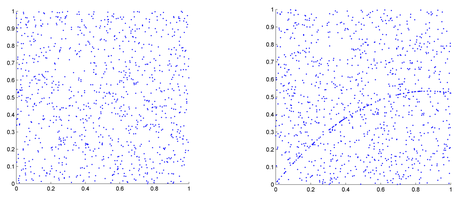

Figure 3: Examples of point clouds in agreement with H_0 (left) resp. H_1 (right)

Piecewise polynomial approximation of Hölder-functions

A function $f \in \mathcal{H}(\alpha, \beta)$ can be approximated in the neighborhood of an expansion point $x_0 \in [0,1]$ using the Taylor series:

$$T_f^{x_0}(x) = \sum_{i=0}^{\lfloor\alpha\rfloor} f^{(i)}(x_0)\frac{(x-x_0)^i}{i!} \tag{5}$$

Motivated by the locality of the approximation provided by (5), the interval $[0,1]$ is divided into several subintervals on which a family of polynomials is defined, such that for each function f within a class $\mathcal{H}(\alpha, \beta)$ and for each subinterval one best fitting approximation polynomial can be identified.

Definition (Approximation Polynomial)**.** *Let $\Delta \in (0,1]$ be the length of one subinterval of the unit interval, $m \in \{1, \dots, \Delta^{-1}\}$, $\delta \in \mathbb{R}^+$, $\delta_{i+1} = \Delta^{-i}\delta$ and $\mathbf{h} \in \mathbb{Z}^{\lfloor\alpha\rfloor+1}$ where $|h_{i+1}\delta_{i+1}| \leq \beta$ and $i \in \{0, \dots, \lfloor\alpha\rfloor\}$. The approximation polynomial $p_{m,\mathbf{h}}$ is then definied by:*

$$p_{m,\mathbf{h}}(x) = \sum_{i=0}^{\lfloor\alpha\rfloor} h_{i+1}\delta_{i+1}\frac{\left(x - \Delta\left(m - \frac{1}{2}\right)\right)^i}{i!} \tag{6}$$

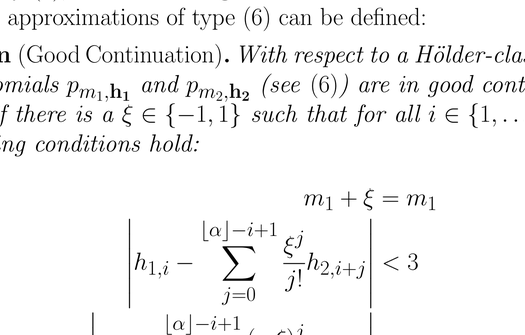

Figure 4: Two approximation polynomials with parameters $\Delta_{\text{left}} = \frac{1}{8}$, $\delta_{\text{left}} = 0.01$, $m_{\text{left}} = 3$, $\mathbf{h}_{\text{left}} = (35, -10)$ and $\Delta_{\text{right}} = \frac{1}{4}$, $\delta_{\text{right}} = 0.01$, $m_{\text{right}} = 1$, $\mathbf{h}_{\text{right}} = (15, 16, -13)$

A principle of 'good continuation' can be used to construct efficient approximation networks

Concatenations of the polynomials defined in (6) can be used to build Taylor-based approximations of arbitrary Hölder-functions on the unit interval. The question reamains, however, which polynomials are suited for concatenation, i.e. which polynomials would - if combined - form a sensible approximation of a function f within a Hölder-class $\mathcal{H}(\alpha, \beta)$. This question can be answered by considering the following estimation of the Taylor coefficients of a function $f \in \mathcal{H}(\alpha, \beta)$ at an arbitrary expansion point $x + \Delta$ given by the coefficients of the same function at expansion point x:

$$\left|f^{(i)}(x+\Delta) - \sum_{j=0}^{\lfloor\alpha\rfloor-i} f^{(i+j)}(x)\frac{\Delta^j}{j!}\right| \leq \frac{\beta}{\lfloor\alpha - i\rfloor!}\Delta^{\alpha-i} \tag{7}$$

Motivated by (7), the notion of 'good continuation' between two piecewise polynomial approximations of type (6) can be defined:

Definition (Good Continuation)**.** *With respect to a Hölder-class $\mathcal{H}(\alpha, \beta)$, two polynomials $p_{m_1,\mathbf{h_1}}$ and $p_{m_2,\mathbf{h_2}}$ (see (6)) are in good continuation if and only if there is a $\xi \in \{-1, 1\}$ such that for all $i \in \{1, \dots, \lfloor\alpha\rfloor + 1\}$ the following conditions hold:*

$$m_1 + \xi = m_1 \tag{8}$$

$$\left|h_{1,i} - \sum_{j=0}^{\lfloor\alpha\rfloor-i+1} \frac{\xi^j}{j!}h_{2,i+j}\right| < 3 \tag{9}$$

$$\left|h_{2,i} - \sum_{j=0}^{\lfloor\alpha\rfloor-i+1} \frac{(-\xi)^j}{j!}h_{1,i+j}\right| < 3 \tag{10}$$

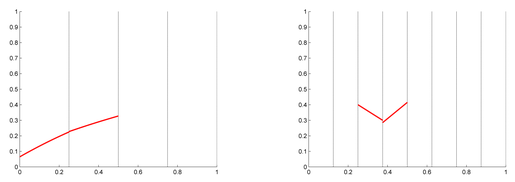

Figure 5: While the left picture shows two polynomials in good continuation, the two polynomials in the right picture are not in good continuation due to their substantial difference in slope.

Hölder functions in point clouds can be detected by computing the longest path of significant approximation polynomials in good continuation

Given a Hölder-class $\mathcal{H}(\alpha, \beta)$ and parameters $\Delta, \delta \in [0,1]$, each of the approximation polynomials $p_{m,\mathbf{h}}$ defined in (6) can be associated with a surrounding region $R \subset [0,1]^2$ given by:

$$R(p_{m,\mathbf{h}}) = \{((x,y) \in [\Delta(m-1), \Delta m] \times [0,1] : \left|y - p_{m,\mathbf{h}}(x) \leq c\delta\right|\} \tag{11}$$

with $c = \sum_{i=0}^{\lfloor\alpha\rfloor} \frac{1}{2^i \cdot i!}$.

Given a point cloud $X_1, \dots, X_n$, let $N(p_{m,\mathbf{h}}) \in \mathbb{N}$ be the number of points within the region $R(p_{m,\mathbf{h}})$, i.e.:

$$N(p_{m,\mathbf{h}}) = \#\{i \in \{1, \dots, n\} : X_i \in R(p_{m,\mathbf{h}})\}$$

After the computation of all values $N(p_{m,\mathbf{h}})$, a threshold $\tau_1 \in \mathbb{N}$ can be used to identify polynomials with a significant amount of points in their proximity. Using only significant polynomials, the notion of the longest significant path can be defined:

Definition (Longest Significant Path)**.** *The longest path of significant polynomials (i.e. polynomials with $N(p_{m,\mathbf{h}}) \geq \tau_1$) that can be constructed using the principle of good continuation (see (8)) is called the 'Longest Significant Path' (LSP).*

By defining a second threshold $\tau_2 \in \mathbb{N}$, the LSP-test can be used to decide between H_0 and H_1.

Definition (LSP-Test)**.** *Let $X_1, \dots, X_n \in [0,1]^2$ be a point cloud on the unit interval, $\mathcal{H}(\alpha, \beta)$ a Hölder class and $\tau_1 \in \mathbb{N}$ the threshold for significant polynomials. Let π denote the longest significant path of corresponding approximation polynomials, then H_1 is chosen if and only if the number of polynomials within π is equal or greater than a second threshold τ_2. That is $\#\pi \geq \tau_2$.*

It has been proven by Arias-Castro et al. that choices for τ_1 and τ_2 can be computed, such that the probability of choosing the correct hypothesis using the LSP-test converges to 1 when the amount of recorded data approaches infinity.

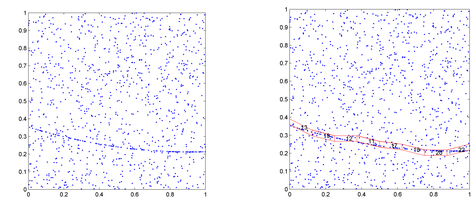

Figure 6: A Hölder-function is successfully detected by computing the longest significant path.

Beamlet-based Filament Tracer (BFiT)

Beamlets

Instead of the approximation polynomials defined in (6), we use beamlets as the elementary building blocks of our approximation network. Beamlets are linear structures that were first introduced by Donoho and Huo [3] and can be described as line segments that connect points lying on a dyadical partition of the 2-dimensional plane. The usage of beamlets has already been considered by Arias-Castro et al. [2] and has mainly two advantages:

- Beamlets make it easier to model structures with high slopes.
- It allows the detection filaments that cannot be modeled as functions of the x-axis.

While in the case of (6), the resolution of the approximation network is controlled by parameters $\Delta, \delta \in \mathbb{R}$, the beamlet-based network is defined by parameters $j, J \in \mathbb{N}$ where $\Delta = 2^{-j}$ and J is a measure of the density, that is $\delta = 2^{j-J}$.

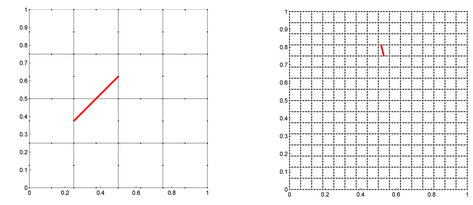

Figure 7: The beamlet in the left picture is part of an approximation network with parameters $j = 2$ and $J = 5$. On the right side, a beamlet is shown that belongs to a network with $j = 4$ and $J = 10$.

Processing of point clouds with BFiT

The beamlet-based filament tracer (BFiT) not only localizes filamentary structures within point clouds. It also renders a picture, using anisotropic gaussian smoothing along significant paths to visualize detected structures. Generally, the following steps lead from a given 2-dimensional point cloud to a rendered picture:

1. Definition of parameters (j,J,τ_1,τ_2,...)
2. Determination of significant beamlets
3. Determination of significant paths
4. Postprocessing and rendering

In order to complete step 2 (determination of significant beamlets) it is necessary to count for each beamlet the number of points within its region. The worst case complexity of this task is $\mathcal{O}(n^2)$, making it the computational bottleneck of the overall procedure. Under the assumption that filaments propagate either along the x-axis or along the y-axis, the determination of significant paths can be done in linear time by dynamic programming.

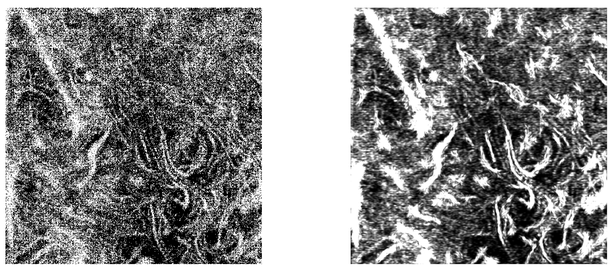

Figure 8: Visualization of a FPALM-recording using BFiT. The analyzed data is a subset of the point cloud shown in figure 2.

References

[1] E. Arias-Castro, D. L. Donoho, and X. Huo. Adaptive multiscale detection of filamentary structures in a background of uniform random points. *The Annals of Statistics*, 34(1):326–349, 2006.

[2] E. Arias-Castro, B. Efros, and O. Levi. Networks of polynomial pieces with application to the analysis of point clouds an images. *Journal of Approximation Theory*, 162:94–130, 2010.

[3] D. L. Donoho and X. Huo. *Multiscale and Multiresolution Methods*, chapter Beamlets and multiscale image analysis, pages 149–196. Springer Lecture Notes in Computational Science and Engineering, 2002.

[4] S. T. Hess, T. J. Gould, M. Gunewardene, J. Bewersdorf, and M. D. Mason. Ultra-high resolution imaging of biomolecules by fluorescence photoactivation localization microscopy (fpalm). 2011. in press.