NMF in LM

Introduction

Localisation microscopy (LM) is a conceptually simple and accessible technique providing super-resolution fluorescent images.[5,7,9,13] The structure of the sample is reconstructed by localising individual fluorophores with precision surpassing[14] the classical resolution limit $\delta = \frac{\lambda}{2\mathrm{NA}}$, where λ is the wavelength of the emission light and NA is the numerical aperture of the objective lens. LM makes use of the fluorophore transition between bright (ON) and dark (OFF) states to discriminate individual sources separated by a distance $d < \delta$. The super-resolution image is achieved by a repetitive localisation of the different individual spatially separated subsets of fluorophores. The optimal number of ON sources in each acquisition frame must be experimentally estimated.[17] A high density of the ON fluorophores results in overlapping sources and complicates localisation whereas a small density leads to a long acquisition time. Standard LM techniques (fPALM,[5] STORM[16]) control the density of the ON sources by photo-switching. However, some recent concepts suggest algorithms which can deal with a high density of ON sources.[6]

Quantum dots (QD) are an order of magnitude brighter compared to the organic dyes used in conventional LM.[8,15] Under continuous excitation the QDs exhibit a stochastic blinking between ON and OFF states.[18] Excellent photo-stability, low cyto-toxicity and distinctive spectral properties make QDs very attractive for biological research.[15] However, the stochastic blinking of QDs is impractical for standard LM as the rate of ON-OFF transition and hence the density of ON sources is difficult to control. QD labeled data typically consist of highly overlapping sources which cannot be localised with standard LM techniques.

Some concepts exploiting the blinking behaviour of the QDs have been proposed. Maximum a-posteriori (MAP) fitting of the positions and the intensities of known point spread functions (PSFs) to blinking QD data has been proposed.[4] Independent component analysis (ICA) of the QD data was suggested.[12] A resolution improvement by analysis of the intensity fluctuation (SOFI) has been demonstrated.[3]

NMF

We used non-negative matrix factorisation (NMF)[10] as a natural model for QD data. NMF decomposes a movie of the blinking QDs into spatial (time independent images of the individual sources) and temporal parts (fluctuating intensities of each emitter). NMF imposes natural non-negativity constraints on both the spatial (the images of the individual sources) and the temporal (intensities) components. Moreover, we used the NMF algorithm[11] which maximises the likelihood of the model for data corrupted with Poisson noise. This makes NMF preferable to ICA[12] for QD data as ICA allows negative entries in the separated components and does not account for noise in the measured data (fig.A.2).

NMF does not put any additional constraints on the shape (point spread function PSF) or blinking behaviour of the individual sources, apart from being non-negative. Therefore NMF can separate overlapping sources of differing shapes and blinking characteristics. Variability in PSFs can, for example, arise in a 3D sample where fluorophores can be in different focal depths and therefore each exhibiting a different PSF.

Model comparison

The NMF algorithm requires prior knowledge about the number of sources K to be separated. Principal component analysis of the data (PCA) can be used as a simple method for dimensionality estimation. However, for noisy data the estimation of K is difficult.

The results of NMF are the maximum-likelihood estimates and therefore a higher K leads to a higher likelihood (or a lower residual error) as more sources are better able to fit the data. However, overestimating K produces spurious sources. On the other hand, underestimation of K results in incorrect source shapes as one component must fit several sources. Therefore a reliable estimation of K is vital for a successful separation.

The posterior distribution $p(K|\mathbf{D})$ represents evidence for a particular value of K given a data set $\mathbf{D}$. This quantity can be estimated by Bayesian treatment of the probabilistic NMF formulation.[1,2] We can then choose the K which maximises $p(K|\mathbf{D})$. However, a test on simulated data suggests that this approach systematically underestimates K.

We achieved more reliable K estimation by analysing the residuals (data-model). Models with low K gives rise to residual correlations as multiple individual emitters have to be represented with fewer components. Correlations decrease with higher K. We selected the smallest K which reduces the residual correlations to a certain

level.

The ability to estimate K correctly becomes increasingly difficult for closely spaced sources. Simulations of the noisy data comprising two QDs show that NMF can reliably separate two closely spaces emitters ($d > \delta/10$) when $K = K_{\text{true}}$. However, the correct estimation of K is possible only when the two sources are separated at least by $d > \delta/5$, where $\delta = \lambda/2\text{NA}$ is the classical resolution limit.

Out of focus PSF

A Supplementary materials

A movie of blinking quantum dots can be regarded as an $N \times T$ data matrix $\mathbf{D}$ where N is the number of pixels and T is the number of time frames in the movie. Each frame in the movie is transformed into a column of the matrix $\mathbf{D}$ by concatenating columns of the 2D image into a $N \times 1$ vector. Non-negative matrix factorisation (NMF) makes an approximative decomposition

$$\mathbf{D} \approx \mathbf{WH}, \tag{A.1}$$

where the $N \times T$ matrix $\mathbf{D}$ is expressed as a multiplication of the $N \times K$ matrix $\mathbf{W}$ and $K \times T$ matrix $\mathbf{H}$ subject to non-negativity constraints on the entries $w_{nk} \geq 0$ and $h_{kt} \geq 0$. Each column $\boldsymbol{w}_k$ of the matrix $\mathbf{W}$ ($N \times 1$ vector) then represents an image of the kth source and each row $\boldsymbol{h}_k$ of the matrix $\mathbf{H}$ ($1 \times T$ vector) represents the time profile of the kth source intensity.

The NMF algorithm[11] makes the decomposition (A.1) such that the likelihood function of the model is maximised under assumption of Poisson noise. The approximative factorisation (A.1) can then be written as

$$\mathbb{E}[\mathbf{D}] = \mathbf{WH}, \tag{A.2}$$

where $\mathbb{E}[.]$ denotes expectation value of the noisy data with respect to the Poisson distribution.

The gamma-Poisson (GaP) model has been proposed as a probabilistic model for NMF.[2] The entries h_{kt} of the intensity matrix $\mathbf{H}$ are regarded as latent variables generated from a Gamma distribution with parameters $\alpha_k,\ \beta_k$ and the data are modelled as Poisson variable with mean $\mathbf{WH}$ as in (A.2). Variables $\theta = \{\boldsymbol{w}_k, \alpha_k, \beta_k,\ k = 1..K\}$ are then parameters of the GaP model. The variational approximation of the GaP model[1] provides a lower bound $\mathcal{L}$ on the likelihood function $p(\mathbf{D}|K, \theta)$ with latent variables $\boldsymbol{h}_k$ integrated out. We can express the posterior distribution $p(K|\mathbf{D}) \propto p(\mathbf{D}|K)p(K)$ where $p(K)$ is a prior distribution over values of K and $p(\mathbf{D}|K) = \int p(\mathbf{D}|\theta, K)p(\theta|K)d\theta$ is a marginal likelihood.

The number of components K_{true} can also be estimated from the $N \times T$ residual matrix $\boldsymbol{S}$ (entries s_{xt}). After evaluating the model (A.2) for different values of K, we compute a standardised residual matrix with entries

$$s_{nt} = \frac{d_{nt} - \sum_{k=1}^{K} w_{nk}h_{kt}}{\sqrt{\sum_{k=1}^{K} w_{nk}h_{kt}}}.$$

The factor $1/\sqrt{\sum_{k=1}^{K} w_{nk}h_{kt}}$ is applied in order to standardise the residuals of Poisson distributed data. We can then compute the $N \times N$ correlation matrix

$$\boldsymbol{C}_S = \boldsymbol{S}\boldsymbol{S}^T,$$

and the $N \times N$ matrix of the correlation coefficients $\boldsymbol{R}_S$ with entries

$$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}. \tag{A.3}$$

Underestimation of the number of sources ($K < K_{\text{true}}$) will lead to correlations between some pixels as the model will try to explain multiple sources with one component. For $K \geq K_{\text{true}}$ the correlations are expected to drop to a base level and the residuals become uncorrelated.

We used simulated noisy data of two QDs separated by d to asses the performance of NMF when we assumed that the number of sources is known $K = K_{\text{true}}$. Data were simulated with parameters shown in A.1 for 10 random geometrical configuration of two sources separated by d. Extracted components were localised as a maximum likelihood estimator of a gaussian centre $\overline{x}_i$. We computed the mean localisation error per separation

$$\varepsilon = \frac{1}{d}\frac{e_1 + e_2}{2}$$

where $e_i = \left|\overline{x}_i - x_i^{\text{true}}\right|$ corresponds to the localisation error of the ith source located at x_i^{true} and $d = \left|x_1^{\text{true}} - x_2^{\text{true}}\right|$ is the true separation of the emitters. For $d > 30$ nm the mean localisation error is less than 20% of the separation d (fig.A.1).

We used the same data to evaluated the model for $K = K_{\text{true}} \pm 1$ and analysed the residual correlations A.3. The maximum correlation in residuals is shown in fig.A.1. For closely spaced sources ($d \leq 30$ nm) the correlations are approximately on the same level for any K. However, for $d \geq 40$ nm there is a steep increase of the correlations for $K = 1$ compared to $K = 2$. The model with $K = 3$ does not lead to any further improvement. For this noise level we can estimate the correct $K = 2$ for emitters with $d > 50$ nm. This can be the actual limiting factor for the resolution of the method.

Tables

Parameter	Value	Description
T	10^3	Number of time slices in the sequence
K_{true}	10	Number of sources in the simulated data
b	10^2 photons	Uniform background added to each time slice
I_{max}	$1.5 \cdot 10^3$ photons	Maximum intensity of a single source in one time slice
blinking	uniform	Blinking behaviour of the individual sources.
λ_{em}	655 nm	Emission wavelength
NA	1.2	Numerical aperture of the objective
pixel-size	106 nm	Size of a pixel in the sample plane
δ	333 nm (3.1 pixels)	Radius of the region containing the sources ($\delta = 0.61\frac{\lambda_{em}}{NA}$)

Table A.1: Parameters of the simulation

Figures

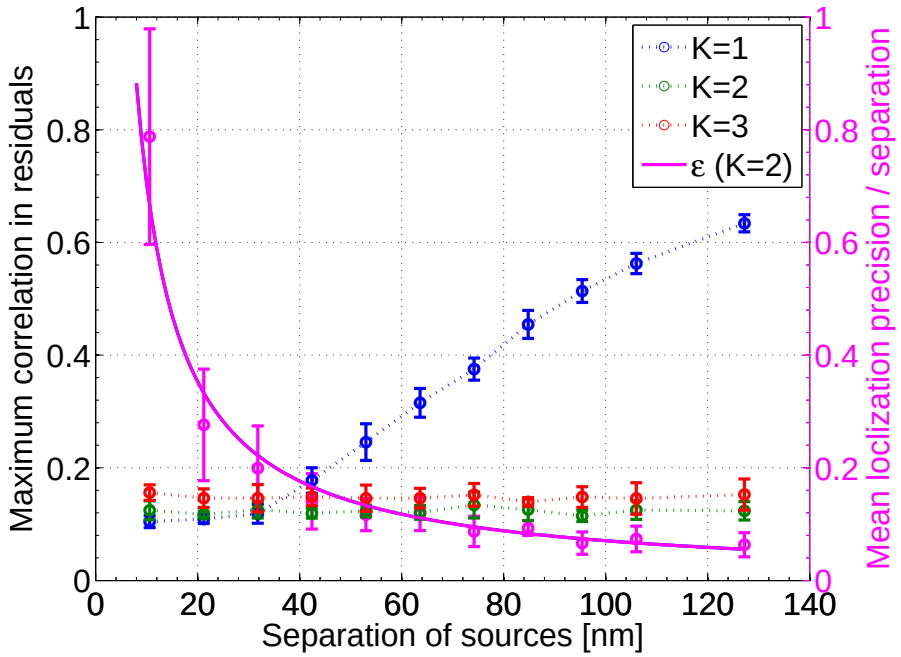
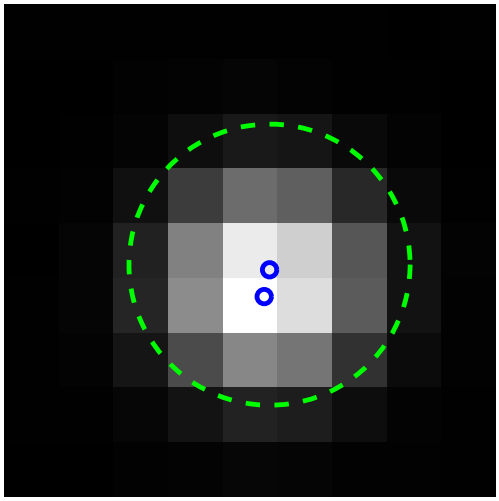
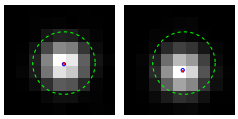


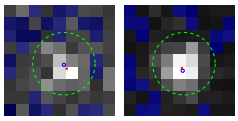
Figure A.1: The maximum value of correlations in residuals for $K = 1$ (blue), 2 (green) and 3 (red). Mean localisation error per separation ε for $K = 2$ (magenta) with fitted curve $\propto 1/d$. Simulated data of two sources ($K_{\text{true}} = 2$). The difference in residual correlations for $K = 1$ and $K = 2$ are apparent only when the two sources are separated by at least 50 nm. The classical resolution limit $\delta = 273$ nm.



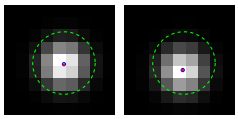
(a) wide field



(b) NMF (noise)

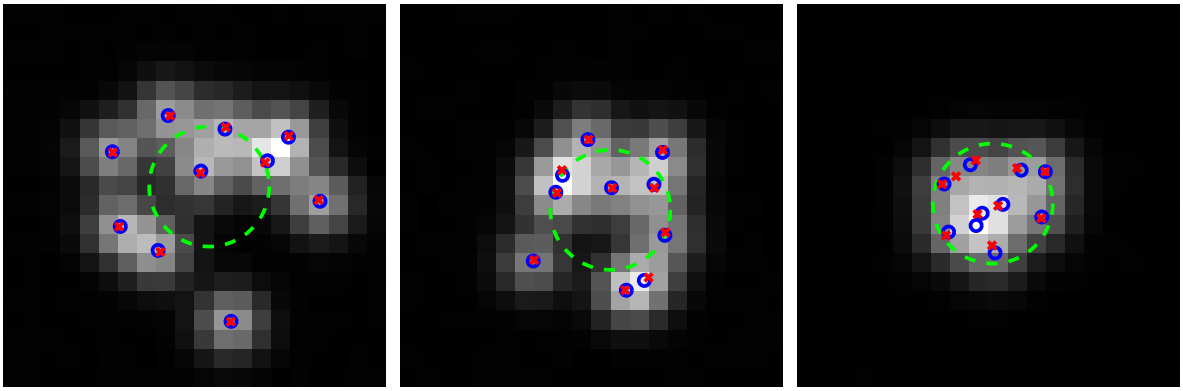


(c) ICA (noise)

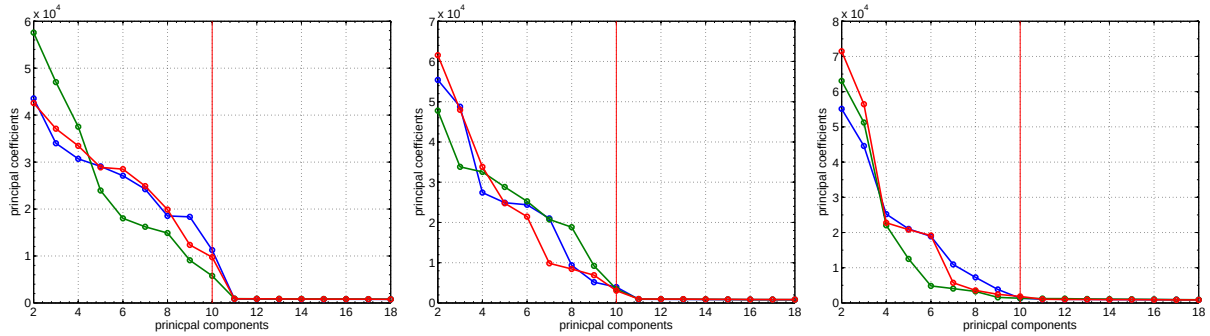


(d) ICA (noise free)

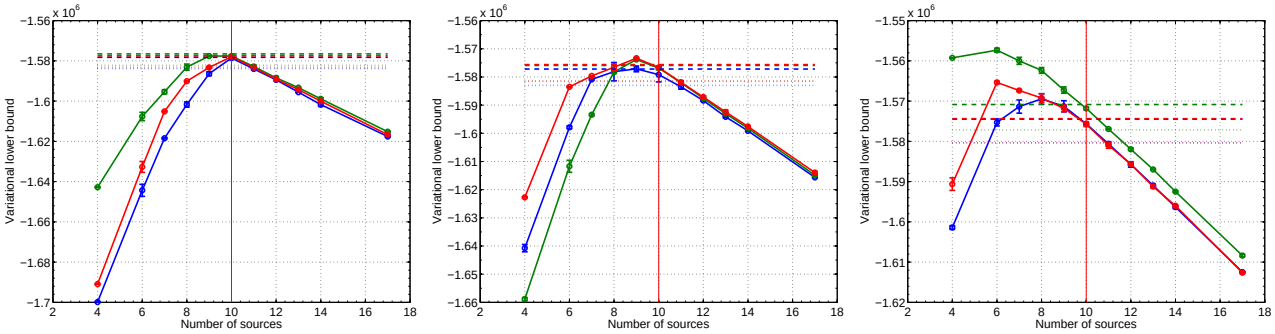
Figure A.2: Comparison of the components separated with NMF (b) and ICA (c) for noisy data of two blinking QDs separated by $d = 50$ nm (mean value shown in (a)). ICA for noise free data shown in (d). The true and the estimated positions are shown as blue circle and red crosses, respectively. The radius of the green circle is the resolution limit δ. Blue pixels contain negative values.



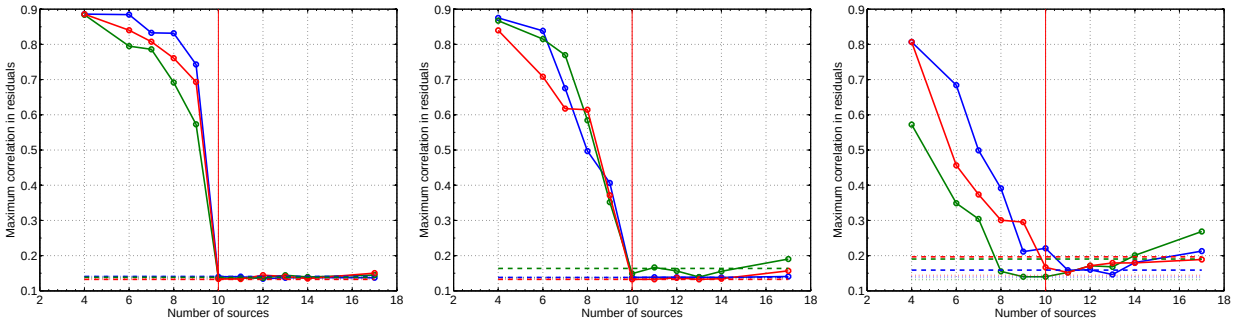
(a) 2δ **(b)** 1.5δ **(c)** δ



(d) 2δ, PCA **(e)** 1.5δ, PCA **(f)** δ, PCA

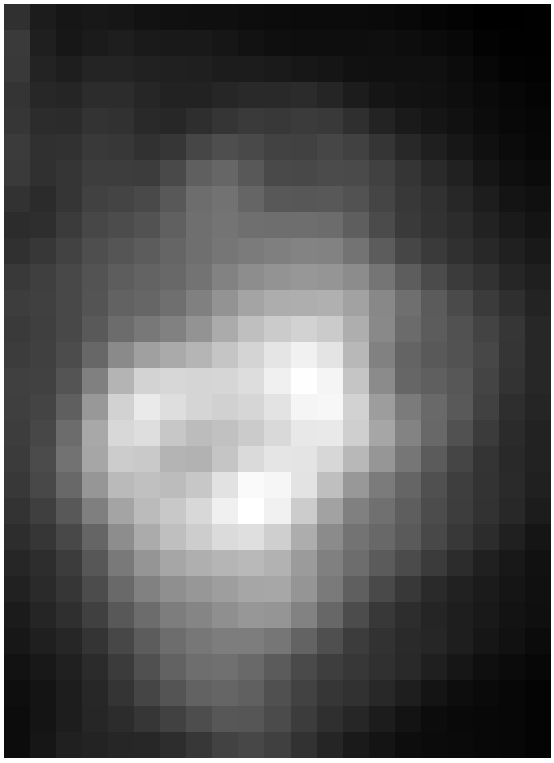


(g) 2δ, lower bound **(h)** 1.5δ, lower bound **(i)** δ, lower bound

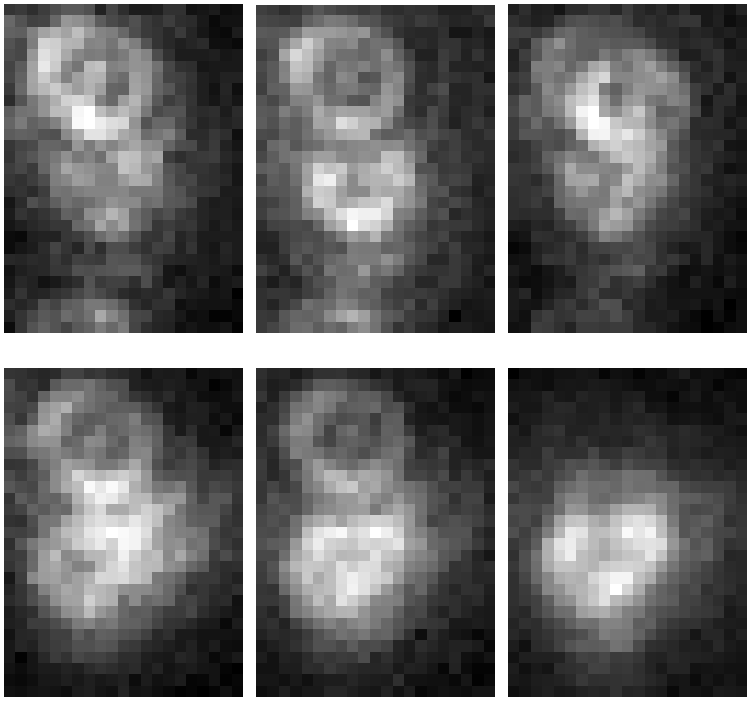


(j) 2δ, maximum correlation **(k)** 1.5δ, maximum correlation **(l)** δ, maximum correlation

Figure A.3: K estimation for sources contained in a circular area with radius 2δ (left column), 1.5δ (middle column) and δ (right column).

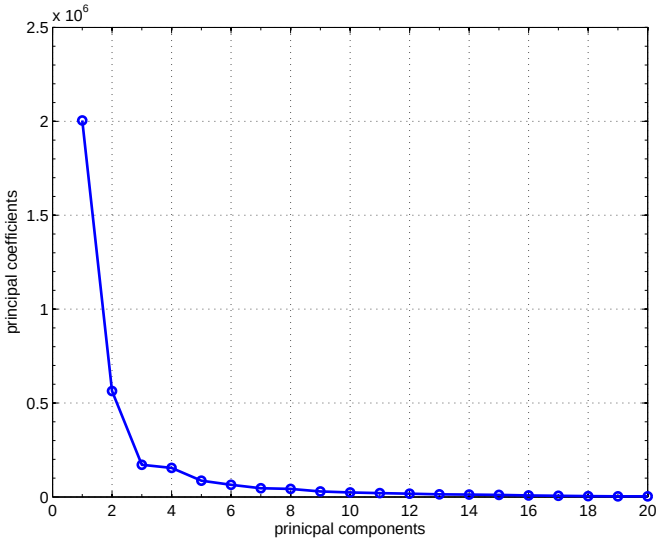


(a) Mean intensity image

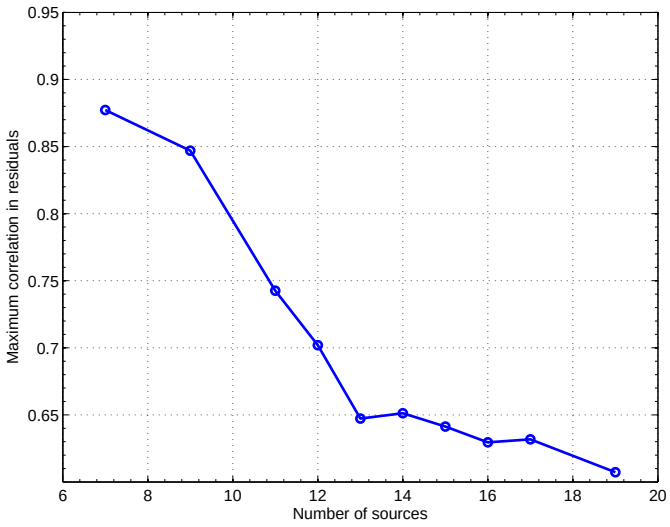


(c) Movie of the blinking sources (six frames out of 10^3)

Figure A.4: Real QD data.

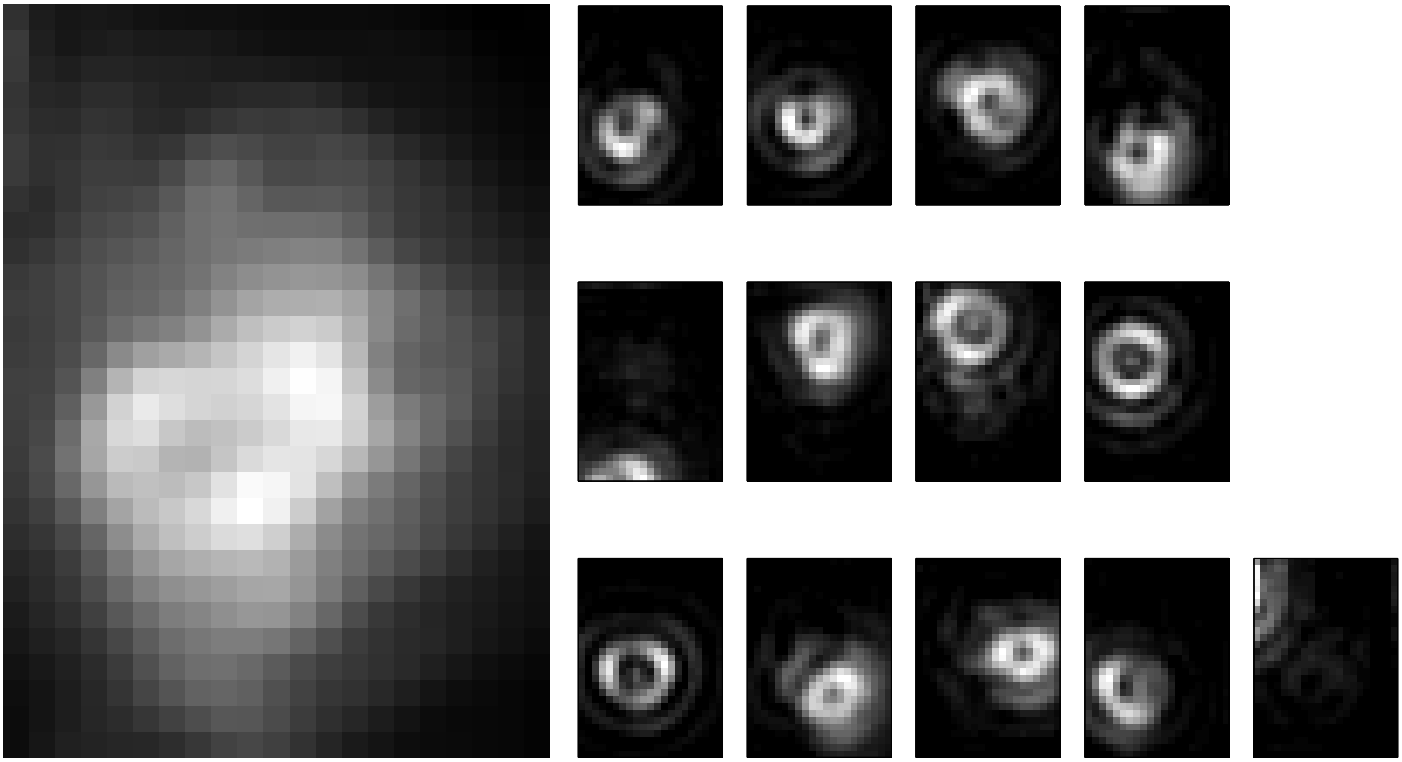


(a) PCA



(b) Correlations

Figure A.5: Estimation of K



(a) Mean intensity image **(b)** Separated sources

Figure A.6: Left: mean intensity image correspond to a standard wide-field image. Right: separated individual sources for $K = 13$.

References

[1] Wray Buntine and Aleks Jakulin. Discrete component analysis. In C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors, *Subspace, Latent Structure and Feature Selection*, pages 1–33. Springer, 2006.

[2] John Canny. GaP: a factor model for discrete data. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 122–129. ACM, 2004.

[3] Thomas Dertinger, Ryan Colyer, Robert Vogel, Jörg Enderlein, and Shimon Weiss. Achieving increased resolution and more pixels with Superresolution Optical Fluctuation Imaging (SOFI). *Optics Express*, 18(18):18875, August 2010.

[4] Paul Harrington, Jonas Anderson, Bernd Rieger, Diane Lidke, and Keith A Lidke. Poster: A Bayesian Approach to Fluorescence Intermittency Based Localization Microscopy. *Supplement of Biophysical Journal*, 96:20–20, 2008.

[5] Samuel T Hess, Thanu P K Girirajan, and Michael D Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–72, December 2006.

[6] Seamus J Holden, Stephan Uphoff, and Achillefs N Kapanidis. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature methods*, 8(4):279–80, April 2011.

[7] Bo Huang, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (New York, N.Y.)*, 319(5864):810–3, February 2008.

[8] Jyoti K Jaiswal and Sanford M Simon. Potentials and pitfalls of fluorescent quantum dots for biological imaging. *Trends in cell biology*, 14(9):497–504, September 2004.

[9] Sara a Jones, Sang-Hee Shim, Jiang He, and Xiaowei Zhuang. Fast, three-dimensional super-resolution imaging of live cells. *Nature methods*, 8(6):499–505, June 2011.

[10] D D Lee and H S Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91, October 1999.

[11] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13, 2001.

[12] Keith A. Lidke, Bernd Rieger, Thomas M. Jovin, and Rainer Heintzmann. Superresolution by localization of quantum dots using blinking statistics. *Optics Express*, 13(18):7052, 2005.

[13] S. Linde, R. Kasper, M. Heilemann, and M. Sauer. Photoswitching microscopy with standard fluorophores. *Applied Physics B*, 93(4):725–731, October 2008.

[14] Raimund J Ober, Sripad Ram, and E Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–200, February 2004.

[15] Ute Resch-Genger, Markus Grabolle, Sara Cavaliere-Jaricot, Roland Nitschke, and Thomas Nann. Quantum dots versus organic dyes as fluorescent labels. *Nature methods*, 5(9):763–75, September 2008.

[16] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–5, October 2006.

[17] Alexander R Small. Theoretical limits on errors and acquisition rates in localizing switchable fluorophores. *Biophysical journal*, 96(2):L16–8, January 2009.

[18] Fernando D. Stefani, Jacob P. Hoogenboom, and Eli Barkai. Beyond quantum jumps: Blinking nanoscale light emitters. *Physics Today*, 62(2):34, 2009.