Support Inference in linear Statistical Inverse Problems

Frank Werner¹

Statistical Inverse Problems in Biophysics Group
Max Planck Institute for Biophysical Chemistry, Göttingen

and

Felix Bernstein Institute for Mathematical Statistics in the Biosciences
University of Göttingen

Applied Inverse Problems Conference 2017, Hangzhou

¹joint work with Katharina Proksch and Axel Munk
Outline

1. Introduction
2. Theory
3. Simulations and real data example
4. Conclusion
Outline

1 Introduction

2 Theory

3 Simulations and real data example

4 Conclusion
Setting and goal

• $\mathcal{X}$ and $\mathcal{Y}$ Hilbert spaces
• $\{\varphi_i\}_{i \in \mathbb{N}} \subset \mathcal{X}$ a dictionary
• $T : \mathcal{X} \to \mathcal{Y}$ bounded and linear

Can we identify the active components (\(\hat{\text{support}}\)) of an unknown $f$ from noisy measurements of $Tf$?
Setting and goal

- $\mathcal{X}$ and $\mathcal{Y}$ Hilbert spaces
- $\{\varphi_i\}_{i \in \mathbb{N}} \subset \mathcal{X}$ a dictionary
- $T : \mathcal{X} \rightarrow \mathcal{Y}$ bounded and linear

Can we identify the active components (\(\hat{\text{\{support\}}\)) of an unknown $f$ from noisy measurements of $Tf$?

More precisely:

Given noisy measurements $Y \approx Tf$ we want to generate a set $\mathcal{I}_a$ such that at a controlled error level all $i \in \mathcal{I}_a$ satisfy $\langle \varphi_i, f \rangle_{\mathcal{X}} > 0$
Support inference?

Special situation: Suppose that

- $\mathcal{X}$ is a space of functions (i.e. $\mathcal{X} = L^2(\Omega)$),
- and the functions $\varphi_i$ have compact support (e.g. wavelet dictionary)

Then:

$$\langle \varphi_i, f \rangle_\mathcal{X} > 0 \quad \Rightarrow \quad f|_{\text{supp}(\varphi_i)} \neq 0$$

Consequently, we obtain information about the support of $f$!
Introduction

Related problems and methods

Suppose $f$ is sparse w.r.t. $\{\varphi_i\}_{i \in \mathbb{N}}$. Then for the recovery of $f$ many methods are used:

- $\ell^1$-penalized Tikhonov regularization / LASSO

$$
\hat{f}_\alpha = \arg\min_{f \in \mathcal{X}} \left[ \| Tf - Y \|_Y^2 + \alpha \sum_{i=1}^{\infty} |\langle \varphi_i, f \rangle_\mathcal{X}| \right]
$$

- Residual method / Danzig selector

$$
\hat{f}_\alpha = \arg\min_{f \in \mathcal{X}} \sum_{i=1}^{\infty} |\langle \varphi_i, f \rangle_\mathcal{X}| \quad \text{subject to} \quad \| Tf - Y \|_Y \leq \rho
$$

- Orthogonal matching pursuit

- ...

But none of these methods can identify the true support at a controlled error level!
Inference for a single $i$:

- compute a function $\Phi_i$ such that $\langle \Phi_i, Tf \rangle_Y = \langle \varphi_i, f \rangle_X$, i.e. $\varphi_i = T\Phi_i$
Inference for a single $i$:

- compute a function $\Phi_i$ such that $\langle \Phi_i, Tf \rangle_Y = \langle \varphi_i, f \rangle_X$, i.e. $\varphi_i = T^* \Phi_i$
- compute the (asymptotic) distribution of $\langle \Phi_i, Y \rangle_Y$ (test statistic) under the hypothesis $\langle \varphi_i, f \rangle_X = 0$
Methodology

Inference for a single $i$:

- compute a function $\Phi_i$ such that $\langle \Phi_i, Tf \rangle_Y = \langle \varphi_i, f \rangle_X$, i.e. $\varphi_i = T^* \Phi_i$
- compute the (asymptotic) distribution of $\langle \Phi_i, Y \rangle_Y$ (test statistic) under the hypothesis $\langle \varphi_i, f \rangle_X = 0$
- with the $(1 - \alpha)$-quantile $q_{1-\alpha}$ of this (asymptotic) distribution it holds under $\langle \varphi_i, f \rangle_X = 0$ (asymptotically)

$$P \left[ \langle \Phi_i, Y \rangle_Y \geq q_{1-\alpha} \right] \leq \alpha$$
Methodology

Inference for a single $i$:

- compute a function $\Phi_i$ such that $\langle \Phi_i, Tf \rangle_Y = \langle \varphi_i, f \rangle_X$, i.e. $\varphi_i = T^* \Phi_i$
- compute the (asymptotic) distribution of $\langle \Phi_i, Y \rangle_Y$ (test statistic) under the hypothesis $\langle \varphi_i, f \rangle_X = 0$
- with the $(1 - \alpha)$-quantile $q_{1-\alpha}$ of this (asymptotic) distribution it holds under $\langle \varphi_i, f \rangle_X = 0$ (asymptotically)

$$\mathbb{P} \left[ \langle \Phi_i, Y \rangle_Y \geq q_{1-\alpha} \right] \leq \alpha$$

- consequently if $\langle \Phi_i, Y \rangle_Y \geq q_{1-\alpha}$ we have $\langle \varphi_i, f \rangle_X > 0$ with probability $\geq 1 - \alpha$. 
**Methodology (cont’)**

Inference for a all $i$:

- if we infer for each $i$ individually, the multiplicity adjustment makes the statements weak (if the statements are true for each single $i$ with probability 90%, then they hold true for two $i$ at the same time only with probability 81% etc.)
Methodology (cont’)

Inference for all $i$:

- if we infer for each $i$ individually, the multiplicity adjustment makes the statements weak (if the statements are true for each single $i$ with probability 90%, then they hold true for two $i$ at the same time only with probability 81% etc.)

$\Rightarrow$ remedy: simultaneous testing

- consider the test statistic

$$M = \max_i \left[ w_i \left( \frac{\langle \Phi_i, Y \rangle_Y}{\sqrt{V[\langle \Phi_i, Y \rangle_Y]}} - w_i \right) \right].$$

- compute the asymptotic distribution of $M$ under $f \equiv 0$ and its $(1 - \alpha)$-quantile $q_{1-\alpha}$

- mark each $i$ for which $\langle \Phi_i, Y \rangle_Y > (q_{1-\alpha}/w_i + w_i) \sqrt{V[\langle \Phi_i, Y \rangle_Y]}$ as active
Methodology (cont’)

- by taking the max in the statistic the statements become uniform in $i$:

$$\mathcal{I}_a := \left\{ i \bigg| \langle \Phi_i, Y \rangle_Y > \left( \frac{q_{1-\alpha}}{w_i} + w_i \right) \sqrt{\text{V}[\langle \Phi_i, Y \rangle_Y]} \right\}$$

satisfies (asymptotically)

$$\mathbb{P}[\langle \varphi_i, f \rangle_X > 0 \text{ for all } i \in \mathcal{I}_a] \geq 1 - \alpha.$$  

\implies we can infer on all $i$ at a controlled level!
Methodology (cont’)

- by taking the max in the statistic the statements become uniform in $i$:

$$\mathcal{I}_a := \left\{ i \mid \langle \Phi_i, Y \rangle_Y > \left( \frac{q_{1-\alpha}}{w_i} + w_i \right) \sqrt{\mathbb{V} \left[ \langle \Phi_i, Y \rangle_Y \right]} \right\}$$

satisfies (asymptotically)

$$\mathbb{P} [ \langle \varphi_i, f \rangle_X > 0 \text{ for all } i \in \mathcal{I}_a ] \geq 1 - \alpha.$$  

=>$\Rightarrow$ we can infer on all $i$ at a controlled level!

Are the quantiles $q_{1-\alpha}$ well-defined? How to compute them?
Outline

1. Introduction
2. Theory
3. Simulations and real data example
4. Conclusion
Specific setting

- \( \mathcal{Y} = L^2([0, 1]^d) \), discrete measurements: \( Y_j = (Tf)(x_j) + \xi_j \), 
  \( j \in \{1, \ldots, n\}^d \)
Specific setting

- \( \mathcal{Y} = L^2([0, 1]^d) \), discrete measurements: \( Y_j = (Tf)(x_j) + \xi_j \), \( j \in \{1, \ldots, n\}^d \)
- \( \xi_j \) are independent, centered (\( \mathbb{E}[\xi_j] = 0 \)) and satisfy a moment condition (especially all moments need to exist)
Specific setting

- $\mathcal{Y} = L^2([0, 1]^d)$, discrete measurements: $Y_j = (Tf)(x_j) + \xi_j$, $j \in \{1, ..., n\}^d$
- $\xi_j$ are independent, centered ($\mathbb{E} [\xi_j] = 0$) and satisfy a moment condition (especially all moments need to exist)
- the dictionary has at most $N = N(n)$ elements $\varphi_i$ which satisfy $\varphi_i = T^*\Phi_i$, and there is a transformed mother wavelet $\Phi$ such that
  \[
  \{\Phi_i\} = \left\{ \Phi \left( \frac{\cdot - t_i}{h_i} \right) \mid 1 \leq i \leq N(n) \right\}
  \]
  with scales $h_i \in [0, 1]^d$ and positions $t_i \in [0, 1]^d$
Specific setting

- $Y = L^2([0, 1]^d)$, discrete measurements: $Y_j = (Tf)(x_j) + \xi_j$, $j \in \{1, \ldots, n\}^d$
- $\xi_j$ are independent, centered ($\mathbb{E}[\xi_j] = 0$) and satisfy a moment condition (especially all moments need to exist)
- the dictionary has at most $N = N(n)$ elements $\varphi_i$ which satisfy $\varphi_i = T^*\Phi_i$, and there is a transformed mother wavelet $\Phi$ such that
  $$\{\Phi_i\} = \left\{\Phi\left(\frac{\cdot - t_i}{h_i}\right) \mid 1 \leq i \leq N(n)\right\}$$

  with **scales** $h_i \in [0, 1]^d$ and **positions** $t_i \in [0, 1]^d$
- the function $\Phi$ has compact support in $[0, 1]^d$
Test statistic

- $\langle \Phi_i, Y \rangle_Y$ is not available, approximate it by

$$\langle \Phi_i, Y \rangle_n := n^{-d} \sum_{j \in \{1, \ldots, n\}^d} Y_j \Phi_i(x_j)$$
Test statistic

- $\langle \Phi_i, Y \rangle_Y$ is not available, approximate it by

$$\langle \Phi_i, Y \rangle_n := n^{-d} \sum_{j \in \{1, \ldots, n\}^d} Y_j \Phi_i(x_j)$$

- the variance $\sigma_i^2 := \text{Var}\left[\langle \Phi_i, Y \rangle_Y\right]$ might also be unknown, consider a family of uniformly consistent estimators $\hat{\sigma}_i^2$
Test statistic

• \( \langle \Phi_i, Y \rangle_Y \) is not available, approximate it by

\[
\langle \Phi_i, Y \rangle_n := n^{-d} \sum_{j \in \{1, \ldots, n\}^d} Y_j \Phi_i(x_j)
\]

• the variance \( \sigma_i^2 := \text{Var} \left[ \langle \Phi_i, Y \rangle_Y \right] \) might also be unknown, consider a family of uniformly consistent estimators \( \hat{\sigma}_i^2 \)

• we have to investigate the asymptotic distribution of

\[
S(Y) := \max_i \left[ w_i \left( \frac{\langle \Phi_i, Y \rangle_n}{\hat{\sigma}_i} - w_i \right) \right]
\]
Test statistic

- $\langle \Phi_i, Y \rangle$ is not available, approximate it by
  \[
  \langle \Phi_i, Y \rangle_n := n^{-d} \sum_{j \in \{1, \ldots, n\}^d} Y_j \Phi_i(x_j)
  \]

- the variance $\sigma_i^2 := \mathbb{V} \left[ \langle \Phi_i, Y \rangle \right]$ might also be unknown, consider a family of uniformly consistent estimators $\hat{\sigma}_i^2$

- we have to investigate the asymptotic distribution of
  \[
  S(Y) := \max_i \left[ w_i \left( \frac{\langle \Phi_i, Y \rangle_n}{\hat{\sigma}_i} - w_i \right) \right]
  \]

- the calibration values are only scale-dependent and chosen as
  \[
  w_i = \sqrt{2 \log \left( \frac{C_{\Phi_i}}{\prod h_i} \right)} + C_d \frac{\log \left( \sqrt{2 \log \left( \frac{C_{\Phi_i}}{\prod h_i} \right)} \right)}{\sqrt{2 \log \left( \frac{C_{\Phi_i}}{\prod h_i} \right)}}
  \]
Gaussian Approximation

Suppose that

- there are only polynomially many probe functions in the dictionary \( N(n) \leq n^\kappa \) for some \( \kappa > 0 \)
- the smallest scale tends to zero not too fast \( \min_i \min_{\text{entries}} h_i \geq \log(n)^p / n \) with some specific \( p \)
- the largest scale tends to zero sufficiently fast \( \max_i \max_{\text{entries}} h_i \leq n^{-\delta} \) for some \( \delta > 0 \)
Gaussian Approximation

Suppose that

- there are only polynomially many probe functions in the dictionary ($N(n) \leq n^\kappa$ for some $\kappa > 0$)
- the smallest scale tends to zero not too fast ($\min_i \min_{\text{entries}} h_i \geq \log(n)^p/n$ with some specific $p$)
- the largest scale tends to zero sufficiently fast ($\max_i \max_{\text{entries}} h_i \leq n^{-\delta}$ for some $\delta > 0$)

Then there are i.i.d. standard normal random variables $\zeta_j$ such that under $f \equiv 0$ it holds

$$\lim_{n \to \infty} |\mathbb{P} [S(Y) > q] - \mathbb{P} [S(\zeta) > q]| = 0$$

for all $q$. 
Gaussian Approximation (cont’)

Continuous Gaussian Approximation

There exists a Brownian sheet $W$ such that $S(Y)$ can be approximated

$$S(W) := \max_i \left[ w_i \left( \frac{\int \Phi_i(x) \, dW_x}{\|\Phi_i\|_{L^2}} - w_i \right) \right],$$

i.e. under $f \equiv 0$ it holds

$$\lim_{n \to \infty} |\mathbb{P}[S(Y) > q] - \mathbb{P}[S(W) > q]| = 0 \quad \text{for all } q.$$

Moreover

- $S(W)$ is a.s. bounded from below and above,
- $S(W)$ is asymptotically non-degenerate, i.e. does not concentrate to any point,
- $S(W)$ does not depend on any unknown quantities.
Gaussian Approximation - Implications

This means that we can use quantiles $q_{1-\alpha}$ from

$$S(\zeta) = \max_i \left[ w_i \left( \frac{\langle \Phi_i, \zeta \rangle_n}{\| \Phi_i \|_2} - w_i \right) \right]$$

to hold the asymptotic level.

- $S(\zeta)$ is 'distribution free', i.e. it depends only on known quantities
- quantiles can be simulated easily
- quantiles are meaningful as $n \to \infty$
Gaussian Approximation - Implications

This means that we can use quantiles $q_{1-\alpha}$ from

$$S(\zeta) = \max_i \left[ w_i \left( \frac{\langle \Phi_i, \zeta \rangle_n}{\|\Phi_i\|_2} - w_i \right) \right]$$

to hold the asymptotic level.

- $S(\zeta)$ is 'distribution free', i.e. it depends only on known quantities
- quantiles can be simulated easily
- quantiles are meaningful as $n \to \infty$

If $\langle \Phi_i, Y \rangle_n > \left( \frac{q_{1-\alpha}}{w_i} + w_i \right) \hat{\sigma}_i$ mark $i$ as active (i.e. $i \in I_a$)!
Detection properties

So far: whenever $i \in \mathcal{I}_a$, then asymptotically $\mathbb{P} [\langle \varphi_i, f \rangle \chi > 0] \geq 1 - \alpha$. 
Detection properties

So far: whenever $i \in I_a$, then asymptotically $\mathbb{P} [\langle \varphi_i, f \rangle_X > 0] \geq 1 - \alpha$.

But how large must $\langle \varphi_i, f \rangle_X$ be to be detected?
Detection properties

So far: whenever $i \in I_a$, then asymptotically $\mathbb{P} [ \langle \varphi_i, f \rangle_X > 0 ] \geq 1 - \alpha$.

But how large must $\langle \varphi_i, f \rangle_X$ be to be detected?

Lower detection bound

If $\langle \varphi_i, f \rangle_X \geq 2 \left( \frac{q_{1-\alpha}}{w_i} + w_i \right) \sigma_i$, then

$$\lim_{n \to \infty} \mathbb{P} [ i \in I_a ] \geq 1 - \alpha$$

uniformly in $i$. 
Special case: deconvolution

- Suppose $d = 2$ and $T$ is a convolution operator, i.e.

$$
(Tf)(y) = (k * f)(y) := \int_{[0,1]^2} k(x - y)f(y) \, dy.
$$

This implies that if we choose $\{\psi_i\}$ of Wavelet-type we obtain $\{\Phi_i\} = \{\hat{\Phi}_h(\cdot - t_ih) \mid 1 \leq i \leq N(n)\}$ where $\hat{\Phi}_h$ depends on $h$.

- Suppose the Fourier transform of the kernel $k$ has a polynomial decay, i.e.

$$c \left( 1 + \|\xi\|^2 \right)^{-a} \leq |F_k(\xi)| \leq C \left( 1 + \|\xi\|^2 \right)^{-a}.
$$

This implies that the functions $\psi_i$ can be chosen such that they are non-negative and have compact support, and $\|\hat{\Phi}_h\|_{L^2}$ behaves like $\max_h \|\cdot\|^a$.
Special case: deconvolution

- Suppose $d = 2$ and $T$ is a convolution operator, i.e.

$$
(Tf)(y) = (k * f)(y) := \int_{[0,1]^2} k(x - y) f(y) \, dy.
$$

This implies that if we choose $\{\varphi_i\}$ of Wavelet-type we obtain

$$
\{\Phi_i\} = \left\{ \tilde{\Phi}_h \left( \frac{\cdot - t_i}{h_i} \right) \mid 1 \leq i \leq N(n) \right\}
$$

where $\tilde{\Phi}_h$ depends on $h$. 
Special case: deconvolution

- Suppose $d = 2$ and $T$ is a convolution operator, i.e.

\[
(Tf)(y) = (k * f)(y) := \int_{[0,1]^2} k(x - y) f(y) \, dy.
\]

This implies that if we choose $\{\varphi_i\}$ of Wavelet-type we obtain

\[
\{\Phi_i\} = \left\{ \tilde{\Phi}_h \left( \frac{\cdot - t_i}{h_i} \right) \mid 1 \leq i \leq N(n) \right\}
\]

where $\tilde{\Phi}_h$ depends on $h$.

- Suppose the Fourier transform of the kernel $k$ has a polynomial decay, i.e.

\[
c(1 + \|\xi\|_2^2)^{-a} \leq |\mathcal{F}k(\xi)| \leq \overline{c}(1 + \|\xi\|_2^2)^{-a}.
\]
Special case: deconvolution

- Suppose $d = 2$ and $T$ is a convolution operator, i.e.

$$ (Tf)(y) = (k * f)(y) := \int_{[0,1]^2} k(x - y)f(y) \, dy. $$

This implies that if we choose $\{\varphi_i\}$ of Wavelet-type we obtain

$$ \{\Phi_i\} = \left\{ \tilde{\Phi}_h \left( \frac{\cdot - t_i}{h_i} \right) \mid 1 \leq i \leq N(n) \right\} $$

where $\tilde{\Phi}_h$ depends on $h$.

- Suppose the Fourier transform of the kernel $k$ has a polynomial decay, i.e.

$$ c(1 + \|\xi\|^2_2)^{-a} \leq |\mathcal{F}k(\xi)| \leq C(1 + \|\xi\|^2_2)^{-a}. $$

This implies that the functions $\varphi_i$ can be chosen such that they are non-negative and have compact support, and $\|\tilde{\Phi}_h\|_{L^2}$ behaves like $\max_{\text{entries}} h^{2a}$. 
Special case: deconvolution (cont’)

- $S(Y)$ can be approximated by a Gaussian version which is a.s. bounded and non-degenerate

$\Rightarrow$ whenever $i \in I_a$, then asymptotically $\mathbb{P} [\langle \varphi_i, f \rangle_X > 0] \geq 1 - \alpha$.

- If $\langle \varphi_i, f \rangle_X$ is sufficiently large, then $i$ will be detected with probability $\geq 1 - \alpha$. 
Special case: deconvolution (cont’)

- $S(Y)$ can be approximated by a Gaussian version which is a.s. bounded and non-degenerate

\[ \Rightarrow \quad \text{whenever } i \in I_a, \text{ then asymptotically } \Pr[\langle \varphi_i, f \rangle_X > 0] \geq 1 - \alpha. \]

- If $\langle \varphi_i, f \rangle_X$ is sufficiently large, then $i$ will be detected with probability $\geq 1 - \alpha$.

**Optimality of the lower detection bound**

In $d = 1$ this lower detection bound is optimal in the sense that no estimator for a $\beta$-Hölder-continuous function $f$ can distinguish between $f|_{[t,t+h]} = 0$ and $f|_{[t,t+h]} \geq h^\beta$ at a faster rate in $h = h(n)$. 
Outline

1 Introduction

2 Theory

3 Simulations and real data example

4 Conclusion
**Considered problem**

- $T$ deconvolution problem, i.e.

\[ Y_j = (k * f)(x_j) + \xi_j, \quad j \in \{1, ..., n\}^2 \]

with an equidistant grid $\{x_j\}$ on $[0, 1]^2$.
Considered problem

- **T** deconvolution problem, i.e.
  
  \[ Y_j = (k \ast f)(x_j) + \xi_j, \quad j \in \{1, \ldots, n\} \]

  with an equidistant grid \( \{x_j\} \) on \([0, 1]^2\).

- The kernel \( k \) is chosen from the family
  
  \[ (\mathcal{F}k_{a, b})(\xi) = (1 + b^2 \|\xi\|^2_2)^{-a}, \quad \xi \in \mathbb{R}^2. \]

Testfunction \( f \)
Considered problem

- $T$ deconvolution problem, i.e.

$$Y_j = (k * f)(x_j) + \xi_j, \quad j \in \{1, \ldots, n\}^2$$

with an equidistant grid $\{x_j\}$ on $[0, 1]^2$.

- The kernel $k$ is chosen from the family

$$(\mathcal{F}k_{a,b})(\xi) = (1 + b^2 \|\xi\|_2^2)^{-a}, \quad \xi \in \mathbb{R}^2.$$  

- The variance is considered to be known.
Considered problem

- $T$ deconvolution problem, i.e.

$$Y_j = (k * f)(x_j) + \xi_j, \quad j \in \{1, \ldots, n\}^2$$

with an equidistant grid $\{x_j\}$ on $[0, 1]^2$.

- The kernel $k$ is chosen from the family

$$(\mathcal{F}k_{a,b})(\xi) = (1 + b^2 \|\xi\|_2^2)^{-a}, \quad \xi \in \mathbb{R}^2.$$ 

- The variance is considered to be known.

- The mother wavelet $\varphi$ is chosen to minimize the variance $\|\Phi\|_{L^2}^2$

($\rightsquigarrow$ Tensor product of Beta-Kernels)
Some empirical levels for $\alpha = 0.1$

<table>
<thead>
<tr>
<th>Noise scenario</th>
<th>Parameters</th>
<th>false positives %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian noise</td>
<td></td>
<td>8.8</td>
</tr>
<tr>
<td>$\nu = 3$</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>$\nu = 6$</td>
<td></td>
<td>94.7</td>
</tr>
<tr>
<td>$\nu = 7$</td>
<td></td>
<td>72.3</td>
</tr>
<tr>
<td>Student’s t noise</td>
<td>$\nu = 11$</td>
<td>21.8</td>
</tr>
<tr>
<td>$\nu = 15$</td>
<td></td>
<td>15.7</td>
</tr>
<tr>
<td>$\nu = 19$</td>
<td></td>
<td>13.0</td>
</tr>
<tr>
<td>$\nu = 23$</td>
<td></td>
<td>13.3</td>
</tr>
<tr>
<td>CCD noise (Sneyder ’93, ’95):</td>
<td>$t = 100, b = 0.5, \sigma = 0.01$</td>
<td>9.8</td>
</tr>
<tr>
<td>obs. time $t$, background $b$, read-out errors $\mathcal{N}(0, \sigma^2)$</td>
<td>$t = 1000, b = 0.005, \sigma = 0.01$</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>$t = 100, b = 0.005, \sigma = 0.01$</td>
<td>14.5</td>
</tr>
</tbody>
</table>
Support recovery - result

data \((\sigma = 0.5)\)
Support recovery - result

data ($\sigma = 0.05$)  

90% significance map
**Support recovery - result**

**Exact solution**

**Data (σ = 0.005)**

**90% significance map**

**Zoom**
Real data example - Setup

- we analyze fluorescent dyes on single DNA Origami
- imaging is performed by a STED microscope
- each of the two strands can at most hold 12 markers
Modeling

The observations can perfectly modeled by

\[ Y_j \overset{\text{independent}}{\sim} \text{Bin}(t, (k \ast f)(x_j)), \quad j \in \{1, \ldots, n\}^2. \]

- Bin \((t, p)\): Binomial distribution with parameters \(t \in \mathbb{N}\) and \(p \in [0, 1]\)
- \(f(x)\): probability that a photon emitted at grid point \(x\) is recorded at the detector in a single excitation pulse
Result
Comparison of the result with the data
Outline

1 Introduction

2 Theory

3 Simulations and real data example

4 Conclusion
Conclusions

Outlook

- Methodology and theory:
  - inference on active coefficients $\langle \varphi_i, f \rangle_X$ w.r.t. a dictionary $\{\varphi_i\}$
  - techniques from multiscale testing yield uniform inference at a controlled (asymptotic) error level
  - detection power is optimal in a suitable sense

- Application:
  - the method can be used to determine the support of a function observed in a convolution model
  - performs well in a real data example
Outlook

- **Methodology and theory:**
  - inference on active coefficients $\langle \varphi_i, f \rangle \chi$ w.r.t. a dictionary $\{\varphi_i\}$
  - techniques from multiscale testing yield uniform inference at a controlled (asymptotic) error level
  - detection power is optimal in a suitable sense

- **Application:**
  - the method can be used to determine the support of a function observed in a convolution model
  - performs well in a real data example

Thank you for your attention!