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 $\delta = 0.61\frac{\lambda}{\mathrm{NA}}$, 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 because 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 because 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 unique 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 $\mathbf{D}$. A lower bound for this quantity can be estimated by Bayesian treatment of the probabilistic NMF formulation.[1,2] We can then choose the K which maximises the lower bound on $p(K|\mathbf{D})$. However, a test on simulated data suggests that this approach systematically underestimates K. (Details in supplementary materials.)

An alternative approach is to analyse the residuals (data-model) for models with different K. Models with low K give rise to correlations in residuals because 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. Simulations suggest that this approach can provide more reliable estimation of K compared to the variational lower bound method. (Details in supplementary materials.)

Results

Test on simulated data with different densities of ten QDs suggests (supplementary materials) that the analysis of the residuals in correlations provides a more accurate estimate of the number of sources K. We used this method in further computations.

Simulations of noisy data comprising two QDs show that NMF can reliably separate two close emitters ($d > \delta/10$) when K_{ture} is known ($K = K_{\text{true}}$). However, the ability to estimate K correctly becomes increasingly difficult for closely spaced sources. The high correlations in residuals for $K < K_{\text{true}}$ arise only when the two sources are separated by at least $d > \delta/5$, where $\delta = 0.61\lambda/\text{NA}$ is the classical resolution limit (fig. A.1).

Out of focus PSF

NMF can extract individual sources with differing shape because apart from the non-negativity there is no further constraints on the separated components. We show a separation of overlapping out-of-focus point spread functions (PSF) from a movie (10^3 frames) of several blinking QDs (fig. A.5). Individual emitters were in slightly different focal depths and therefore each QD produced a distinctive PSF. As can be seen from fig.A.5 individual PSFs were highly overlapping in each frame.

The number of components was estimated from the analysis of the correlations in residual (fig.A.6).

NMF was able to recover images of individual sources which correspond to realistic out-of-focus PSFs (fig. A.5). It also recovered sources with PSFs only partially present in the selected region.

Discussion

We demonstrated that NMF allows to separate images of individual emitters from a movie of blinking QDs. The ability to recover PSFs with different shapes from the data with highly overlapping sources makes the method interesting for imaging of 3D samples where fluorophores can be in different focal depths and therefore produce a variety of different PSFs. Diversity in PSFs for individual sources can also arise with aberrations in the optical system.

NMF allows localisation microscopy with blinking QDs, however, NMF is not limited to QD only. It can separate any source with intermittent intensity as long as the actual fluorophores remain static in space during the acquisition period.

A typical data set consist of 10^3 frames and therefore the acquisition time is typically in order of tens of seconds. During this period the sample must remain still. This is a common limiting factor for localisation microscopy techniques. However, as NMF can separate overlapping sources it can speed up the acquisition compared to the standard LM techniques.[17] Moreover the possibility of using QDs can make the exposure time shorter because QDs are usually brighter than common organic fluorophores used in LM.

NMF does not take into account the temporal structure of the data. Time frames can be permuted without affecting the separated components. However, there is a clear temporal structure in QDs blinking - the periods of ON and OFF times. This information would likely facilitate the factorisation.

Reliable estimation of the number of sources is important for correct recovery of individual components. This is a difficult task for data with high density of overlapping sources. We demonstrated on simulated data that this might be limiting factor for the resolution and applicability of the method.

A Supplementary materials

NMF model of the blinking QDs

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

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

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

$$\mathbb{E}\left[\mathbf{D}\right] = \mathbf{W}\mathbf{H}, \tag{A.2}$$

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

NMF updates

We used iterative NMF updates[11] to compute the factorisation (A.2)

$$\begin{aligned} w_{nk} &= \frac{w_{nk}}{\sum_{t=1}^{T} h_{kt}} \left[(\mathbf{D}./\mathbf{WH})\mathbf{H}^\top\right]_{nk} \\ h_{kt} &= \frac{h_{kt}}{\sum_{n=1}^{N} w_{nk}} \left[\mathbf{W}^\top(\mathbf{D}./\mathbf{WH})\right]_{kt} \end{aligned} \tag{A.3}$$

where the './' operation refers to an element-wise division. The matrix elements w_{nk} and h_{kt} were initialised at random from uniform distribution. The columns of the matrix $\mathbf{W}$ were normalised such that $\sum_n w_{nk} = 1$. One component was added as a homogeneous, static background: $w_{n(K+1)} = \frac{1}{N}$, $h_{(K+1)t} = Nb$, where the background value b was estimated from the dark regions of the data averaged over time $\frac{1}{T}\sum_t d_{nt}$. This component was not updated over the iterations. We run each NMF evaluation with five partial restarts - after each run convergence we restarted the values of h_{kt} while keeping w_{xk}. This procedure helps to reach a better local minimum.

Gamma-Poisson model

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 α_k, β_k and the data are modelled as a 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.

Correlations in residuals

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 a specific 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

$$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}. \tag{A.4}$$

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 can pick up the value of $K = \bar{K}$ for which the residual correlations decrease to a certain level and $K > \bar{K}$ does not give any further improvement.

Simulated data

The parameters used for simulations are shown in table A.1. The individual QDs were generated as in-focus point-spread functions and the intensity of each source was generated from uniform distribution on the interval $[0, I_{max}]$. A uniform constant background was added to each frame and all pixels were corrupted with Poisson noise.

We generated two different collections of data-sets. In the first collection we simulated two sources ($K_{\text{true}} = 2$) separated by a distance $d = 0.1 - 1.2$ pixels. In the second collection we simulated ten sources ($K_{\text{true}} = 10$) randomly distributed within a circular area with diameter $r = \delta$, 1.5δ and 2δ, where $\delta = 0.61\frac{\lambda}{\text{NA}}$ is the classical resolution limit. Each dataset in both collections was generated 10 times with a random geometrical configuration of the individual sources.

Estimation of K

We compared the performance of the the variational approximation of GaP model and the analysis of the residual correlations in K estimation task. We used simulated data comprising 10 sources ($K_{\text{true}} = 10$). The mean image of the data with three different densities of sources is shown in the first line of fig.A.3. We generated data with 10 random geometrical configuration for each density. First 18 principal components (PC) eigenvalues for three different random configurations are plotted in the second row of fig.A.3. For sparse data (left column) we can estimate the number of components from the 'kink' in the spectrum of PC eigenvalues at $K = 10$. However, for dense data (right) this estimation is impossible to make because the PC eigenvalues drip very gradually.

The variational lower bound for models with different K for three different geometrical configuration of the sources is plotted in the third line of fig.(A.3). The lower bound peaks at the correct $K = K_{\text{true}}$ for sparse data set (left), but for the denser data (right) it tends to reach the maximum for $K < K_{\text{true}}$. The variational lower bound systematically underestimates the value of K.

The maximum correlation coefficient (A.4) in residuals $\max_{i,j}(r_{ij})$ is shown in the last line of fig.A.3. The model was estimated using the NMF updates(A.3). Values for three different configurations are shown. We can pick the K were the 'kink' occurs.

A histogram of estimated K for both methods is shown in fig.A.4b. The analyses of the maximum residual correlation provides a better estimate for K.

We used the simulated data of two QDs ($K_{\text{true}} = 2$) separated by $d = 0.1 - 1.2$ pixels to evaluated the model for $K = 1$, 2 and 3 and analysed the residual correlations (A.4). The maximum value $\max_{i,j}(r_{ij})$ in the correlation coefficient matrix (A.4) as a function of d is shown in fig.A.1 for the three different values of K. 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 $K = 1$ (blue line) compared to $K = 2$ (green line). The model with $K = 3$ (red line) does not lead to any further improvement. For this noise level we can therefore estimate the correct $K = 2$ for emitters separated by at least $d > 50$ nm.

Localisation error

To asses the localisation performance of NMF we used dataset of two QDs ($K_{\text{true}} = 2$) separated by d and assumed that the number of sources is known $K = K_{\text{true}}$. Separated components were localised as a maximum likelihood fit of an isotropic 2D gaussian with a centre $\overline{\boldsymbol{x}}_i$. We computed the mean localisation error per separation

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

where $e_i = \left|\overline{\boldsymbol{x}}_i - \boldsymbol{x}_i^{\text{true}}\right|$ corresponds to the localisation error of the ith source located at $\boldsymbol{x}_i^{true}$ and $d = \left|\boldsymbol{x}_1^{\text{true}} - \boldsymbol{x}_2^{\text{true}}\right|$ is the true separation of the emitters. The estimated ε for different values of d is shown as a magenta line in fig.A.1. For $d > 30$ nm the mean localisation error is less than 20% of the separation d ($\varepsilon = 0.2$) and is further decreasing for higher values of d.

Comparison with with the maximum residual correlation curves (blue, green, red in fig.A.1) shows that the determining factor for resolution of the method is the ability to estimate K correctly (here for $d > 50$ nm).

Real samples

Solution of QDs (Qdot 565, invitrogen) was deposited and dried on the coverslip and imaged with a standard wide-field microscope (100x, 1.4NA) on a CCD camera (64nm pixel-size). We acquired 10^3 frames each with 100 ms exposure time. The total acquisition time was about 2 mins.

The camera pixels values were converted into a photon counts with a gain estimated from a calibration measurement.

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 ($\delta = 0.61\frac{\lambda_{em}}{NA}$).
noise	Poisson	Each pixel was corrupted with Poisson noise.

Table A.1: Parameters of the simulation

Figures

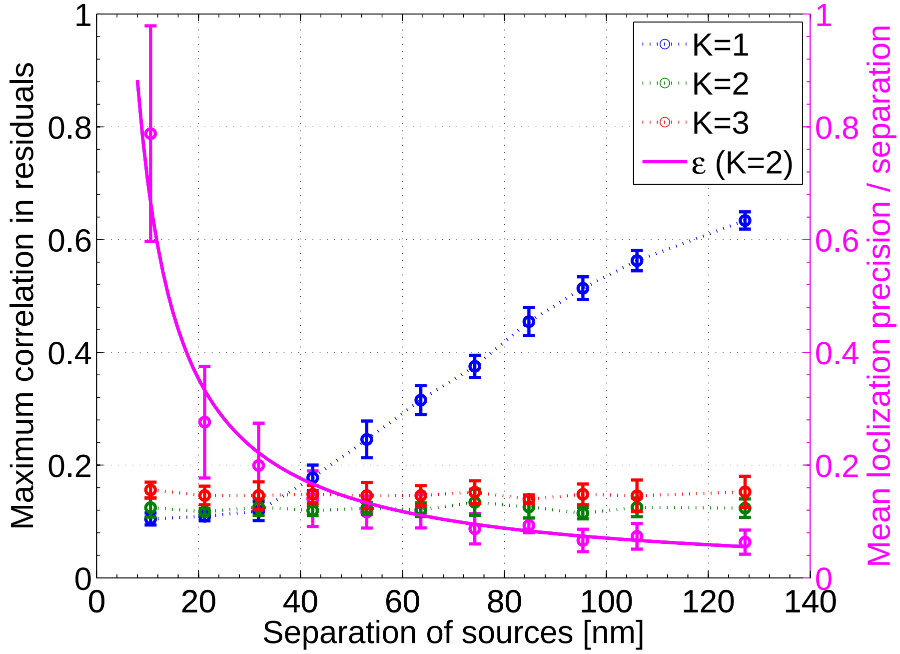
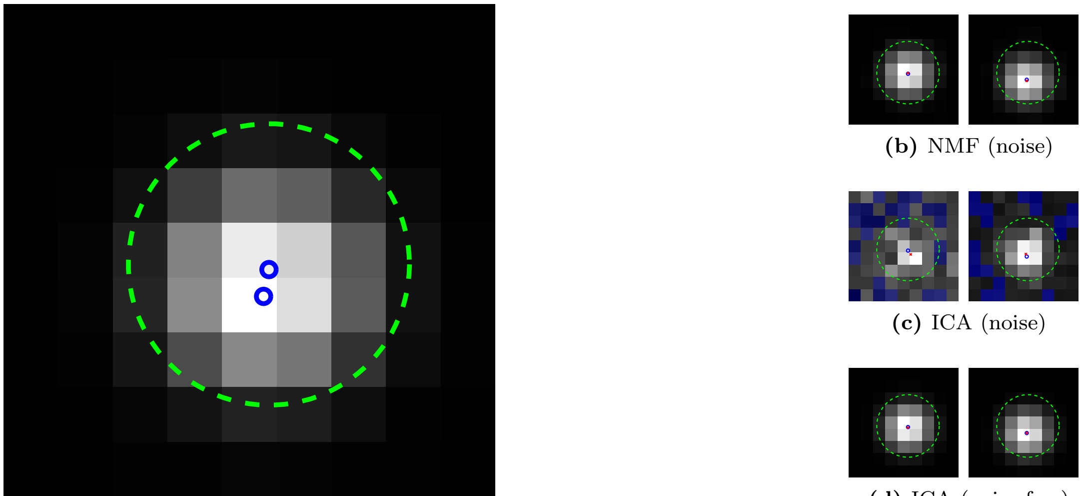


Figure A.1: The maximum value of correlations in residuals for $K = 1$ (blue), 2 (green) and 3 (red). Mean localisation error per separation ε for $K = 2$ (magenta) with fitted curve $\propto 1/d$. Simulated data of two sources ($K_{\text{true}} = 2$). The difference in residual correlations for $K =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



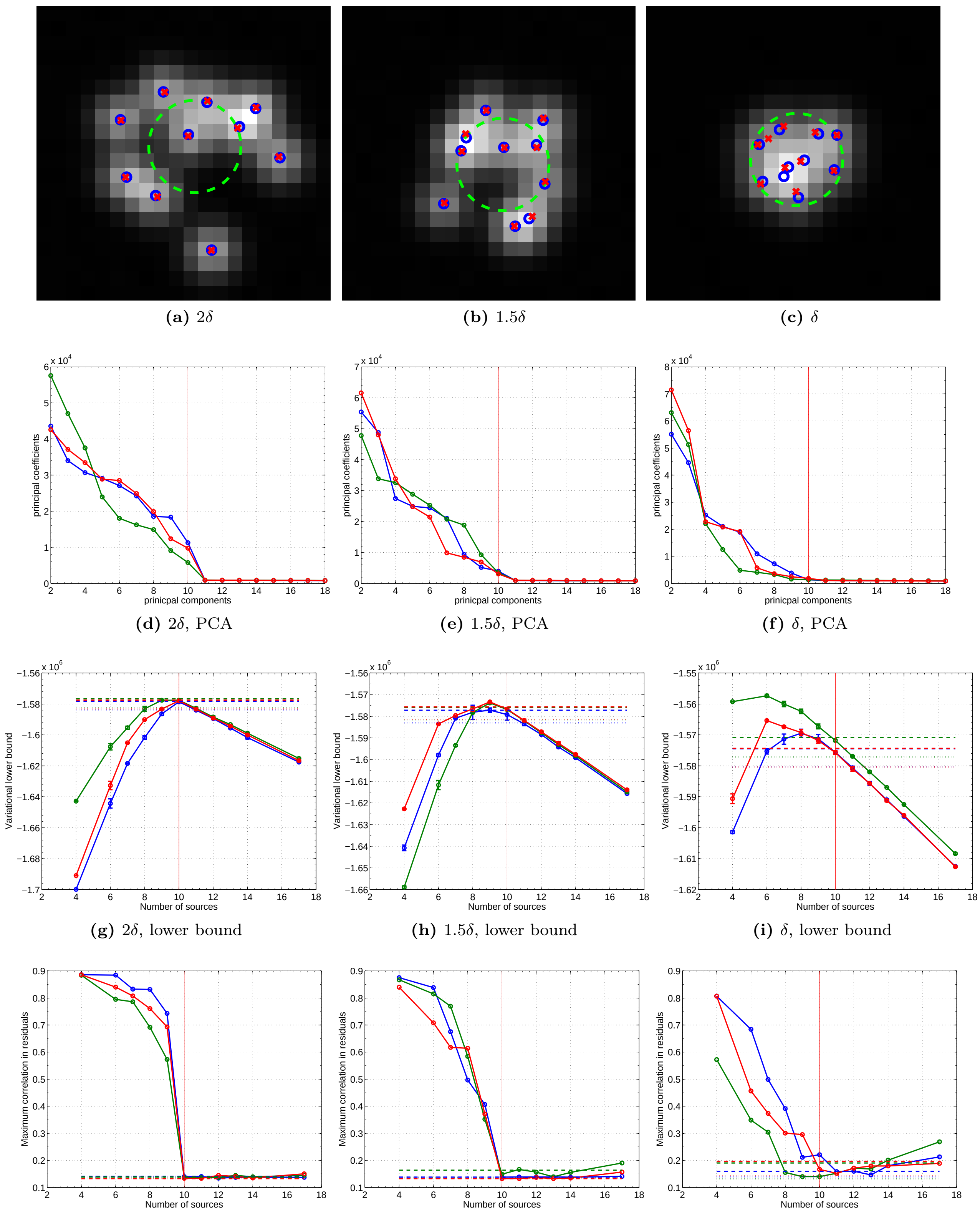
(a) wide field

(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.



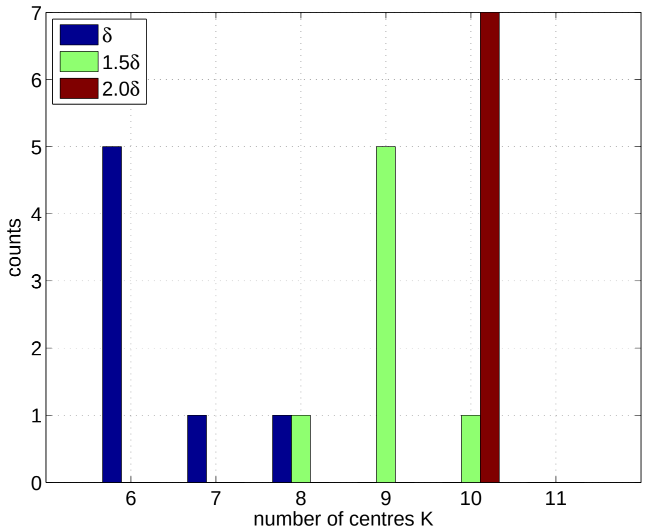
(a) 2δ **(b)** 1.5δ **(c)** δ

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

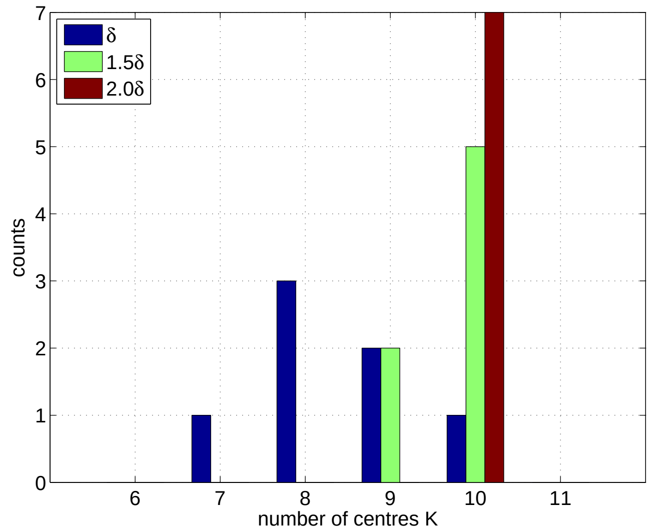
(g) 2δ, lower bound **(h)** 1.5δ, lower bound **(i)** δ, lower bound

(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).

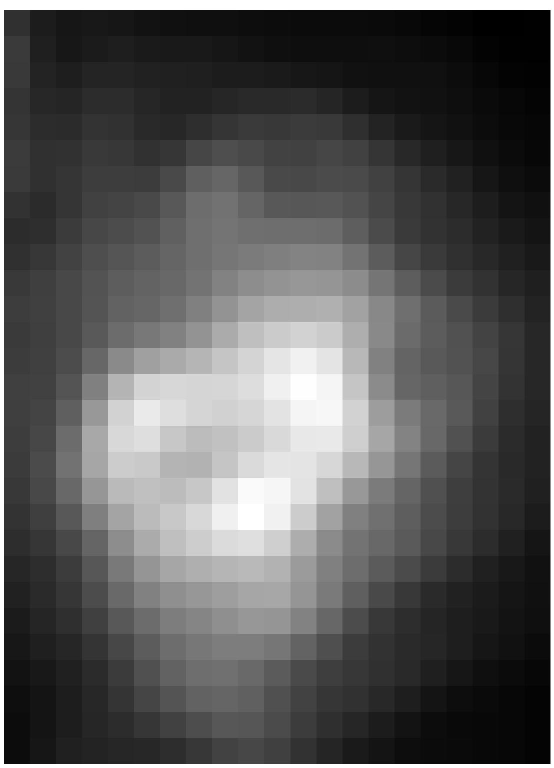


(a) Variational lower bound

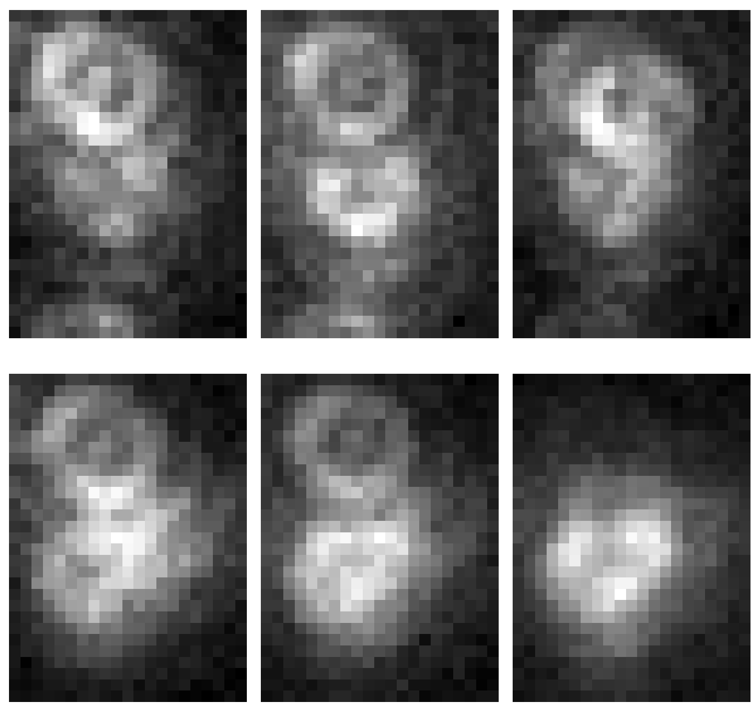


(b) Maximum correlations in residuals

Figure A.4: Histograms of the K estimations for data with three different sources density. ($K_{\text{true}} = 10$)

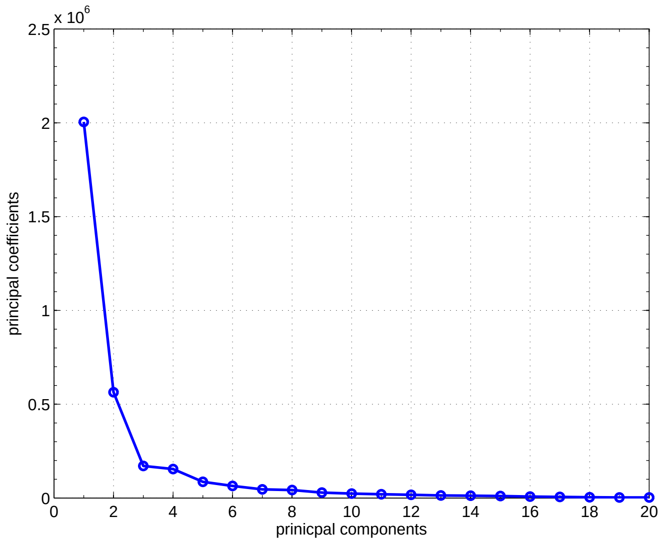


(a) Mean intensity image

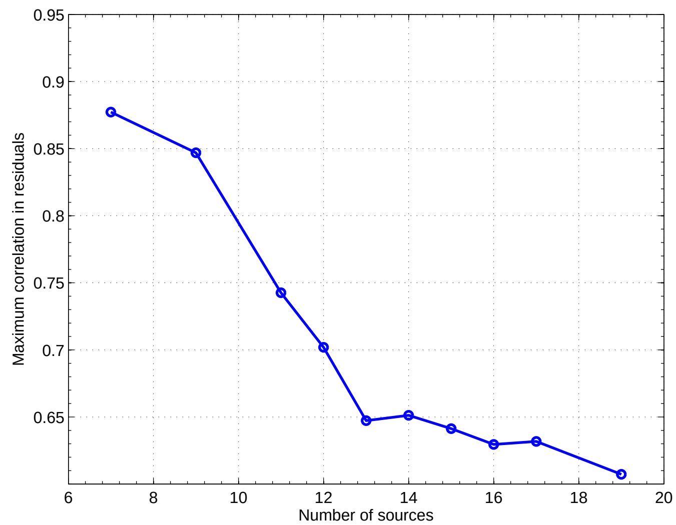


(c) Movie of the blinking sources (six frames out of 10^3)

Figure A.5: Real QD data.

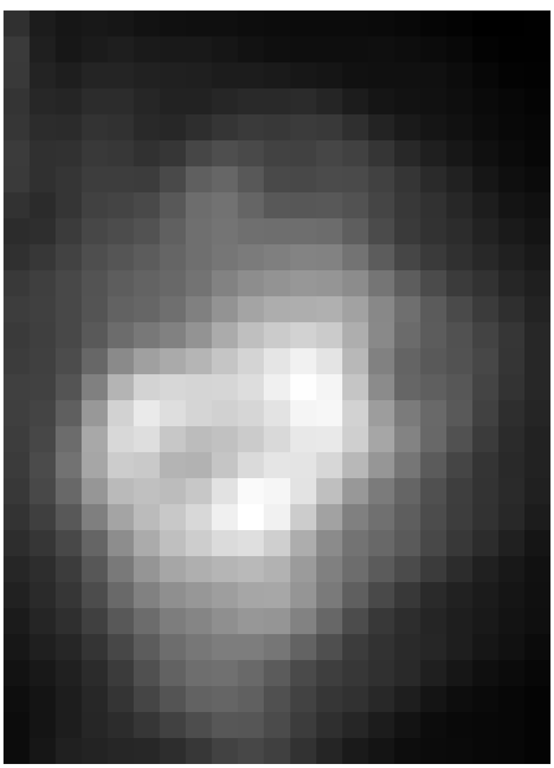


(a) PCA

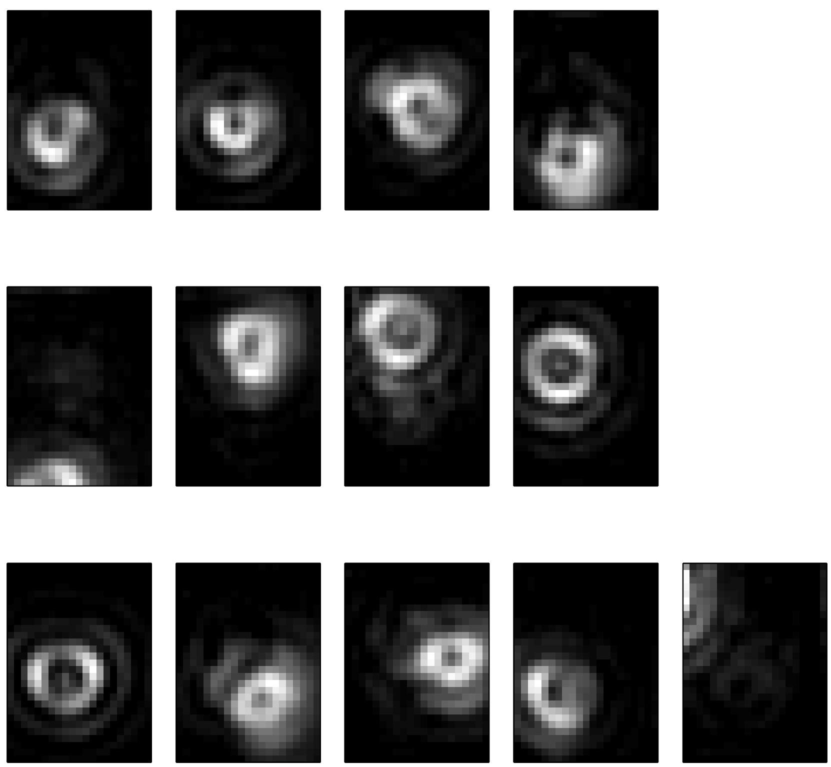


(b) Correlations

Figure A.6: Estimation of K



(a) Mean intensity image



(b) Separated sources

Figure A.7: 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.