A proposal to optimise the penalty parameter in sparse deconvolution super-resolution fluorescence imaging

By: Mohamad Ahmad

Supervised by: Cyril Ruckebusch, Siewert Hugelier

Examined by: Johan Westerhuis, TBD

02 July 2018

Villeneuve d’Ascq, France
Preface

Let me start it off, by thanking my supervisors, Cyril Ruckebusch and Siewert Hugelier for giving me the opportunity to work here and the assistance that I received during my project. I would also like the thank Raffaele Vitale, as he helped me through a couple of fires as well. I enjoyed working with everybody at the office, especially with Lucas Uriarte, Dario Cevoli, Do Mai Trang and Siewert (again). I enjoyed my stay in Lille immensely and learned a lot on both the educational as on the social side of things. I hope to work with these amazing people again in the future.
Summary
A heuristic approach has been developed for the Sparse Image Deconvolution and Reconstruction algorithm. Whereby the approach balance between the reproduction of the original signal and the amount of points utilised to reproduce the signal. The approach has been rigorously validated through localisation performance and robustness testing. The limits of the approach have been determined, achieving a maximum density of 8 emitters per um² and robustness testing showed that the approach is capable of handling minor perturbation of the data. The approach was applied on a real dataset that analysed mitochondria in a live cell. The mitochondria showed movement during the analyses and photo-bleaching occurred. Both phenomenon were captured by SPIDER. The super-resolution images obtained showed a resolution of 61 and 110 nm, while the limit of the imaging system was set to be 267 nm.
Table of Content

Preface ................................................................................................................................. 1
Summary ................................................................................................................................. 2

1 Introduction ......................................................................................................................... 5
  1.1 Super-resolution fluorescence microscopy [1] ................................................................. 5
     1.1.1 Microscopy .................................................................................................................. 5
     1.1.2 Fluorescence ............................................................................................................... 5
     1.1.3 Wide-field epi-fluorescence microscopy ....................................................................... 6
     1.1.4 Super-resolution microscopy ....................................................................................... 7
  1.2 Sparse Image Deconvolution and Reconstruction ............................................................ 10
     1.2.1 Image Deconvolution ................................................................................................. 10
     1.2.2 SPIDER ....................................................................................................................... 10
     1.2.3 Penalty parameter ....................................................................................................... 10
  1.3 Project ............................................................................................................................. 11
     1.3.1 Aim ............................................................................................................................... 11
     1.3.2 Approach ...................................................................................................................... 11

2 Theoretical Background ....................................................................................................... 11
  2.1 Penalised Least squares regression: .................................................................................. 11
  2.2 Various norms ................................................................................................................... 12
  2.3 SPIDER ............................................................................................................................ 13
  2.4 Weight optimisation: .......................................................................................................... 14
  2.5 Sum of the Normalised Terms .......................................................................................... 14

3 Methods .............................................................................................................................. 17
  3.1 Tools & Datasets ............................................................................................................... 17
    3.1.1 Simulation .................................................................................................................... 17
    3.1.2 Real .............................................................................................................................. 19
    3.1.3 Performance tools ....................................................................................................... 19
    3.1.4 The Bootstrap ............................................................................................................. 21
  3.2 Approach ........................................................................................................................ 23
    3.2.1 Localisation performance ............................................................................................ 23
    3.2.2 Structure determination ............................................................................................. 24

4 Results & Discussion .......................................................................................................... 24
  4.1 Localisation performance ................................................................................................. 24
    4.1.1 Densities ...................................................................................................................... 24
4.1.2  Background ................................................................................................. 27
4.1.3  Robustness testing ..................................................................................... 31
4.2   Structure determination ................................................................................ 36
   4.2.1  Simulated .................................................................................................. 37
   4.2.2  Real........................................................................................................... 38
5   Conclusion ........................................................................................................ 41
6   Future perspectives ............................................................................................. 41
7   References ......................................................................................................... 42
8   Appendix ........................................................................................................... 44
8.1  Comparative study on: Λ-optimisation approaches ....................................... 44
   8.1.1  L2-norm .................................................................................................. 44
   8.1.2  L1-norm .................................................................................................. 45
   8.1.3  L0-norm .................................................................................................. 46
1 Introduction

1.1 Super-resolution fluorescence microscopy [1]

1.1.1 Microscopy
The microscope is an essential tool in science. It gives an in-depth look into structures that the human eye would not be able to observe, by magnifying the object of interest. Microscopy has changed drastically and microscopes now produce datasets that are generated by a CCD camera. Resolution has increased dramatically, resolution being the ability to distinguish two objects from one another. The details observed with the first microscopes would be a few hundred micrometres (µm) where now a few hundred nanometres (nm) can be easily reached. To put it in perspective and give an example, that would mean that the investigation has gone from observing a single bacterium to multiple mitochondria in a cell. There are many modalities in the field of microscopy, however to keep it within the scope of the project, the focus will be on wide-field fluorescence microscopy.

1.1.2 Fluorescence
Fluorescence is the phenomenon that occurs when a molecule absorbs a photon at a specific wavelength, putting the molecule in an excited state from which it will go down to either the ground state or to a triple state (from the triple state, phosphorescence can occur). This transition from the excited state to the ground state has multiple pathways. The non-radiative pathway corresponds to any pathway where no light is emitted. This can be e.g. due to quenching by another nearby molecule (such as singlet oxygen). Quenching is the occurrence of energy transfer from one molecule to another, putting the former to the ground state. On the other hand, for the radiative pathway, a photon is emitted by the excited molecule. The energy of the emitted photon is in general less than the absorbed photon, this difference in energy is called the “Stokes Shift”. This is due to the excited molecule dropping to the lowest vibrational excited state (“Kasha’s Rule”). Due to this difference in energy between excitation and emission, detection at a different wavelength can be utilised. This effect is a huge advantage that is exploited by fluorescence spectroscopy. In figure 1 a Jablonksi Energy Diagram is shown. This diagram shows the process of the various pathways, mentioned previously.

![Jablonski Energy Diagram](image.png)

Figure 1: Jablonksi Diagram, taken from [2]. Shows the various pathways of excitation and emission.
Another important factor to consider is that the emission happens uniformly in every direction, regardless of the path the incoming light took, as such, the detector is in most cases, not behind the sample, but either perpendicular to the incoming light or behind the coming light.

Kasha’s rule: Emission only occurs from the lowest excited state, meaning that if the molecule of interest is excited to a higher vibrational (and/or excited) state, it will first fall back to the lowest vibrational (and excited) state before emitting a photon, this is called internal conversion. This is due to the speed of this process compared to fluorescence, as internal conversion is a magnitude of at least 3 times faster. Thus the emitted photon will have a lower energy than the absorbed photon, due to this decrease in energy of the excited state.

1.1.3 Wide-field epi-fluorescence microscopy

To keep within the scope of the project, the focus will be on wide-field epi-fluorescence microscopy. In this particular approach (see figure 2), the whole sample (wide-field) is illuminated. Taking advantage of the “Stokes Shift”, a dichroic mirror placed between the sample and the detector, which allows the excitation light to take the same path as the emission light (in reverse). A dichroic mirror reflects certain wavelengths of light, while letting the other wavelengths through.

![Figure 2: Set-up of an epi-fluorescence microscope, taken from [3].](image)

Behind the dichroic mirror an emission filter can be positioned to further specify the detected light and exclude undesirable wavelengths. The detector is in most cases a Charge-Coupled Device (CCD), which, to keep it short, is a camera that digitises the received signal to a grey-scale pixel-image. Each pixel on the camera detects the photons emitted from a particular area of the sample. A pixel has, within the framework of this project, an area of around 100 by 100 nm (this will be specified if otherwise). This is smaller than the limit imposed by the diffraction limit (250 nm), which will be discussed in Chapter 1.1.4.1. The wavelengths of excitation are usually in the region of visible light (400 to 800 nm), although somewhat closer to the blue side of the spectrum. Emission is within the same region, however, at a longer wavelength than excitation.
1.1.4 Super-resolution microscopy

1.1.4.1 The Abbe limit
Ernst Abbe discovered in 1873 [4] that light with a certain wavelength $\lambda$, beamed through a certain medium which has a refractive index $\eta$, while converging to a point with a half-angle $\theta$ will produce a spot with a certain lateral resolution $d_{x,y}$ (Eq. 1). Abbe found that the limit of discerning two objects from one another depended on the wavelength of light and not on the quality of the optics; this was coined as “The Abbe Limit”. It was also inversely dependent on the Numerical Aperture (NA, Eq.2), which is defined by the refractive index and half-angle of incoming light, giving a dimensionless number that defines the range of angles that a particular system can accept.

$$d_{x,y} = \frac{\lambda}{2NA} \quad \text{(Eq. 1)}$$

$$NA = \eta \sin \theta \quad \text{(Eq. 2)}$$

Due to diffraction, a singular emission point is thus detected as a spot when observed through an imaging system. This spot is called the Point Spread Function (PSF), which has a size $d$. A model of the PSF is a two-dimensional Gaussian shape. It can be seen as a convolved distribution from a singular point (figure 3). Discerning two singular points with a certain PSF requires that they are separated by their width.

From this Lord Rayleigh developed a new formula to better represent wide-field microscopy, this was dubbed “Rayleigh’s Criterion” (Eq. 3) [5], which defined the minimum resolution of discerning two objects from one another. This minimum distance is shown in figure 3, in both one and two-dimensional space.

$$d_{Rayleigh} = \frac{0.61\lambda}{NA} \quad \text{(Eq. 3)}$$

![Rayleigh Criterion](image)

*Figure 3: One- and two-dimensional representation of a PSF and Rayleigh’s Criterion*
Within the framework of wide-field fluorescence microscopy, a resolution of around 200 nm can be determined to be the limit. This is somewhat the resolution limit of fluorescence microscopy, as lower excitation wavelengths than around 400 nm are not used. Working with such a limit means that e.g. mitochondria can be analysed as their sizes are up to around 1 μm. However this forbids minute analysis of the object, and the underlining structure cannot be finely determined due to this limit.

### 1.1.4.2 Nobel Prize 2014

In 2014, the Nobel Prize in Chemistry was awarded to Eric Betzig, Stefan W. Hell and William E. Mourner: “for the development of super-resolved fluorescence microscopy” [6]. Super-resolved (or super-resolution) microscopy is defined as breaking through the wall in optical microscopy (“The Abbe Limit”), meaning that a sub-diffraction limited resolution is achieved (< 200 nm). One of the proposed methods are based on the on/off dynamics of fluorescent probes (fluorophores). Certain fluorophores can have the ability to switch from an emitting state (on) to a dark non-emitting state (off). This can be done through various ways, very often by using different laser illuminations. Dronpa [7] is an example of a photo-switchable protein. Illumination of this protein by a particular wavelength of light (405 nm) makes it switch from a protonated form, which is near non-fluorescent, to a de-protonated form, which is fluorescent. An illustration of the conversion process is shown in figure 4. Other than an “on-” and “off-state”, the protein will also be able to reach a “dark-state”, which is irreversible. As is seen with all fluorescent molecules, photo-bleaching occurs, such that the signal is irreversibly lost. However the point is to work with proteins that are resistant to fatigue and do not bleach too quickly.

![Diagram of fluorophore states](image)

**Figure 4:** Illustration on the conversion process of a fluorophore

Wide-field super-resolved single-fluorophore microscopy encompasses a range of techniques (PALM, STORM and PAINT). All these techniques localise individual points from the convolved signal that is detected by the imaging system (single emitter to PSF). These points are localised by detecting the PSF, from which a point of origin can be determined, this coincides with the maximum of the PSF, in both the one and two dimensional shapes (see figure 5).

![PSF representation](image)

**Figure 5:** One- and two dimensional representation of a fluorophore detected by an imaging system.
However, for the direct determination of the position of a single emitter, no overlap can occur between emitters, as determined by Rayleigh’s Criterion. This limits the applicability of the technique to very low densities of emitters per spatial unit.

The “blinking” property of photo-switchable proteins allow this problem to be ignored. Blinking results from the on/off properties of the molecule and is the process of emitting light with certain intervals of time, stochastically. As such, the imaging system will not detect all the fluorophores simultaneously. An example is shown in figure x, where 2 fluorophores (a / b) are overlapped with a distance \( d \) \(< d_{Rayleigh} \) between them. However, due to this blinking property, they can be localised independently from one another, at different times (t). After localisation, the solutions are summed to get a final results.

![Figure 6: Artistic illustration of the blinking property.](Image)

Different approaches to generating this blinking property have been developed. Points Accumulation for Imaging in Nanoscale Topography (PAINT) generates its blinking by turning certain molecules on through reactions with fluorescent tags. It considers the shift of the emission spectrum of certain tags after binding to specific molecules. We will focus below on the approaches generating this blinking using photo-switchable fluorophores, namely Photo-activated localisation Microscopy (PALM) [8] and Stochastic Optical Reconstruction Microscopy (STORM) [9].

1.1.4.2.1 PALM
PALM activates within a small region of the sample a sparse number of fluorophores. After activation, the fluorophores are illuminated with a wavelength of light that make the molecules fluoresce. This illumination is applied until the activated fluorophores are completely dark (photo-bleached). From here another set of fluorophores is activated and these steps will continue on until the whole sample has been analysed. The solutions are summed up and this results in a singular image that visually represents the structure of the sample. The technique is quite ingenious in the fact that the already analysed fluorophores will not be able to interfere with the measurements that still have to be taken. However the caveat is that only a sparse number of fluorophores can be activated at one point in time, making this not applicable in high density situations.

1.1.4.2.2 STORM
STORM is simpler PALM on the practical side, as over the whole sample fluorophores are stochastically fluorescent, and frames are taken over a specific timeframe, with each frame having different fluorophores illuminated. The fluorophores are localised in each frame independently by means of a post-processing algorithm. The sum of solutions across all frames is taken, to visualise the distribution of the single emitters. The downside to this method is the same as PALM, meaning that the approach is not applicable in high density situations.
1.2 Sparse Image Deconvolution and Reconstruction

In this study, the super resolution approach to single emitter localisation is Sparse Image Deconvolution and Reconstruction (SPIDER) [10], which is an image deconvolution algorithm. This algorithm will be the main subject of this project.

1.2.1 Image Deconvolution

Image deconvolution is defined as removing the optical distortion that occurs when an image of a sample is taken. The distortion being the difference between the actual object and what is shown in the image, this is called the PSF. Although this is one of the main distortions that occur in the imaging process, it is not the only one. One also has to consider the interferences of other compounds in the sample, such as auto-fluorescence, which is the fluorescence that occurs in biological samples [11]. Added to this is the fluorescence of molecules that are outside of the focal plane of the imaging system. The unwanted and unfocused fluorescence and PSF are the distortions happening within the optics and sample, adding to this are the distortions that occur due to the digitisation of the signal. As with most electronic acquisition, there is a background generated which can be due to various factors, such as Poisson and Gaussian noise, which are due to the digitisation of the signal and dark current, respectively. To briefly summarise, with image deconvolution one must consider three things: the convolution function (PSF), the solution (the emitters) and the untargeted/unwanted information, such as noise and auto-fluorescence.

1.2.2 SPIDER

The SPIDER algorithm “separates” these three objects from one another, by applying an iterative algorithm that tries to reproduce the image that is generated by the imaging system with a certain convolution matrix $C$, optimizing coefficients $x$, while discarding the unwanted/untargeted information $E$. In other words, the algorithm will try to fit a particular image with a sum of pre-determined convolved shapes. As such, from the algorithm three pieces of information can be considered: the convolution matrix, which is pre-determined; the positioning of the objects that have the aforementioned convolved structure; the discarded information, see figure 7 for an artistic representation.

![Figure 7: Artistic representation of the deconvolution framework.](image)

SPIDER applies penalised least squares regression (PLSR) with an $L_0$-norm to achieve this. The $L_0$-norm restricts the number of coefficients in $x$, regardless of their individual intensities. The $L_0$-norm tries to minimise the reproduction error utilising the least amount of data points. To say it concisely, SPIDER attempts to fit the image with the convolution matrix utilising the minimum number of objects.

1.2.3 Penalty parameter

The caveat of PLSR is that a penalty parameter has to be chosen, which differs depending on the data that are utilised. The penalisation of the $L_0$-norm is on the number of coefficients that are
utilised to reproduce the image. The penalty parameter defines the severity of this penalisation, meaning that a higher value for the penalty parameter translates to utilising fewer coefficients to reproduce the data. In the framework of image deconvolution, the correct value for this penalty parameter translates to utilising the same number of coefficients as there are emitters. This however cannot be done without a priori information and observing the solutions that are generated by applying a particular value for the penalty parameter. The approach to utilising SPIDER as of now is filling in an arbitrary value for the penalty parameter and reviewing the solution. Expertise on SPIDER or a priori information on the sample is necessary to properly review the solutions.

1.3 Project

1.3.1 Aim
The aim of the project is to develop an approach that can optimise the previously mentioned penalty parameter to removing the need for an expert or a priori information to review the generated solutions, while also giving a more objective view on the matter. Furthermore, the optimisation approach has to be rigorously tested to determine its validity and has be applied in real situations.

1.3.2 Approach
Within this framework one could look towards the $L_1$- and $L_2$-norms, to understand the underlining reasoning behind the particular approaches and see if it could be translated to the $L_0$-norm. To test the approach, a point of reference has to been made, as to what would dictate a good result while also defining the limits of the approach. As such, a method for validation has to be developed.

How the approach was developed is explained in Chapter 2, while the method to rigorous testing will be explained in Chapter 3. The results of the tests will be shown and discussed in Chapter 4, and in Chapter 5 a conclusion will be made from the results.

2 Theoretical Background

2.1 Penalised Least squares regression:
Regression is one of the oldest and most well-known approaches for defining a relationship between multiple variables [12]. This approach, together with the least squares method, will try to model the relationship between the variables, whereby it will minimize the sum of the squared differences between the model and data.

Although this approach for regression is used in many areas of science, it does have its shortcomings, such as the possibility of overfitting and the fact that the problem one tries to resolve with regression might be ill-posed. An ill-posed problem is defined as not fulling the criteria for being well-posed [14], which is defined as: (1) a solution exists, (2) the solution is unique and (3) the solution is continuous with the data. The meaning behind the first two criteria is apparent. However with the third point, an approach fulfils this criterion if with a slight change in the data, no large change is seen in the solution. The least squares approach does fulfil the first criterion, for the second and third, this is not the case. To alleviate this ‘problem’ and broaden the capabilities of the approach, a variation of the least squares approach can be utilised. Such a variation is penalised least squares regression, where one applies a regularisation parameter on the least squares minimisation. This regularisation parameter is
a weighted penalisation on the coefficients of the model, added to the least squares minimisation function.

\[
\min \|y - Cx\|_2^2 + \lambda\|x\|_{L^q}
\]  
(Eq. 4)

Equation 4 shows the general minimisation function that is applied in penalised least squares regression, whereby \(y\) represents the data, \(x\) the coefficients, \(C\) the convolution matrix, \(L^q\) the norm that is applied to the coefficients and \(\lambda\) the weight of the penalisation. The model that is generated is formulated as \(Cx\). In the function, the \(\|x\|_{L^q}\) term represents the “size” of the coefficients, while the \(\|y - Cx\|_2^2\) term represents the similarity of the model relative to the original data. In addition, the parameter \(\lambda\) applies the weight to the penalisation, defining the balance between these two terms. The \(\|y - Cx\|_2^2\) term is a squared two-norm, while with the \(\|x\|_{L^q}\) term the \(q\)-norm is applied. PLSR has many forms, which are defined by a particular value for \(q\) that can have any real positive number (or 0).

### 2.2 Various norms

The most known norms of PLSR are the \(L^2\)-, \(L^1\)- and \(L^0\)-norm, which have a value of \(q = 2, 1\) and \(0\), respectively. The various norms have different capabilities, e.g. the \(L^2\)-norm can be utilised in smoothing [14], the \(L^1\)-norm, in regression shrinking [15] and the \(L^0\)-norm in deconvolution. There are some other forms as well, such as Elastic-net [16], which combines the \(L^2\)- and \(L^1\)-norm, adding two penalisation forms to the least squares minimisation function, to try to get the benefits of both norms, without the disadvantages.

\[
L_2\text{ norm } \rightarrow \sum_{i=1}^{n} x_i^2
\]  
(Eq. 5)

\[
L_1\text{ norm } \rightarrow \sum_{i=1}^{n} x_i
\]  
(Eq. 6)

\[
L_0\text{ norm } \rightarrow \sum_{i=1}^{n} x_i^0
\]  
(Eq. 7)

The \(L_2\)-norm (Eq. 5) penalises on the two-norm of the coefficients, which indirectly penalises on the variance of the coefficients, and as a consequence, on overfitting. For the \(L_1\)-norm (Eq. 6), the penalisation is on one-norm of the coefficients. Due to the nature of the constraint, it gravitates towards creating coefficients with zeros, making the norm suitable for regression shrinking, as it simplifies the model. The \(L_1\)-norm has also been utilised in the deconvolution context. As introduced in Chapter 1, deconvolution is an inverse problem, as one seeks to find the causal factors that are producing the data, which are in this case the coefficients. The \(L_1\)-norm is capable of generating these coefficients, due to the indirect creation of coefficients which have very small values. The \(L_0\)-norm (7), compared to the \(L_1\)-norm, directly penalises on the non-zero coefficients and due to this characteristic, it is more suitable for deconvolution. It will try to find the least amount of points to fit the data with a certain convolution function, depending on the penalty parameter \(\lambda\) that is used. If the correct \(\lambda\) is chosen, then the coefficients \(x\) can be extracted from the data.

One could argue to apply the \(L_1\)-norm, as this also forces some coefficients to be close to zero. However one crucial benefit that the \(L_0\)-norm has over the \(L_1\)-norm, is the sparseness of the solution.
that it produces, as the penalisation of the former is on the total number of non-zero coefficients, while the latter is on the sum of all the coefficients. A red data-line with 10 (convolved) data-points is shown in figure 8, the convolution is done by a simple Gaussian function. The blue dots represent the true positions of the data-points, whereby for each of the various norms the solution is shown with the green lines. These generated solutions display the coefficients that were determined by each norm.

![Figure 8: One-dimensional example of the solutions that are generated by the various norms.](image)

From this one can see that although the L1-norm produce results with zeros, the solution is not as sparse as the one returned by the L0-norm, as the L0-norm penalises on the total number of non-zero values directly. S. Hugelier et al. compared both norms from a deconvolution point of view, utilising the SPIDER (L0) and STORM (L1) algorithms and showed that the use of the L0-norm resulted in sparser and more quantitatively correct results [10].

### 2.3 SPIDER

Although it was mentioned that SPIDER applies the L0-norm, it is done so by utilising some mathematical prowess. Applying the L0-norm is computationally demanding, due to the non-convex nature of the 0-norm, this being a summation of binary values. The SPIDER algorithm applies an iterative weighted regression procedure [17], whereby the assumption is made that:

\[
\|\mathbf{x}\|_0 \approx \frac{x^2}{\hat{w}^2} \approx \hat{w} \ast x^2
\]  

(Eq. 8)

with x representing the vector of coefficients and \(\hat{x}\) being an approximation. From this, a transformation is observed, making the process of generating a solution computationally viable. Equation 9 shows how the coefficients (\(\hat{x}\)) are generated, whereby \(\hat{W}\) represents the diagonal matrix of \(\hat{w}\).

\[
\hat{x} = (\mathbf{C}^T\mathbf{C} + \Lambda\hat{W})^{-1}\mathbf{C}^T\mathbf{y}
\]  

(Eq. 9)

Two important points to note, firstly, the fact that although SPIDER is utilised in image deconvolution, the algorithm transforms the image to a singular vector (\(\mathbf{y}\)). This vector is then utilised in the algorithm. This is done due to the computational difficulties as a 3-dimensional convolution matrix has to be developed if an image is utilised instead of a vector. As such the convolution matrix has the form of figure 9, which is constructed from multiple 2D Gaussian distributions.
Figure 9: Illustration of a convolution matrix (C), which consists of multiple 1D-Gaussian distributions.

Secondly, the localisation is applied on a super-resolution grid, meaning that the solution obtained will contain more pixels than the original data, this is done so that the localisation accuracy is elevated. Because if this would not be applied, the limit of resolution would be capped at the size of the pixel of the imaging system. As such, a “zoom” is applied, whereby the solution vector \( \hat{x} \) is \( n \) times larger than the data vector \( y \). Within this framework a zoom of 4 was applied, meaning that a singular camera pixel will be transformed to a 4 by 4 pixel.

To summarise: An image is inserted into the SPIDER algorithm, which transforms the image to a vector, applying the deconvolution approach on a super-resolution grid. The obtained solution vector is transformed back to an (sparse) image. The matrix calculations will not be further discussed, as it is outside the scope of the project, however, this can be read in [17].

2.4 Weight optimisation:
As seen in equation 4, the parameter \( \lambda \) defines the weight of the penalisation and balances the fit of the model. If \( \lambda \) is equal to 0, no penalty is applied on the coefficients and the model will be similar to a normal least squares minimisation. There are ways to go about finding the optimal \( \lambda \) value for the \( L_2 \) - and \( L_1 \)-norm, such as (generalized) cross-validation, Bayesian Information Criterion, Akaike’s information Criterion and the \( L(\lambda) \) – curve [14 18]. For the \( L_0 \)-norm, however, this is not the case, as the solution is non-convex. This discontinuous nature of the solution, due to the binary solutions that it generates causes the existing methods for optimisation to collapse, see Chapter 8.1 for a comparative study. This also applies to the \( L_2 \)-norm, but due to the difference in the nature of the solutions some methods to circumvent this issue are available for the \( L_1 \)-norm, as the \( L_1 \)-norm is not a summation of binary value, but a summation of the total intensity of the coefficients [18]. As of writing, no optimisation of \( \lambda \) for the \( L_0 \)-norm has been found in the literature. The approach applied for the \( L_0 \)-norm as of now is to use an arbitrary (can be estimated with high user experience) value for \( \lambda \) and determine by visual inspection whether the solution is acceptable or not. If there is no prior knowledge on the correct solution, then one has to determine the ‘correctness’ of the solution through user expertise, making the entire method for optimisation very subjective and dependant on the user.

2.5 Sum of the Normalised Terms
Looking towards the optimisation approaches of \( \lambda \) for the other various norms, one can observe certain similarities. The main point that is seen across all optimisation approaches is the balance
between the ‘size’ of the coefficients and the fit of the model on the provided data [19], which translates to making a trade-off between the solutions of the L0-norm and the least squares term as a function of $\Lambda$. This balance is determined by $\Lambda$, as having a high value for this parameter will result in a high weight added to the penalisation function, penalising heavier on the norm that is utilised. While a low value will result in an indirect higher weight added to the least squares term, which determines the fit of the model on the provided data. What these various approaches for this balance differ in, is how these two terms are calculated and how one decides to determine the optimal trade-off.

A new approach to the $\Lambda$-optimisation problem for the L0-norm is proposed in the context of deconvolution whereby a balance is made between the number of non-zero coefficients (NNZ, Eq. 10) and the root mean squared error (RMSE, Eq. 11) of the model with respect to the original data. The NNZ represents the penalisation, while the RMSE represents the fit of the model. The formula are:

\[
NNZ(\Lambda) = \sum_{i=1}^{n} (\hat{x}_i(\Lambda))^0
\]  
\text{(Eq. 10)}

\[
RMSE(\Lambda) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - C\hat{x}_i(\Lambda))^2}{n}}
\]  
\text{(Eq. 11)}

$\hat{x}$ being the solution generated by SPIDER, $n$, the total number of data-points, $y$, the original data and $C\hat{x}$ the reconstructed data. In figure 10 a representation of the NNZ and the RMSE as a function of $\Lambda$ can be seen. As one would expect, the NNZ decreases as a function of $\Lambda$, due to the higher penalisation on this term. As a consequence, the RMSE rises, due to the number of points to model the data decreasing. As a model utilising 20 points to explain the data will have a better fit than a model utilising 2 points.

![Figure 10: Representation of the RMSE and NNZ, as a function of $\Lambda$](image)

Similar to the L-curve used in the optimisation of $\Lambda$ for the L2-norm, whereby the least squares term is plotted against the penalisation term, the RMSE and NNZ are plotted against each other for the L0-norm. When doing so, an L-shaped trend is observed (figure 11), similar to the other norms.
From this, the optimal value can be deduced to be the closest point to the origin, as with the $L_0$-norm one tries to achieve the best reproduction of the data utilising the least amount of coefficients. This translates to the lowest RMSE and NNZ, respectively, which is in this case the origin (0,0). To find the closest point to the origin, the distance is calculated by taken the Manhattan distance from any particular point on the graph to the origin, however due to the differences between the values of the RMSE and NNZ, as the NNZ is a summation of binary values and the RMSE is dependent on the intensity of the data itself these terms cannot be compared directly. As such, a normalisation is applied, between 0 and 1:

$$N^* = \frac{N - \min(N)}{\max(N) - \min(N)}$$

(Eq. 12)

where $N$ is the term to be normalised and $N^*$, the product of normalisation. To calculate the Manhattan distance, the sum has to be taken of the two terms of a particular point on the graph, this is indicated with the green and red lines in figure 11. By which the formula will become:

$$SNT(\Lambda) = NNZ^*(\Lambda) + RMSE^*(\Lambda)$$

(Eq. 13)

From this the Sum of the Normalised Terms (SNT) as a function of $\Lambda$ can be plotted (see figure 12). One will observe a minimum that will be determined as $\Lambda_{\text{min}}$. The minimum will represent the $\Lambda$-value that will penalise the algorithm is such a manner that it will achieve the best reproduction of the signal, utilising the least amount of coefficients.
3 Methods

3.1 Tools & Datasets
During this project a number of tools and datasets were utilised, these will be discussed within this section.

3.1.1 Simulation

3.1.1.1 SOFI Simulation Tool
A major part of the methodology testing was applied on simulations. The SOFI Simulation Tool [20] was utilised for the creation of the simulated data. The tool was initially developed for testing Super-Resolution Optical Fluctuation Imaging (SOFI), however the simulated datasets could also be applied in SPIDER or any super resolution imaging algorithm. This is due to the fact that the tool simulates blinking emitters (Chapter 1.1.4.2), which is in most cases crucial to the deconvolution process, as it increases the capability of the applied algorithm. As trying to deconvolve 100 emitters on 1 image is much more difficult than having 100 emitters blink stochastically over 1000 frames, which e.g. could have an average of only 10 emitters on any particular frame. With real datasets (and the SOFI simulation tool) this blinking is in most cases not controlled, as such the number of emitters on any particular frame will be random.

3.1.1.1.1 Main settings
Although the tool has the capability to simulate blinking emitters, this parameter was not utilised due to computational limitations. SPIDER is computationally heavy, as such, the luxury of applying SPIDER on a 1000 frame dataset of a particular simulation was not available. The blinking parameter was shut down by setting the on-state lifetime to $10^6$ ms and the off-state lifetime to $10^{-6}$ ms. The number of frames to take of any particular simulation was set to 1 and to further decrease computation times, only a 32 by 32 pixel image was taken. In table 1 the standard settings are displayed. The density and S/B parameter are not included, as these settings were varied.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density ($\mu m^{-2}$)</td>
<td>X</td>
<td>Dark (elections/pixel/sec)</td>
<td>0.06</td>
</tr>
<tr>
<td>Acquisition time (sec)</td>
<td>1</td>
<td>Quantum efficiency (-)</td>
<td>0.7</td>
</tr>
<tr>
<td>Intensity peak (-)</td>
<td>1000</td>
<td>Gain (-)</td>
<td>6</td>
</tr>
<tr>
<td>S/B (-)</td>
<td>X</td>
<td>Pixel size ($\mu m^2$)</td>
<td>$0.1 \times 0.1$</td>
</tr>
<tr>
<td>On-state lifetime (ms)</td>
<td>$10^6$</td>
<td>Pixel number (pixels)</td>
<td>$32 \times 32$</td>
</tr>
<tr>
<td>Off-state lifetime (ms)</td>
<td>$10^6$</td>
<td>Numerical aperture (-)</td>
<td>1.3</td>
</tr>
<tr>
<td>Average bleaching time (sec)</td>
<td>$10^6$</td>
<td>Wavelength (nm)</td>
<td>600</td>
</tr>
<tr>
<td>Acquisition rate (frames/sec)</td>
<td>1</td>
<td>Magnification</td>
<td>1</td>
</tr>
<tr>
<td>Readout noise (rms)</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The simulation tool also had the capability to simulate structures, however this was not utilised and the emitters were randomly positioned on the pixel image. However to remove the possibility of an unlucky draw where e.g. all the emitter would be near each other and not spread over the whole image, which would increase the local density to a much higher degree than was initially given, multiple images of the same density would be simulated, only differing in the positioning. From here
The assumption is made that the solutions would not be solely dependent on the randomisation of the positioning. The size of the distribution of a singular emitter that is simulated is calculated from pixel size, NA, wavelength of light and magnification factor. From these parameters the standard deviation ($\delta_c$) of the Gaussian distribution is calculated, this refers back to Chapter 1.1.4. Rayleigh’s Criterion determines this distribution, however $d_{\text{Rayleigh}}$ considers the full width at half maximum (FWHM), having the value in nanometres, to transform this to pixels and $\delta_c$, equation 14 is utilised:

$$
\delta_c = \frac{0.61 \lambda}{NA} \frac{\text{magnification}}{2 \sqrt{2 \ln(2)} \ast \text{pixel width}}
$$

(Eq. 14)

Dividing by the pixel width, which is within this framework, 100 nm, transforms the value to pixels and dividing by $2 \sqrt{2 \ln(2)}$ transforms the FWHM to $\delta_c$. A $\delta_c \approx 1.19$ was determined with the chosen parameters, these settings were determined to be of an average imaging system.

The average bleaching time was set to $10^6$ seconds as the goal of the simulations was to investigate only a singular time point, as such the effect of bleaching was not considered. The acquisition time and rate was set to 1 to simulate only 1 image. The readout noise, Dark current, quantum efficiency and gain were not changed from the standard settings chosen by the tool, as these were deemed appropriate and were not of high priority for investigation. These terms simulated the noise and loss of signal by the imaging system. As such, although an intensity peak, which determined the maximum of the singular emitters, of $10^3$ was chosen, the actual perceived signal was $\approx 420$, due to these parameters that caused the loss of signal.

3.1.1.1.2 Variables

Two terms in table 1 were left out, these were the parameters that were varied to test the heuristic approach that was developed.

The first parameter is the density, which considers the amount of singular points that are to be simulated within the pre-determined image boundaries. The values are given in the # points / per $\mu m^2$. The translates to:

$$
density = \frac{\text{# of points}}{\text{# of pixels} \ast \text{pixel size (um$^2$)}}
$$

(Eq. 15)

As such, for a given image size and density the tool will simulate a certain number of points. Depending on the situation, either the density or # of emitters (points) will be given. In figure x an example is given of a density of 2.5 $\mu m^2$ or 25 emitters with the parameters previously mentioned.

*Figure 13: Simulated example of 25 emitters.*
The second parameter is the S/B term, this stands for Signal / Background. This parameter considers the ratio between the intensity of a singular emitter and the average background signal. A value of 0.5 for S/B will heighten the background signal to 50% of the maximum signal of a singular emitter. The tools tries to simulate unwanted/untargeted signals that are due to the imaging system (see Chapter 1.2.1). However added to this is the readout noise, which is dependent on the intensity of the signal that is perceived by imaging system, as such, a higher background will cause the signal to be not only more intense but also noisier as well, this is shown in figure 14.

![Image of figure 14](image1.png)

**Figure 14:** Singular emitter simulated with an increasing background.

### 3.1.1.2 Complex structures
Other than images with randomised positioning, complex simulated structures were also investigated. Two datasets were taken from a 2013 single molecule localisation microscopy challenge, given by EPFL. The datasets are the “Contest Dataset 1 (HD)” and “Contest Dataset 2 (HD)”. No ground truth of the emitters was available, however the pixel size and $\delta_C$ were taken from [21 22]. These were determined to be 125 nm and 1, respectively. Each dataset has an image size of 128 x 128 pixels, which simulated blinking emitters, with the first dataset having 1001 frames while the other has 204.

### 3.1.2 Real
As a final hurdle for the heuristic approach that was developed, it was applied to a real dataset. Specifically, on the analysis of mitochondria in a live cell, utilising epi-illumination [10].

A live HEK293-T cell labelled with DAKAP-Dronpa, which targets the outer-membrane of mitochondria [7], was analysed utilising Total Internal Reflection Fluorescence Microscopy (TIRF). The analysis was over the course of 30 seconds, in which 1000 frames were taken, utilising a 488 nm excitation laser, NA of 1.2 and a pixel size of 100 nm, the detection wavelength was not mentioned, as such it was assumed to be 525 nm, as this is the maximum of the emission spectrum of DAKAP-Dronpa. The image resolution is 40 by 80 pixels, more on the imaging system and sample preparation can be found in [10]. An added note, over the course of the 30 seconds of analysis time, photo bleaching and movement of the mitochondria were observed.

### 3.1.3 Performance tools
As mentioned previously, the validation or determination of the correctness of the solution for the $L_0$-norm is done so by either, the use of the ground truth, if available, or user expertise. From the ground truth point of view, one means to utilise the information of the true emitter positions to determine
the performance of the solution. From the user expertise point of view, the solutions will be visually inspected and determined by the user for its validity.

For the performance testing, four parameters will be utilised. These were developed by Hugelier et al. [10], which are the Recall (Eq. 16), Accuracy (Eq. 17), False Positive count and Sparsity (Eq. 18).

\[
Recall \,(\%) = \frac{\# \text{True emitters found}}{\text{Total } \# \text{ of true emitters}} \times 100\% \quad (\text{Eq. 16})
\]

\[
\text{Accuracy} \,(\text{nm}) = \frac{\sum_{i=1}^{n} \sqrt{(v - \hat{v})^2 + (b - \hat{b})^2}}{n} \quad (\text{Eq. 17})
\]

\[
\text{Sparsity} \,(\%) = \frac{\text{Total } \# \text{ of generated emitters}}{\text{Total } \# \text{ of true emitters}} \times 100\% \quad (\text{Eq. 18})
\]

The recall rate gives the percentage of the discovered true emitters. A true emitter is discovered when a generated emitter is within a certain distance of the true position. This distance was determined to be 100 nm, as this was the width of a singular pixel. If one has a recall rate of 0 %, then no emitter were discovered and having a 100 % means the discovery of all the simulated emitters.

The accuracy gives the average distance of the generated emitters to an assigned true emitter. These tools allow multiple emitters that were generated by SPIDER to belong to a singular true position. The accuracy therefore gives the average distance of all the emitters that belong to a particular true emitter. The distance is calculated by taken the Euclidean distance from the generated point ($\hat{x}$) to the true position ($x$), with ($\hat{v}, \hat{b}$) and ($v,b$) being the coordinates of their respective emitters on the pixel image. The sum is taken over n emitters that were generated to belong to the true emitter and this is divided by n to get the average distance.

Continuing with the false positive count, this is the number of emitters generated by SPIDER that do not belong to a true position. Meaning that the generated emitter tries to fit the background/noise of a particular image. This means that no true emitter is near the generated false positive within a 100 nm radius, as this was determined to be the maximum distance one generated emitter can have from a true emitter.

The sparsity considers all the generated emitters (including false positives) and gives the % relative to the total number of true emitters. This gives one an indication on the total number of points needed to fit the image. This could mean that having a sparsity of above 100%, more points than there are true emitters are necessary to fit the data.

Although on their own, the performance parameters give a considerable amount of information on the “correctness” of the obtained solutions. If the performance parameters on a whole are observed more information can be extracted. As an example, if there is a recall rate of 100 % and no false positive are observed, this does not necessarily mean that the solution obtained is perfect, due to the sparsity term. If the sparsity is 1000%, that would mean that on average 10 points are needed for each true emitter. This would defeat the purpose of applying an $L_0$-norm, as one of the main benefits is the
sparse solutions that it should obtain. As such, one must consider all four performance parameters as a whole and not as individual terms.

3.1.4 The Bootstrap

The bootstrap [23] is a randomised resampling approach, to determine the confidence interval of a population from a particular dataset without any a priori information. It is applied in robustness testing, as it can give an indication on the robustness of a method. The bootstrap applies a randomised (minor) perturbation on the data. This is done a number of times (depending on the computational limitations, 10² to 10⁴ times) and the solutions that are obtained for these new datasets are investigated. Since this is done a considerable amount of times, a distribution can be obtained, which can shed some light on the approach in question. There are multiple pathways of applying a bootstrap, depending on the information that is available. Within this framework the non-parametric bootstrap will be utilised, as the assumption is made that there is no information on the distribution of the noise. The non-parametric bootstrap resamples from itself to create new datasets. Specifically within the framework of image analysis, the difference (\( \hat{e} \)) between the reconstructed and the original data is resampled. The \( \hat{e} \) is resampled due to the fact that original data cannot be resampled, as resampling the original data will deform the convolved signals, removing the ability to localise them, as they are deformed randomly, the same goes for the reconstructed dataset, as resampling that will cause the convolved signal to deform and the algorithm will no longer be able to localise them, as such \( \hat{e} \) is resampled. In table, the steps to the bootstrap approach are displayed:

**Table 2: Step by Step approach of the Bootstrap**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>( \hat{e} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>( y \rightarrow \text{SPIDER(SNT)} \rightarrow \Lambda_{\text{min}} )</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>( \text{SPIDER}(\Lambda_{\text{min}}) \rightarrow C\hat{x} )</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>( \hat{e} = y - C\hat{x} )</td>
<td></td>
</tr>
<tr>
<td>Step 4</td>
<td>( \hat{e} \rightarrow \text{Bootstrap}^* \rightarrow \hat{e}_b^* )</td>
<td></td>
</tr>
<tr>
<td>Step 5</td>
<td>( \hat{e}_b^* + C\hat{x} = \hat{y}_b^* )</td>
<td></td>
</tr>
<tr>
<td>Step 6a</td>
<td>( \hat{y}<em>b^* \rightarrow \Lambda</em>{\text{min}} \rightarrow C\hat{x}_b^* )</td>
<td>( \hat{y}<em>b^* \rightarrow \text{SNT} \rightarrow \Lambda</em>{\text{min}}^* )</td>
</tr>
<tr>
<td>Step 6b</td>
<td>( \hat{y}<em>b^* \rightarrow \Lambda</em>{\text{min}}^* \rightarrow C\hat{x}_b^* )</td>
<td></td>
</tr>
<tr>
<td>Step 7</td>
<td>( \bar{\hat{x}}^* \pm \delta^* )</td>
<td></td>
</tr>
</tbody>
</table>

Starting with step 1, on the data \( y \), SPIDER, utilising the SNT approach is applied, from which \( \Lambda_{\text{min}} \) is obtained, that one assumes to be the correct \( \Lambda \)-value. SPIDER is applied again in step 2, utilising \( \Lambda_{\text{min}} \), which results in a solution \( C\hat{x} \). This solution is deemed by the SNT approach to be the optimal solution, containing all the essential information, with no addition of noise or background. The reconstruction is subtracted from the original data, which results in \( \hat{e} \) (step 3). Which one assumes to contain only untargeted information, as all the essential information is subtracted from the original data. In step 4 the bootstrap is applied, of which an illustration of the approach can be seen in figure 15.
The resampling is performed by taking a set of pixels within a certain window and randomly choose a percentage of pixels that will be exchanged for a different set of pixels. It is sectioned into windows, from which a percentage of the cells is changed to another particular cell in that same window. The use of windows is due to the structural information that is created in the vector, see figure 16 for an artistic representation of this problem.

![Artistic representation of the applied bootstrap approach](image)

*Figure 15: Artistic representation of the applied bootstrap approach, where A: Transformation to vector; B: Resampling; C: Transformation to image.*

Due to the “hole” that is left behind in the vector, it becomes structured, as such, if one would resample over the whole vector, spikes could be created in these “hole”, due to chance that e.g. point I could be resampled to be positioned at point II. However the chosen window size, should at least encompass the size of a single emitter, if not, the perturbation could have little to no impact on the overall image, e.g. point III could be exchange for point II, having no impact on the overall vector. As such, a window size of 4 by 4 pixels was chosen, with 4 pixels being exchange in each window, this translates to changing 25% of $\hat{e}$. After $\hat{e}$ is bootstrapped $b$ times, $\hat{e}^b$ is added to the original reconstructed data, to crease a bootstrapped $y$, which is $\hat{y}^b$. From this point on, 2 paths can be taken. The first one is applying the same $\Lambda_{\text{min}}$ that was previously determined with the original data $y$, or the second path is recalculateing $\Lambda_{\text{min}}$, utilising the SNT approach to create $\Lambda_{\text{min}b}$. Both pathways will produce the same amount of data, however, the information in the solutions is slightly different. With the first path (left-side in table 3), being more constrained to the solutions that it produces, due to the utilisation of the same $\Lambda$-value. The second approach, however, is more free to utilise a new $\Lambda$-value. Both approaches will be investigated and compared.

The solutions will be plotted in a different way that the other results, as with the bootstrap the distribution of the solutions is crucial. As such, a boxplot is utilised, displaying the range of the data with black disconnected lines, while putting a box around the interquartile range (IQR). The IQR is

![Artistic representation of the structure of the “hole” that is left behind](image)

*Figure 16: Artistic representation of the structure of the “hole” that is left behind.*
specified to be all the solutions below the 90\textsuperscript{th} percentile and above the 10\textsuperscript{th} percentile. Outliers will be dependent on this range. Two forms of outliers will be calculated, the first being an “Outlier”, which is assigned to a particular solution if that solution deviates from the mean with the IQR multiplied with certain factor, this factor being 1.5 and 3. Meaning that, if a particular solution deviates from the mean between 1.5*IQR and 3*IQR, it will be deemed an “Outlier”. The second form is an “Extreme Outlier”, which considers everything above or below 3*IQR, as such.

### 3.2 Approach

The testing procedure of the heuristic approach will be split into two sections. Beginning with performance testing, whereby the four performance parameters will be utilised to determine the performance of the solution obtained from the SNT approach. The simulations discussed in Chapter 3.1.1.1 will be utilised here, as only with these datasets the ground truth is known beforehand. The second section will be from the users’ point of view, visually inspecting the overall structure of the solutions that were given by the SNT approach. In this section the simulations in Chapter 3.1.1.2 and the real dataset in Chapter 3.1.2 will be used as the ground truth is not known of these datasets.

#### 3.2.1 Localisation performance

As mentioned before, the four performance parameters will be utilised to determine the performance of the obtained solutions. Localisation performance testing has two main sections. Firstly, the testing on singular images with varying densities and backgrounds, to determine the performance and limits of the approach. The second section will go over the robustness testing that was performed, by utilising the bootstrap approach. By applying the bootstrap, one can determine the robustness of the approach and how the method copes with perturbation of the data.

For each subsection, an example will be first shown to demonstrate how the solutions are obtained. Secondly, a preliminary study will be shown whereby repetitions of the example shown previously will be investigated and lastly an extended study will be discussed.

#### 3.2.1.1 Densities

A range of 0.5 to 15 emitters per um\textsuperscript{2} was simulated, which each density having 50 repetitions, to resolve the unlucky draw problem. The values of the performance parameters are compared to the work of Hugelier et al., which performed similar testing utilising the same parameters. However the approach to obtain a valid $\lambda$-value was done by inspecting the four performance parameters and choosing the best result, as such one can assume the results obtained by Hugelier et al. to be the limit of the SPIDER algorithm, as the ground truth was utilised to determine the optimal $\lambda$-value.

#### 3.2.1.2 Background

The height of the background was varied from 0 to 70 \% of the intensity of a single emitter. This was done to determine the limits of the approach, as the ratio between the essential and non-essential information decreases. It was tested on 5 different densities and for each density and background 10 repetitions were simulated, each having a different positioning of the emitters.

#### 3.2.1.3 Robustness testing

For the robustness investigation, 3 different densities (1, 5 and 10 um\textsuperscript{2}) were simulated, each having 5 repetitions. Each repetition was 250 times bootstrapped, creating 250 solutions for each repetition and density. The Model Robustness and SNT Robustness approach were both investigated and compared to one another. One has to take into account that in most cases with the non-parametric
bootstrap one would try to create 1000 or 10000 bootstrapped sample, this however could not be done due to computational limitations.

3.2.2 Structure determination
This investigation is less objective than the previous investigation, as the ground truth is not known. Thus the solutions that are obtained will be investigated through visual inspection. Due to computational limitations, only the first 10 image were investigated for the two simulated complex structures. This was not the case for the real dataset and all 1000 images were investigated.

4 Results & Discussion
Within this chapter the results achieved are discussed. Chapters 4.1 and 4.2 utilise the performance parameters discussed in Chapter 3.2, while with Chapter 4.3 only structural information is investigated, as the ground truth is not known. Sub-chapters 4.1.1, 4.1.2, 4.1.3.1 and 4.1.3.2, are each divided into three segments. Starting with an singular example in the first segment, from there the second segment will show a preliminary study and the final section will go over an extended study. Chapter 4.2.1 discusses the results achieved with Complex Simulations, while Chapter 4.2.2 extensively investigates a real dataset that was obtained (see Chapter 3.1).

4.1 Localisation performance
The localisation performance of SPIDER utilising the SNT approach is discussed within this sub-chapter. The performance testing is separated into three sections, first looking towards the effect of higher densities, and secondly, observing the effect of an increased background signal and lastly, discussing the solutions of the bootstrapping approaches.

4.1.1 Densities
A density range of 0.5 to 15 um$^2$ was taken and for each density 50 images were simulated, with randomised positioning.

4.1.1.1 Example
As an example, a singular image with a density of 2.5 um$^2$ is investigated, on which SPIDER, utilising the SNT approach, is applied. Figure (left) shows the original image map, while in figure 17 (middle), the SNT curve as a function of $\Lambda$ is displayed. The solution at $\Lambda_{\text{min}}$ is shown in figure 17 (right). A clear minimum is seen in the SNT curve and the accompanying $\Lambda_{\text{min}}$ solution show an exceptional result. Observing it only by eye, one can see that most if not all the emitters have been localised.

![Figure 17: Example of 25 emitters, showing the original image, SNT curve as function of $\Lambda$ and the sparse image.](image)
4.1.1.2 Preliminary study
From this point, the localisation performance will be investigated. Taking the same density shown previously, 50 repetitions are observed in the results below. The results are great, looking towards the recall rate, one can see a near perfect result, as almost all emitters are localised, and with an accuracy of around 30 nm on average. No false positives were observed and although not all the emitters were localised using only one 1 point for each emitter, it is not far off, taking an average of 110 % for the sparsity, means that for every 10 emitters, 1 true emitter utilises 2 points.

![Figure 18: The four performance parameters applied on 50 images with a density 2.5 μm⁻²](image)

4.1.1.3 Extended study
With the extended study, a considerable range of densities is investigated. Comparing the results to the work of Hugelier et al. [10], that did similar testing, utilising the same performance parameters, however it was applied on a different dataset. Different simulation parameters were chosen, as such the results should be taken with a grain of salt. This comparison is just to give an idea in what ball park the limit of the SPIDER algorithm is. The results of Hugelier et al. are seen in figure 20, and when compared, similar results are observed up to a density of 8 μm⁻². After which the SNT approach seems to collapse. The is due to a number of factors. Starting with the fact that no false positives are allowed with the SNT approach, it is already known that with SPIDER, one has to allow false positives to surface, to allow for more true emitters to be discovered, as such the perfect reconstruction will not be feasible at higher densities. The SNT approach is based on the balance between reconstruction and the utilisation of the least amount of points to perform the reconstruction, as such, it will not fit the signal that originates from something else that is not an emitter. The second significant point is the sparsity.
Comparing the sparsity and recall rate one observes that with the results of Hugelier et al. a reasonable gap is seen, meaning that multiple points are necessary to fit either the signal of the emitters or the signal of another source. The former only increases the sparsity, while the latter increases the false positive count and sparsity. This phenomenon is not seen in the SNT approach, where the recall rate follows, although no exactly but closely, the sparsity. As such from this we can conclude that although not all true emitters have been found, the points that have been generated by the approach are most likely the true positions. These two factors most likely cause the collapse at a density of $9\ \mu m^{-2}$. Considering the high density, the signal of multiple emitters can be fitted, utilising only one point. This also explains the accuracy change, as having 1 point in between two emitter signals will increase the distance than having 2 points closer to their respective maxima. A slight remark on the standard deviation of the sparsity, considering the relative nature (%) of the parameter, a bigger deviation is seen with lower densities. This is due to the fact that if 11 emitters are generated with a density of 1 ($10$ true emitters), an increase of 10% in sparsity is observed, while, if 101 emitters are generated with a density of 10 ($100$ true emitters) an increase of only 1% is observed. This also considers the recall rate, as such the standard deviation is not considered, when comparing across densities.

Figure 19: Four performance parameters across a range of densities, repeated 50 times.
4.1.2 Background
Within this section the effect of a higher background will discussed.

4.1.2.1 Example
A singular example is shown below, whereby an image map with a density of 2.5 um$^2$ is simulated with three different background %'s. With the % representing the relative height of the background to the maximum intensity of a single emitter. Showing the original images with a background of 5, 25 and 50%, in figure 21. It is quite clear that the information that is available per image decreases as the background increases. Looking at the SNT curve, it seems to have shifted its minimum towards a lower $\lambda$-value. This is due to the fact that, since a normalisation is applied, the background will have a larger impact on the SNT curve than the emitters. Observing the sparse images at $\lambda_{\text{min}}$ shown in figure 21, one can see that the 5 and 25% solutions show a reasonable localisation of emitters, while with 50% the background is “localised” as well. What one can take from this is that if the balance of signal between the background and emitters is shifted towards the background, the SNT approach will try to accommodate for that and fit the background with emitters as well.
Figure 21: A singular simulation at three different backgrounds (5, 25 and 50%), showing the original image, SNT curve and sparse image.

4.1.2.2 Preliminary study

The previous results have been repeated 10 times, and the four performance parameters show an interesting result. There is no significant change in the recall rate or accuracy. However, the false positives and sparsity show for the 50% background similar curves, meaning that most if not all the added emitters are trying to localise the background. This means that the true emitters that were localised were not impacted in anyway by this increased background. This is due to the fact that the emitters still have twice the height of the background.
4.1.2.3 Extended study

Extending this to a greater range of densities (0.5 to 5 um⁻²) and backgrounds (0 to 70%), the results observed are quite remarkable, making the point of collapse for the SNT curve quite apparent, while also finding that the collapse is dependent on the density. Starting with the recall rate, nothing is out of the ordinary, only observing the slightly lower values for the 50 emitters, which was observed in Chapter 4.1.1.3 as well. For the accuracy however, there seems to be a slight increase after a background of 40%. Which is most likely due to the case that more “emitters” that localise the background are observed, giving a higher probability of assignment to a true position. Moving on to the false positives and sparsity, the previous collapse of the SNT approach is observed again. However the interesting point from this is the fact that the point of collapse starts earlier with lower densities, meaning that the threshold of the height of the background for the collapse is lower for lower densities. This is quite counterintuitive as one would expect the higher densities to be more susceptible to worse solution at a higher background. One could argue that with a higher densities it is more likely for a generated emitter that “localises” the background to be assigned to a true position, this however is not the case, as the false positives are near equal to the sparsity, looking at one particular density at a time. As due to the sparsity being relative, comparing across densities in not appropriate.
Figure 23: Four performance parameters at 5 densities, 15 backgrounds, repeated 10 times.

This collapse was more closely investigated, by looking towards the SNT curves. Two densities (0.5 and 2.5 \( \mu m^2 \)) were investigated and the curves of a range (10 to 70%) backgrounds is shown in figure 23. From this, the cause of the collapse is quite noticeable. It would seem that a second minimum in generated in the curve at a lower \( \lambda \)-value. This is the case for both densities, however, for the higher density the minimum becomes gradually apparent, while with the lower density this is not the case.

Figure 24: SNT curves at densities 0.5 and 2.5 \( \mu m^2 \), with a background range of 10 to 70%.

Considering the foundation of the SNT approach being: “The SNT approach will find the best reproduction utilising the least amount of coefficients”. As such, when one considers the ratio between the signal of the emitters and the background:

\[
\text{Ratio} = \frac{\sum_{i=1}^{n} C x_i^2}{\sum_{i=1}^{n} e_i^2}
\]

(Eq. 19)

This collapse was more closely investigated, by looking towards the SNT curves. Two densities (0.5 and 2.5 \( \mu m^2 \)) were investigated and the curves of a range (10 to 70%) backgrounds is shown in figure 23. From this, the cause of the collapse is quite noticeable. It would seem that a second minimum in generated in the curve at a lower \( \lambda \)-value. This is the case for both densities, however, for the higher density the minimum becomes gradually apparent, while with the lower density this is not the case.

Figure 24: SNT curves at densities 0.5 and 2.5 \( \mu m^2 \), with a background range of 10 to 70%.

Considering the foundation of the SNT approach being: “The SNT approach will find the best reproduction utilising the least amount of coefficients”. As such, when one considers the ratio between the signal of the emitters and the background:

\[
\text{Ratio} = \frac{\sum_{i=1}^{n} C x_i^2}{\sum_{i=1}^{n} e_i^2}
\]

(Eq. 19)
Whereby the two-norm is taken as this gave a good representation of the results, \( \mathbf{x} \) defined as the true positions, and \( \mathbf{C}\mathbf{x} \) being the true reconstructed image without unwanted/untargeted signals. This ratio will decrease, as the background increases. When the term is plotted as a function of background, the curve in figure 25 is observed. The black stars indicate the point of collapse for the SNT approach, the remarkable point being that a somewhat linear trend is seen. Meaning that even though the signal to background ratio for a particular emitter is high, the approach will still collapse if the summation of the signal for the background is higher than that of the emitters.

\[
\text{Ratio} = \frac{\sum_{i=1}^{n} \mathbf{C}\mathbf{x}_i^2}{\sqrt{\sum_{i=1}^{n} 0^2}}
\]

(Eq. 20)

4.1.3 Robustness testing

Within the section the results of the robustness testing are discussed. With the first subsection considering Model robustness, whereby the original \( \Lambda_{\text{min}} \)-value, that was determined by the SNT approach, is utilised on the bootstrapped images. While with the second subsection a new \( \Lambda_{\text{min}} \)-value is determined by applying the SNT approach on the bootstrapped images.
4.1.3.1 Model Robustness

4.1.3.1.1 Example

Starting with an example that has a density of 5 um\(^2\). Figure 26 shows the original and an example of a bootstrapped image. If one observes closely a minor difference can be observed. Below, the solutions are shown. However for the solution of the bootstrapped image, the mean is taken over all 250 bootstrapped images. To give an indication on where the most variation is localised. Comparing both images, if there would be no variation in the mean solution, this would mean that the perturbation of the data has not impacted the solutions of the bootstrap at all. This however is not the case, as one who would compare the mean solution of the bootstrap to the original, in the denser areas of the image more variation in the solution is seen. Especially in the dense bottom-left area of the image, where multiple emitters are more densely packed compared to the overall image. This is to be expected, as applying a perturbation to this area will have a more dramatic impact on the image.

![Image of original and bootstrapped images](image)

*Figure 26: Bootstrapped example at density 5 um\(^2\).*

4.1.3.1.2 Preliminary study

Looking at the accompanying performance parameters, nothing remarkable is to not here, except the fact that there is a slightly higher average accuracy and sparsity than the original image. This shift in accuracy is most likely due to the maxima of a emitter being shifted due to the perturbation of the data, causing the localisation of the generated emitter to be shifted as well. The shift in sparsity could be due to the deformation of the convolved signal, increasing the necessity for multiple points to fit 1 emitter.
4.1.3.1.3 Extended study
This investigation has been extended to 5 repetitions and various densities (1, 5, 10 um²). The first point that has to be stressed is the lack of data. Meaning that 250 bootstrapped samples should be considered the bare minimum to apply a bootstrapping approach and repeating this 5 times is not enough to be considered an extensive study. Due to computational limitations, as the SPIDER calculations are very heavy, this is the limit of the amount of data that can be investigated, to obtain the results within an adequate amount of time.

Continuing with the obtained results, starting with the sparsity distribution for the lower density, this significant difference compared to the other densities is due to the term being relative. Another point of interest is the outlier that was determined in the recall rate for the lower density, this means for a particular bootstrapped image the perturbation caused SPIDER to properly localise all the emitters. The probability of this occurring is most likely very small, as such only 1 bootstrapped sample showed this occurrence, added to this, it was determined to be an extreme outlier. If one observes the overall results, then in all except for the sparsity, the original solution is inside the IQR. Meaning that the solutions of the bootstrap are not significantly different from the original. However when looking at the sparsity, one can see that in all, except 2 cases the average sparsity is higher than the original. This is due to the creation of deformed convolved signal. The Gaussian distribution is changed and to accommodate for this change, more points will be necessary to explain 1 true emitter.
4.1.3.2 SNT Robustness

4.1.3.2.1 Example
Other than Model robustness, the robustness of the SNT approach was investigated. This was compared to the Model Robustness, to get a point of reference. From the mean solution previously shown, when compared to the mean solution of the SNT approach, one observes a little to no difference in the solutions, from this one can conclude that reapplying the SNT approach will give a similar solution.

4.1.3.2.2 Preliminary study
When the four performance parameters are investigated for the particular example (density 5 um²), again little to no difference is observed. The only noteworthy point is that the sparsity seem to be slightly higher than the average for Model robustness.
4.1.3.2.3 Extended study
Observing the extended study, the significant differences that are observed between the two approaches are the 2nd sample of the 50 emitters and the 2nd, 4th and 5th sample of the 100 emitters, considering only the IQR.
The distribution of the solutions has increased as well, as such a $t$-test was applied to determine whether this was significant or not. The results of the null-hypothesis (based on a 90% confidence interval) are shown in table 3. The false positives were not included, as there is no difference in the performance of this parameter between the approaches. The same goes for the recall rate at a density of 1 um$^2$, as such no values are given. With table x, having a value for 1 means that there is a significant difference between the two approaches for that particular sample. While if a 0 is given, this would mean that is no significant difference.

Table 3: Null-hypothesis of the performance parameters comparing the SNT robustness and Model robustness

<table>
<thead>
<tr>
<th>Sample</th>
<th>1 um$^2$</th>
<th>5 um$^2$</th>
<th>10 um$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4.</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4.</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sparsity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

It is quite clear from these results that there is a significant difference in the solutions, except for 4 samples with the accuracy, that are produced by the SNT approach. This is most likely due to the change in signal that is generated by the perturbation. It is known that $\lambda$ is dependent on signal intensity, see equation 4. As such the change in intensity due to the perturbation of the data causes a new $\lambda_{\text{min}}$-value to be utilised.

For this it can be said that it is most likely better to do multiple measurements and apply the mean $\lambda_{\text{min}}$-value that is determined by the SNT approach across all the samples.

4.2 Structure determination

Within this chapter the localisation performance will no longer be investigated, as the ground truth of the next datasets are not known. First, the 2 contest datasets are explored after which the real dataset is investigated.
4.2.1 Simulated
For these 2 datasets, only the first 10 frames were considered, due to computational limitations. Although a small number of frames were utilised, this had little to no impact on the determination of the complex structures.

4.2.1.1 Contest dataset 1
With this particular dataset two points of have to be considered, the first being the untargeted/unwanted fluorescence that is observed on the right middle side and the second point is the overlapping tubulins. From the first point the solution shows the complete removal of the background, while the second point, the overlapping tubulins in the zoomed image shows that the overlapping emitters are properly localised.

Figure 32: Contest Dataset 1, the original image, with the accompanying sparse image, with a zoomed addition on both images.

4.2.1.2 Contest dataset 2
This second dataset is quite more complex, but shows similar results. Observing the zoomed images, one can see that the two tubulins that are near each other are reasonably separated. Given a higher resolution.
The investigation has mainly been superficial as only 10 frames were investigated. These datasets were used to get a more accurate grasp on the capabilities of the SNT approach.

### 4.2.2 Real

Now the most crucial dataset will be investigated. The most crucial, as this is the only real dataset, on which the SNT approach is applied. This gives us an indication, whether or not the approach is applicable under real conditions. Figure 34 (left) shows the mean image of the entire datasets (1000 frames). In the next figure (34, right), the summation of the intensity of a particular image is shown as a function of frame number. This gives an idea, as to what occurs in the image over time. As mentioned in Chapter 3.1.2, photo-bleaching is observed.

This is known to be photo-bleaching, due to the characteristic exponential decay. It is also known that $\lambda$ is dependent on the signal intensity, as such one can assume that the same $\lambda$-value cannot be
applied for both frames 50 and 500. The SNT approach was applied on these particular frames and the corresponding SNT curves are shown in figure 35 (left).

![SNT curve of frames 50 and 500 (Left) and the SNT map as a function of $\Lambda$ and Frame # (Right)](image)

**Figure 35:** SNT curve of frames 50 and 500 (Left) and the SNT map as a function of $\Lambda$ and Frame # (Right)

A clear shift is observed in the curve, added to this, the relative value for $\Lambda_{\text{min}}$ had also shifted to a higher value. The SNT approach has been applied on all 1000 frames and the corresponding map of the curves is shown in figure 35 (right). Frames 50 and 500 are indicated with a white line. Quite remarkably, the minimum seems to gradually decrease as a function of the frames, while also giving a more ambiguous minimum. In figure 36, $\Lambda_{\text{min}}$ is plotted against the frame #. This curve seems to near identically follow the curve of the total intensity. Meaning that the SNT approach is directly proportional to the intensity of the image.

![$\Lambda_{\text{min}}$ as a function of Frame #](image)

**Figure 36:** $\Lambda_{\text{min}}$ as a function of Frame #

A noteworthy point is the horizontal lines that are observed in the first 300 frames, this is the due to the step – size between the $\Lambda$-values on which the SNT approach was based on. A logarithmic scale was applied, meaning that at higher $\Lambda$-values bigger steps were taken. This was applied to reduce computation times.

The SNT approach is applied on the entire dataset, utilising a specific $\Lambda_{\text{min}}$ value for each frame, the resulting sparse image is shown in figure x.
Figure 37: Mean sparse image of all 1000 frames, applying the SNT approach independently on each frame.

This is an immense increase in resolution, compared to the original image. The hollow mitochondrial structure has become very clear. A minor point is the “double strings” seen in the left structure, indicated by a white arrow. This is due to the movement of the mitochondrion, which blurs the resulting image. The analysis of the live cell was done over 30 seconds, so movement in the image is to be expected. To further showcase the improvement, the white lines indicated in the original and sparse image are plotted over each other in figure 38. This clearly shows that the resolution increase dramatically. The width of the x-axis is 1000 nm, this shows that the resulting sparse image truly reached the super-resolution. However to make sure of this, the FWHM was calculated: the left peak has a FWHM of 61 nm, while the right peak has a FWHM of 110 nm.

Figure 38: Zoomed comparison of increase in resolution before and after the SNT approach.

The lowest possible number deemed by Rayleigh’s Criterion is: 267 nm, considering 525 nm emission and a NA of 1.2. From this we can safely say that super-resolution has been achieved utilising the SNT approach on SPIDER.
5 Conclusion

Multiple conclusions can be made from the results that were achieved. Starting off with the fact that although the SNT approach is heuristic, it achieves its goals, within certain confines. That goal being, finding the $\Lambda$-value for SPIDER that produces a model which gives the best reproduction of the original signal utilising the least amount of points to fit the data. This causes however that at higher densities, less emitters will be found, as multiple emitters will most likely be fitted with a singular data-point. However, this approach will not allow false positives to be created (in most cases) and will be able to localise the overall structure of complex datasets, as seen with the simulated structures. Added to this is the fact that it could be utilised as a post-processing tool, since it can account for the photo-bleaching effect and removes any background, as it only localises single emitters, if the data is modelled appropriately. With the final results, the conclusion was that super-resolution was achieved, having a FWHM of 110 and 61 nm, while Rayleigh’s Criterion deemed the limit by to be 267 nm.

6 Future perspectives

For further research, multiple pathways can be taken.

Starting with SPIDER, the development of a graphical user interface with the SNT approach implemented will increase the accessibility of the algorithm for non-experts, removing the necessity of knowledge on coding and / or image analysis. From here, investigation on the understanding of $\Lambda$ with respect to the data that is utilised will be crucial. Although a relationship has been shown between signal intensity and $\Lambda$, this has to be further developed.

The SNT approach, although specifically utilised in SPIDER within the framework of image deconvolution of fluorophores, can be extended to other deconvolution methods, such in the field of astronomy, where galaxies and star clusters are localised by means of deconvolution. Also in the field of seismology, where 1D-deconvolution is applied to find certain reading of the vibrations in the earth. This approach can be extended to any deconvolution problem, opening the gates for the use of the $L_0$-norm. There is no known method for $\Lambda$-optimisation for this norm specifically, as such the development of the SNT approach will prove to be crucial to understanding this specific norm.

Within deconvolution there is the side-track called blind-deconvolution, whereby the convolution matrix is estimated. This was not possible previously, as there was no criterion for the information on the correctness of the solution. However the SNT approach will be able to make use of this, as it can try to fit a range of convolution matrices, whereby the matrix that utilises the least amount of points to reproduce the data in the best way can be determined.

The SNT approach has a broad range of possibilities to be further developed.
7 References

[1] An introduction to optical super-resolution microscopy for the adventurous biologist by Vangindertael J - DOI:10.1088/2050-6120/aaae0c


[16] Regularization and variable selection via the elastic net by Hui Zou and Trevor Hastie - 2005 Royal Statistical Society


[18] L1 regularization in electrical impedance tomography by using the L1-curve (Pareto frontier curve) by Nasehi Tehrani, J. et al. - Applied Mathematical Modelling March 2012


[22] http://bigwww.epfl.ch/smlm/index.html#&panel1-8 - Visited on 02-07-2018


8 Appendix

8.1 Comparative study on $\Lambda$-optimisation approaches

Two approaches (GCV, L-curve) to $\Lambda$-optimisation are compared against the SNT approach. The two approaches are applied on the L$_2$-norm, as such, to keep the comparison fair, all three approaches will be applied on the L$_{2,1,0}$-norms. The approaches will not be discussed and can be found in [14 25]. The optimal $\Lambda$ will be determined through two ways, for the SNT / GCV approach, a clear minimum should be seen, which will correspond to the optimal value determined by the approach. However for the L-curve, the step-size between points is taken and from here a minimum should be observed. This minimum will correspond to the optimal $\Lambda$-value.

To keep the comparison concise, a singular 1D example will be shown. A data-line of a 100 points with 5 emitters is convolved with a Gaussian distribution and the various norms are applied on a range of $\Lambda$-values. The curve of optimisation are displayed with point and is not a line, it would seem like that, but that is due to the very small steps that were taken in the $\Lambda$-range on which the approaches were applied.

8.1.1 L$_2$-norm

Starting with the L$_2$-norm, the data-line is on the top left, with while the approaches to $\Lambda$-optimisation is displayed below. One can notice that both the GCV and Step-size show similar curves, the SNT approach does not. The SNT approach seems to shift its minimum towards a lower $\Lambda$-value. In the next figure the three approaches are more closely investigated, whereby the minimum is highlighted with a vertical blue line. This shows that the SNT approach does not give the same or similar value as the other two approaches. This also seen in the solution, as it is a lot more varied.

**SI Figure 1: L$_2$-norm on various optimisation approaches**
8.1.2 \textbf{L}\textsubscript{1}-norm

The same was applied on the \textbf{L}\textsubscript{1}-norm, from this one observes a collapse of the Step-size approach, while the GCV has two local minima. Only the SNT approach seems to have retained its curve.

However when observing more closely in SI figure 4, one can see the collapse of the GCV approach, while the Step-size gives a seemingly appropriate \(\Lambda\)-value. However this could be due to the fact that a too small range of \(\Lambda\) has been chosen, as it would seem that going to a lower \(\Lambda\)-range could give a new minimum. This is not the case for the SNT approach, however, if a bigger steps in the \(\Lambda\)-range were taken, the \(\Lambda_{\text{min}}\) found by the approach could be missed.
8.1.3 $L_0$-norm

Ending this comparative study with the $L_0$-norm, a similar collapse is observed with the Step-size and also with the GCV, like with the $L_1$-norm a second local minimum is seen. The $\Lambda_{\text{min}}$ values for each approach look completely different and this is also seen when further investigated with the next figure (5).

One can see that looking at the solutions for the various $\Lambda_{\text{min}}$ values that were determined by the approaches, that the SNT approach localises the emitters perfectly. This is not the case for the other two approaches.
**SI Figure 6: Comparison of $\Lambda$-optimisation for $L_0$**

As a safety measure, the test on the $L_0$-norm, was repeated several times, and showed similar curves.

**SI Figure 7: Repeated testing of $\Lambda$-optimisation for $L_0$**

From this we can conclude that it is unlikely that utilising the two approaches with the $L_0$-norm will result in appropriate solutions.