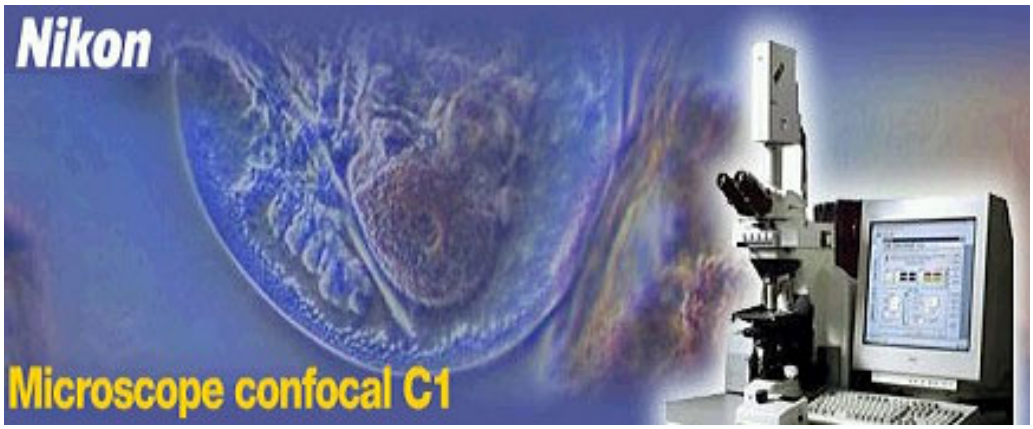
A review on :
3-D reconstruction from confocal scanning microscope images: Deterministic and maximum likelihood reconstruction

Written by: Jose-Angel Conchello and Eric W. Hansen
review by: Sarit Shwartz



1 Introduction

The purpose of this work was to improve three dimensional images of biological specimens stained with fluorescent probes taken with confocal scanning microscope. The article describes two approaches–an iterative deterministic deconvolution and maximum likelihood estimator.

1.1 Acquiring the images

Three dimensional images are obtained from optical microscope by collecting a series of two dimensional images focused at different planes through the specimen. Each optical section shows the plain in focus plus contributions from out of focus plains, which obscures the desired image.

Better images can be obtained by using a confocal scanning microscope. In confocal scanning microscope only a small portion of the specimen is illuminated at a given time and light from other points in the specimen is rejected by a small aperture in front of the detector, but the contributions from out of focus plains are reduced but not completely eliminated.

1.2 Operators and notations

Symbol	meaning
(x_D, y_D, z_D)	Detector (image) position
(x_s, y_s, z_s)	Scan position
(x, y, z)	Coordinates for a point in the specimen
λ_f	Fluorescent light wave length
λ	Excitation light wave length
d	Distance of focus plane from the lens
$sin(\alpha)$	The numerical aperture of the lens
$o(\cdot)$	Object intensity
$i(\cdot)$	Recorded image brightness
$h_1()$	The point spread function of the lens at wave length λ
$h_2()$	The point spread function of the lens at wave length λ_f
$h()$	The point spread function of the microscope
$i_f()$	The fluorescent intensity distribution in the object space
$I()$	The fluorescent image of $i_f()$ through the lens
$OTP()$	The optical transfer function of the microscope
$T_1()$	Optical transfer function of the lens at wave length λ
$T_2()$	Optical transfer function of the lens at wave length λ_f
(v_x, v_y, u)	Optical coordinates
(s, t, w)	Normalized frequencies
$T_{n1}()$	Optical transfer function of the lens at wave length λ in optical coordinates
$T_{n2}()$	Optical transfer function of the lens at wave length λ_f in optical coordinates
$h_{n1}()$	The point spread function of the lens at wave length λ in optical coordinates
$h_{n2}()$	The point spread function of the lens at wave length λ_f in optical coordinates
$\hat{i}^{(k)}(x, y, z)$	Image of the reconstructed image at iteration k
$\hat{o}^{(k)}(x, y, z)$	The reconstructed image at iteration k
ϵ	Arbitrary convergence criterion
$\gamma(x, y, z)$	The step size of iteration k for point (x,y,z) in the object
A	Half of the maximum allowed value of the object
$p()$	Probability density
$\mathcal{Y}$	Incomplete data space
$\mathcal{Z}$	Complete data space
$f : \mathcal{Z} \rightarrow \mathcal{Y}$	Mapping between incomplete data space and complete data space
θ	Parameter vector to be estimated
Θ	Parameters space
$hat\theta^{(k)}$	Parameter vector estimation in iteration k

$\mathcal{L}$	Log likelihood
$\mathcal{L}_{id}$	Incomplete log likelihood
$\mathcal{L}_{cd}$	Complete log likelihood
$Q(o(1), o(2), ...\|\hat{o(1)}^{(k)}, \hat{o(2)}^{(k)}, ...)$	Conditional expectation of the complete log likelihood
η	Arbitrary learning rate constant
$\mathcal{X}_{lt}$	Arbitrary parameters characterizing correlation of neighboring source elements in the object around element l
$\bar{o}(l)$	Mean strength value of voxel l in the object
σ_l	Standard deviation of source element l in the object
$\otimes$	Convolution operator

2 Optical transfer function and point spread function of a confocal scanning fluorescent microscope

2.1 Optical system

The imaging property of a confocal scanning microscope that uses induced fluorescence is different from that of one that uses transmitted or reflected light. This is because the fluorescence is incoherent. Thus the bandwidth of its optical transfer function is twice as large as that of a confocal coherent microscope.Moreover, the missing cone in the optical transfer function for this microscope is smaller, and if the excitation and the fluorescence wavelength is close, the missing cone disappear.

There are two possible configurations for the optical system. Simplified diagrams of both are shown in Fig. 1. The optical transfer function calculations are identical for both, so mathematical analysis will be given only for the one lens configuration.

Excitation light with wavelength λ that comes from a point source is reflected by half mirror and focused by the lens at a distance d from the lens. The point in the specimen space is denoted by coordinates x, y and z. The direction of the z axis is on the optical axis. The distribution of a fluorescent object may be expressed as $o(x - x_s, y - y_s, z - z_s)$.

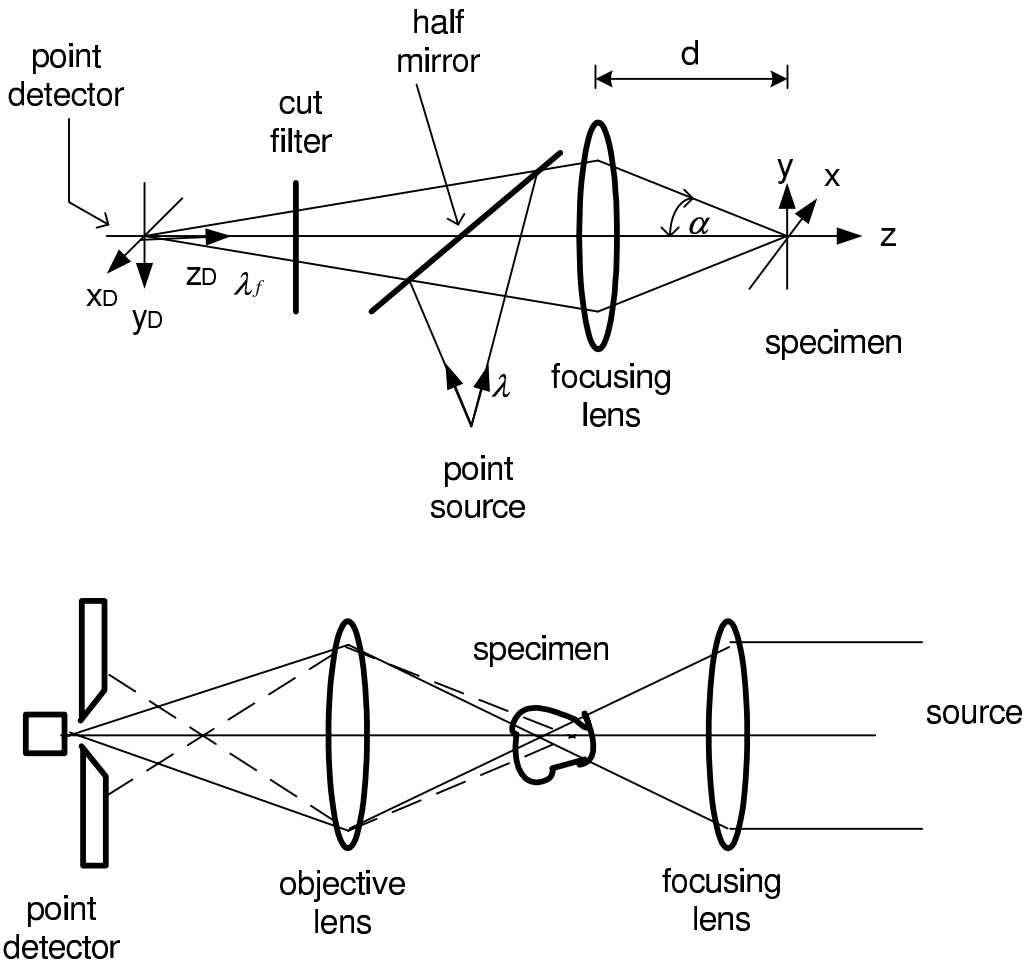


Figure 1: Confocal optical path configuration - Up:one lens configuration. Down:two lens configuration

Coordinates x_s, y_s and z_s represents the scan position. The image of the induced fluorescence is formed through the focusing lens, the half mirror, and a cut filter. Coordinates x_D, y_D and z_D denote the position of the image. The cut filter is inserted into the optical path to block the excitation light. for simplicity assume that only fluorescence with wavelength λ_f is allowed to pass through the filter. and that the a point detector is positioned at the origin of the coordinate system of the image.

2.2 Calculation of Optical transfer function

Following Shigeharu and Chusuke [1], let a complex amplitude of the point source image reflected through the focusing lens at a wavelength of λ be $h_1(x, y, z)$, and let the same amplitude at a fluorescence wavelength of λ_f be $h_2(x, y, z)$. Assume that the induced fluorescence at some object point is proportional to the distribution $o(x - x_s, y - y_s, z - z_s)$ and to the incidence intensity of the excitation light at the same point. then the fluorescent intensity distribution $i_f(x, y, z)$, in the object space is:

$$i_f(x, y, z) = |h1(x, y, z)|^2 o(x - x_s, y - y_s, z - z_s). \tag{1}$$

Therefore the fluorescent image of this distribution through the focusing lens is:

$$I(x_D, y_D, z_D, x_s, y_s, z_s) = \iiint_{-\infty}^{\infty} \left| h_2\left(\frac{x_D}{M} - x, \frac{y_D}{M} - y, \frac{z_D}{M} - z\right)\right|^2 |h_1(x, y, z)|^2 \times o(x - x_s, y - y_s, z - z_s)\, dx\, dy\, dz \tag{2}$$

where M is the magnification of the system. the point detector is placed at the origin of the image coordinate system, and h_2 is an even function therefore the detected fluorescence is expressed as

$$I(0, 0, 0, x_s, y_s, z_s) = \iiint_{-\infty}^{\infty} |h_2(x, y, z)|^2 |h_1(x, y, z)|^2 o(x - x_s, y - y_s, z - z_s)\, dx\, dy\, dz \tag{3}$$

from Eq.3, the point spread function $psf(x, y, z)$ for the microscope is expressed as

$$psf(x, y, z) = h(x, y, z) = |h_2(x, y, z) h_1(x, y, z)|^2 \tag{4}$$

The 3-D optical transfer function can be calculated with the 3-D fourier transform of the point spread function from Eq.4.

$$OTF(f_x, f_y, f_z) = \iiint_{-\infty}^{\infty} |h_2(x, y, z)|^2 |h_1(x, y, z)|^2 exp(-2\pi i(x f_x + y_f y + z f_z))\, dx\, dy\, dz = \int_{-\infty}^{\infty} [T_1(f_x, f_y, z) \otimes T_2(f_x, f_y, z)] exp(-2\pi i z f_z)\, dz \tag{5}$$

where f_x, f_y and f_z are spatial frequencies in the x, y and z directions. and $\otimes$ denotes a convolution. The functions T_1 and T_2 are 2-D optical transfer function with defocus and are expressed as

$$\begin{aligned} T1(f_x, f_y, z) &= \iint_{-\infty}^{\infty} |h_1(x, y, z)|^2 exp(-2\pi i(xf_x + y_f y))\, dx\, dy, \\ T2(f_x, f_y, z) &= \iint_{-\infty}^{\infty} |h_2(x, y, z)|^2 exp(-2\pi i(xf_x + y_f y))\, dx\, dy. \end{aligned} \tag{6}$$

By introducing the optical coordinates

$$\begin{aligned} v_x = \frac{2\pi}{\lambda} x \sin(\alpha) \approx \frac{2\pi}{\lambda}\frac{a}{d}x,\ \ v_y = \frac{2\pi}{\lambda} y \sin(\alpha) \approx \frac{2\pi}{\lambda}\frac{a}{d}y, \\ u = \frac{2\pi}{\lambda} z \sin^2(\alpha) \approx \frac{2\pi}{\lambda}\left(\frac{a}{d}\right)^2 z. \end{aligned} \tag{7}$$

Where λ is the excitation light wavelength, $\sin(\alpha)$ is the numerical aperture, a is the radius of the exit pupil of the focusing lens, assuming the refractive index $n = 1$, and introducing the normalized frequencies

$$s = \frac{\lambda}{\sin(\alpha)} f_x,\ \ t = \frac{\lambda}{\sin(\alpha)} f_y,\ \ w = \frac{\lambda}{\sin^2(\alpha)} f_z. \tag{8}$$

The 2-D optical transfer function represented by T_1 can be replaced with

$$T_{n1}(s, t, u) = \iint_{-\infty}^{\infty} |h_{n1}(v_x, v_y, u)|^2 exp(-i(u_x s + v_y t))\, dv_x\, dv_y \tag{9}$$

Where h_{n1} is the complex amplitude of the point source represented in the optical coordinates. Hopkins [2] calculated $T_{n1}(s, 0, u)$ analytically. However, his approach involved a series of Bessel functions converges slowly and it is time consuming. Later Stokseth [3] presented an empirically derived analytic approximation of the optical transfer function to make the calculation easy. According to this approximation, $T_{n1}(s, 0, u)$ can take the following form

$$T_{n1}(s, 0, u) = \begin{cases} g_1(s)\left\{\frac{J_1[ug_2(s)]}{ug_2(s)}\right\} & 0 < s < 2 \\ 0 & 2 \le s \end{cases} \tag{10}$$

where

$$g_1(s) = 2(1 - 0.69s + 0.0076s^2 + 0.043s^3),$$
$$g_2(s) = s - 0.5s^2 \tag{11}$$

and J_1 is a Bessel function of the first kind of order 1.

In similar way the 2-D optical transfer function at the fluorescence wavelength T_{n2} is represented in the optical coordinates with λ_f as the fluorescence wavelength and h_{n2} the complex amplitude. expressing $\lambda_f = \beta\lambda$ leads to

$$h_{n2}(v_w, v_y, u) = h_{n1}\left(\frac{v_x}{\beta}, \frac{v_y}{\beta}, \frac{u}{\beta}\right) \tag{12}$$

Hence from the similarity theorem of the Fourier transform we get the relationship between T_{n1} and T_{n2}.

$$T_{n2}(s, t, u) = T_{n1}\left(\beta s, \beta t, \frac{u}{\beta}\right) \tag{13}$$

By applying Eq. 9, 12 and 13 to Eq. 5 we can write it as

$$OTP(s, t, w) = \mathcal{F}[h_{n1}(v_x, v_y, u)|^2] \otimes \mathcal{F}[h_{n2}(v_x, v_y, u)|^2] \tag{14}$$

in which the symbol $\mathcal{F}$ denotes a 3-D fourier transformation. Eq. 14 shows that the 3-D optical transfer function for the confocal fluorescent microscope is represented by the convolution of incoherent wide field microscope 3-D optical transfer functions at the wavelengths of the excitation and fluorescence light. the incoherent wide field microscope 3-D optical transfer functions itself is a convolution of the coherent 3-D optical transfer functions.

2.3 The optical transfer function bandwidth

From Eq. 14 we can see qualitatively how the bandwidth change for the 3-D optical transfer function. The bandwidth of wide field microscope optical transfer function becomes narrower as the wavelength of the fluorescence light becomes longer. When the wavelength of the fluorescence is equal to the excitation wavelength, the confocal optical transfer function

bandwidth is twice as wide as the wide field microscope bandwidth. When the fluorescence wavelength becomes longer, the bandwidth approach the bandwidth of the wide field microscope. in that case a missing cone region appears.

3 Additive Deconvolution

The reconstruction is carried out by constrained iterative deconvolution, a purely deterministic algorithm which is a modification to the Jansson-van-Cittert method of successive convolution [4]. The correction step size for each voxel at each iteration is determined by a quadratic form. This quadratic form has maximum at half the maximum value allowed for the object $\gamma(x,y,z) = 1$ if $o_old(x,y,z) = A$, and minimum at the maximum value allowed for the object and at zero $\gamma(x,y,z) = 0$ if $o_old(x,y,z) = 2A$ or 0. The quadratic form of the step size is disabling drastic corrections at the saturation and cutoff regions of the image, were such corrections could spoil the estimation.

3.1 Algorithm

1. $\hat{i}^{(k)}(x,y,z) = \hat{o}^{(k)}(x,y,z) \otimes h(x,y,z)$
2. $\hat{o}^{(k+1)}(x,y,z) = \hat{o}^{(k)}(x,y,z) + \gamma(x,y,z)[i(x) - \hat{i}^{(k)}(x)]$

with $\gamma(x,y,z) = 1 - \left[\frac{\hat{o}^{(k)}(x,y,z)-A}{A}\right]^2$

3. Apply constraints.
4. k=k+1.
5. Repeat 1–4 until $max(|i(x,y,z) - \hat{i}^{(k)}(x,y,z)|) < \epsilon$

where $\hat{o}^{(k)}(x,y,z)$ is the reconstructed object of iteration k, $\hat{i}^{(k)}(x,y,z)$ is the image of the reconstructed object at iteration k and $h(x,y,z)$ is the point spread function. A is half of the maximum allowed value of the specimen or object, ϵ is an arbitrary convergence criterion and $\gamma(x,y,z)$ is the step size. The 3-D convolution at step 1 is implemented by FFT. At step 3, prior knowledge about the specimen can be incorporated to the algorithm. In this

article simulation the only constraint that was used was boundedness i.e.

$$\hat{o}^{(k)}(x,y,z) \leftarrow \begin{cases} 0 & if \hat{o}^{(k)}(x,y,z) < 0 \\ 2A & if \hat{o}^{(k)}(x,y,z) > 2A \\ \hat{o}^{(k)}(x,y,z) & else \end{cases}$$

The recorded image $i(x,y,z)$ is used as an initial guess.

3.2 1D and 2D Simulation Results

The simulations tested several several convergence criterions and several objects smooth,sharp bright and sharp dim. The first simulation checked the convergence as function of the stoping condition. In Fig. 2 the system parameters (PSF, object) are showed. The results of the simulation are showed in Fig. 3, 4 and 5. We can see for these figures that the algorithm is very sensitive to the stoping condition ϵ. The range of good values is very small. If we will choose a value too large, the improvement in the image will be minor were as if we will choose a value too small the algorithm won't converge at all. In addition the smaller ϵ is, or the more iteration done, more artifacts will be introduced to the estimation. The artifacts will be seen as bright and dark spots or grid on the background of the image. This can be clearly seen in Fig. 5 as local minimums and maximums.
The second simulation checked different 2D objects that can be seen in Fig. 6, 8 and 10, The algorithm results for the different objects are in Fig. 7, 9 and 11. for all the objects and for $\epsilon = 10,5$ the algorithm didn't converge but was stop after 100 iterations. For all objects there is an improvement in the reconstructed image but also some artifacts were introduced to the background. In the sharp and dim object, although the object is less blurred it is still not easily separable from its background.

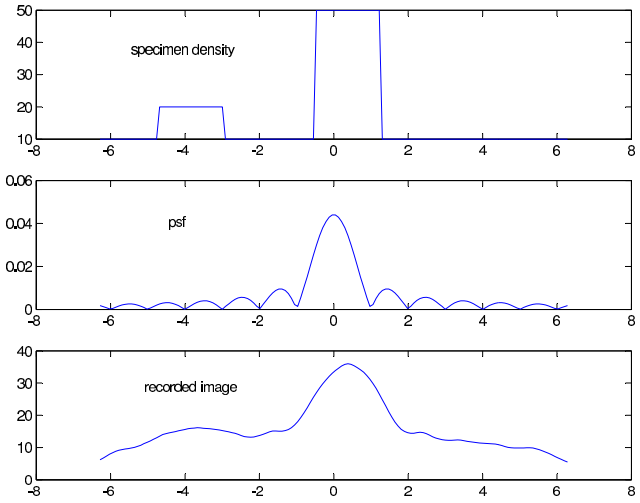


Figure 2: 1D system parameters for sharp object

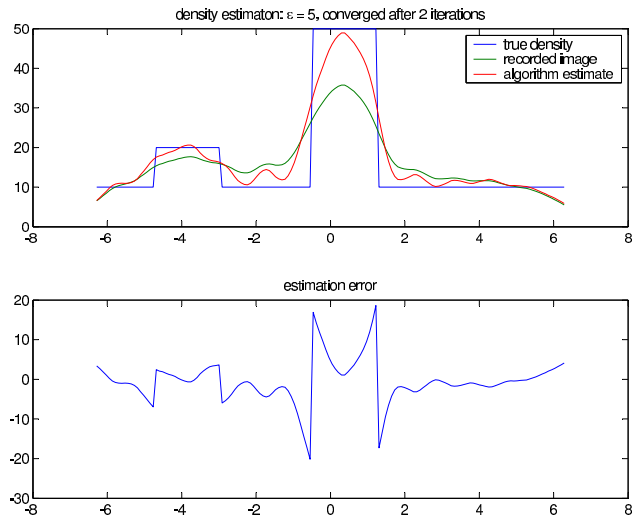


Figure 3: 1D additive deconvolution results $\epsilon = 5$: convergence after two iterations.

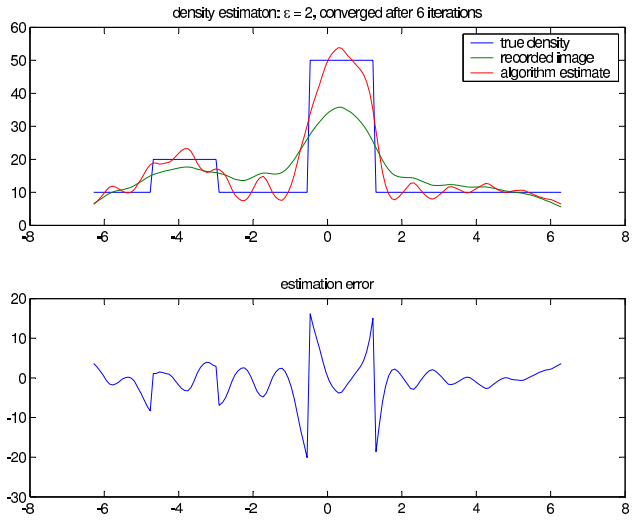


Figure 4: 1D additive deconvolution results $\epsilon = 2$: convergence after six iterations

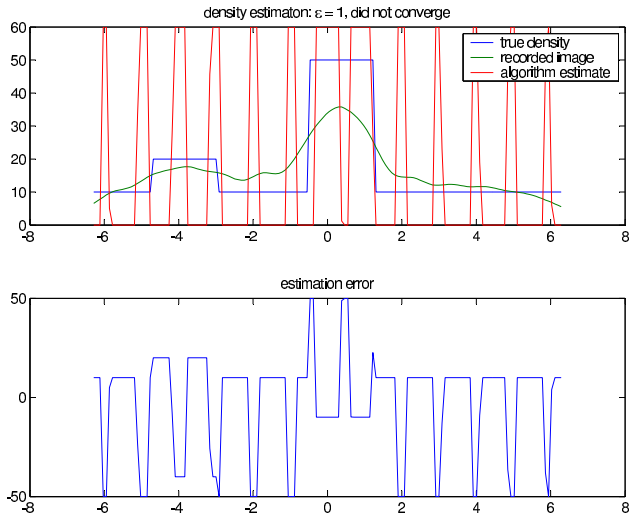


Figure 5: 1D additive deconvolution results $\epsilon = 1$: did not converge and was stopped after 1000 iterations

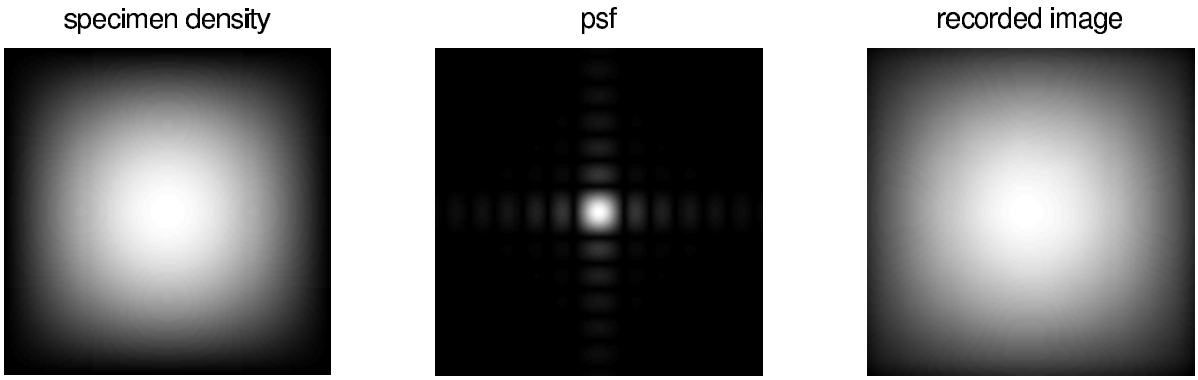


Figure 6: 2D system parameters for smooth object

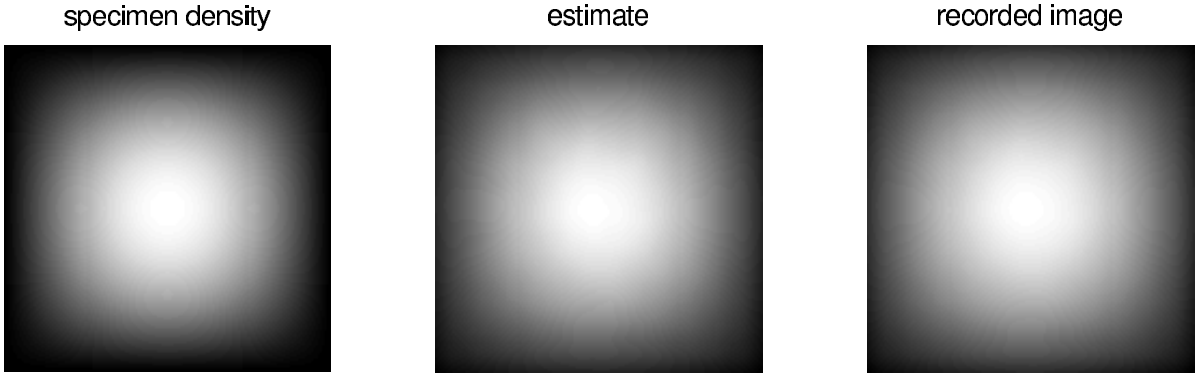


Figure 7: 2D additive deconvolution results for smooth object: did not converge for $\epsilon = 10, 5$ and was stopped after 100 iterations

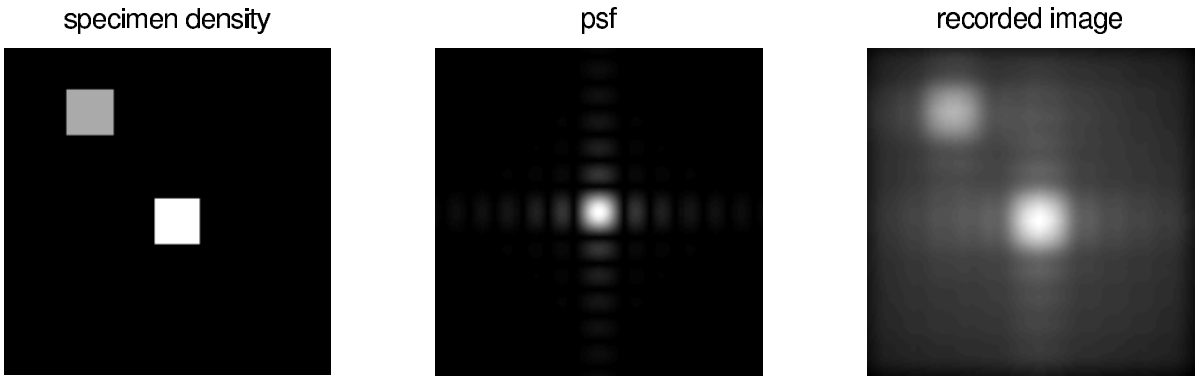


Figure 8: 2D system parameters for sharp and bright object

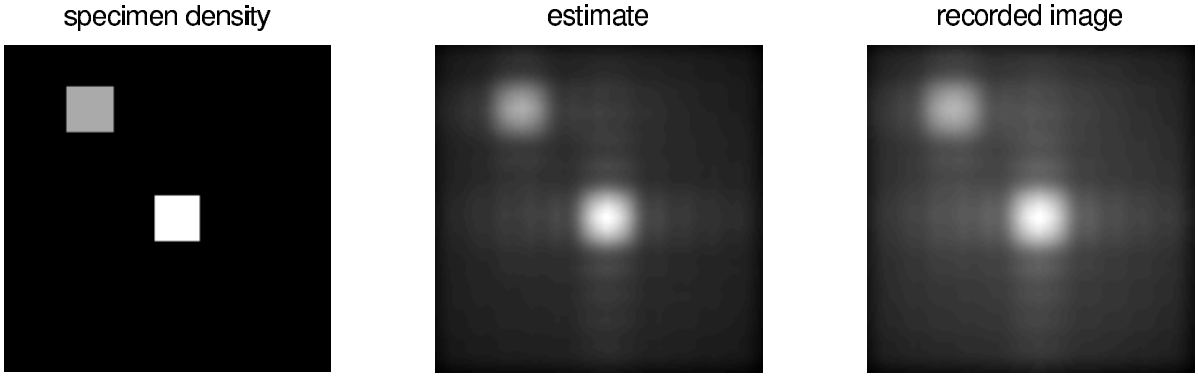


Figure 9: 2D additive deconvolution results for sharp and bright object: did not converge for $\epsilon = 10,5$ and was stopped after 100 iterations

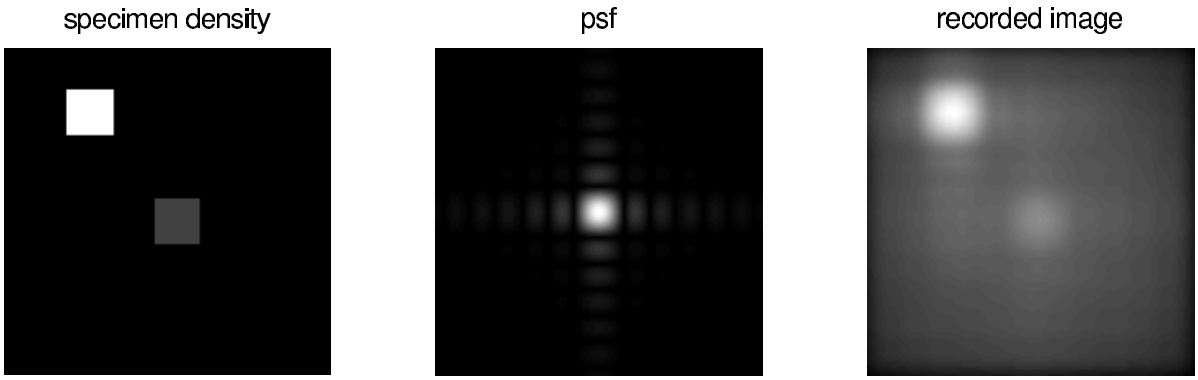


Figure 10: 2D system parameters for sharp and dim object

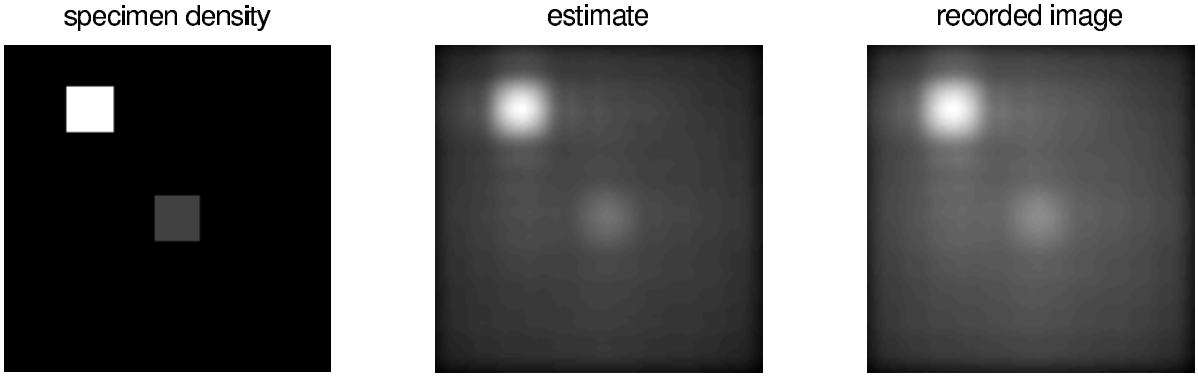


Figure 11: 2D additive deconvolution results for sharp and dim object: did not converge for $\epsilon = 10, 5$ and was stopped after 100 iterations

3.3 Article Simulation Results

The simulation was done on three different objects. a cross and a circle 200 nm apart and with equal brightness, a cross and a circle 50 nm apart and with equal brightness and a cross and a circle 200 nm apart with the cross strength of 1/10 from the circle strength. The first simulation was applied to the case where the cross and the circle have equal strength, are 200 nm apart and without noise. The result of this simulation are in Fig. 12(left) from this figure we can see that the out of focus contribution are greatly reduced, after 10 iterations the objects are bright in the slice where they are in focus, and dim only in the slices or planes immediately above and below. Some ringing artifacts are seen in the planes immediately above and below where the object is in focus. in the same figure (right) the results for the case when the cross and circle are 50 nm apart are shown. we can see that the out of focus contribution from the two objects are reduced in only 10 iteration to one slice above and below the location of the circle. The contribution to these two slices does not totally go away even after 40 iterations. The cross, on the other hand, can only be seen at the second and third planes.

Figure 13(left) shows the case where the cross have 1/10 the intensity of the circle. in the recorded image, the contributions of the out of focus circle to the second panel outweigh the in focus contribution from the cross. As the iteration progresses, the relative strength of the cross increases. One plane above the cross, the contributions from the circle are quickly outweighed by those of the cross. Although the algorithm is not able to resolve clearly the two objects, it is capable of bringing out dim detail that could barely be seen in the recorded image.

The algorithm was tested then in the presence of signal dependent poisson noise. the case where the two objects are with equal intensity and 200 nm and 50 nm apart. for both cases the SNR at the brightest region was set to 3.16($\sqrt{10}$ corresponding to the mean intensity of 10 pixels per voxel). The results for the first case are on Fig. 13(right). The algorithm was able to reduce the out of focus contributions to only the optical slices immediately above and below the two objects. However, at the same time, the noise is amplified as in evidenced by the salt and pepper noise in the reconstructed image.

The results for the second case are shown in Fig. 14 In only 10 iterations, the out of focus contributions of the circle are reduces to the slices above and below the plane where it is in focus. Unlike the noiseless case, the out of focus contributions from the cross to the bottom panel still shows even after 40 iterations. As in the previous case, the high frequency noise grows as iteration progress.

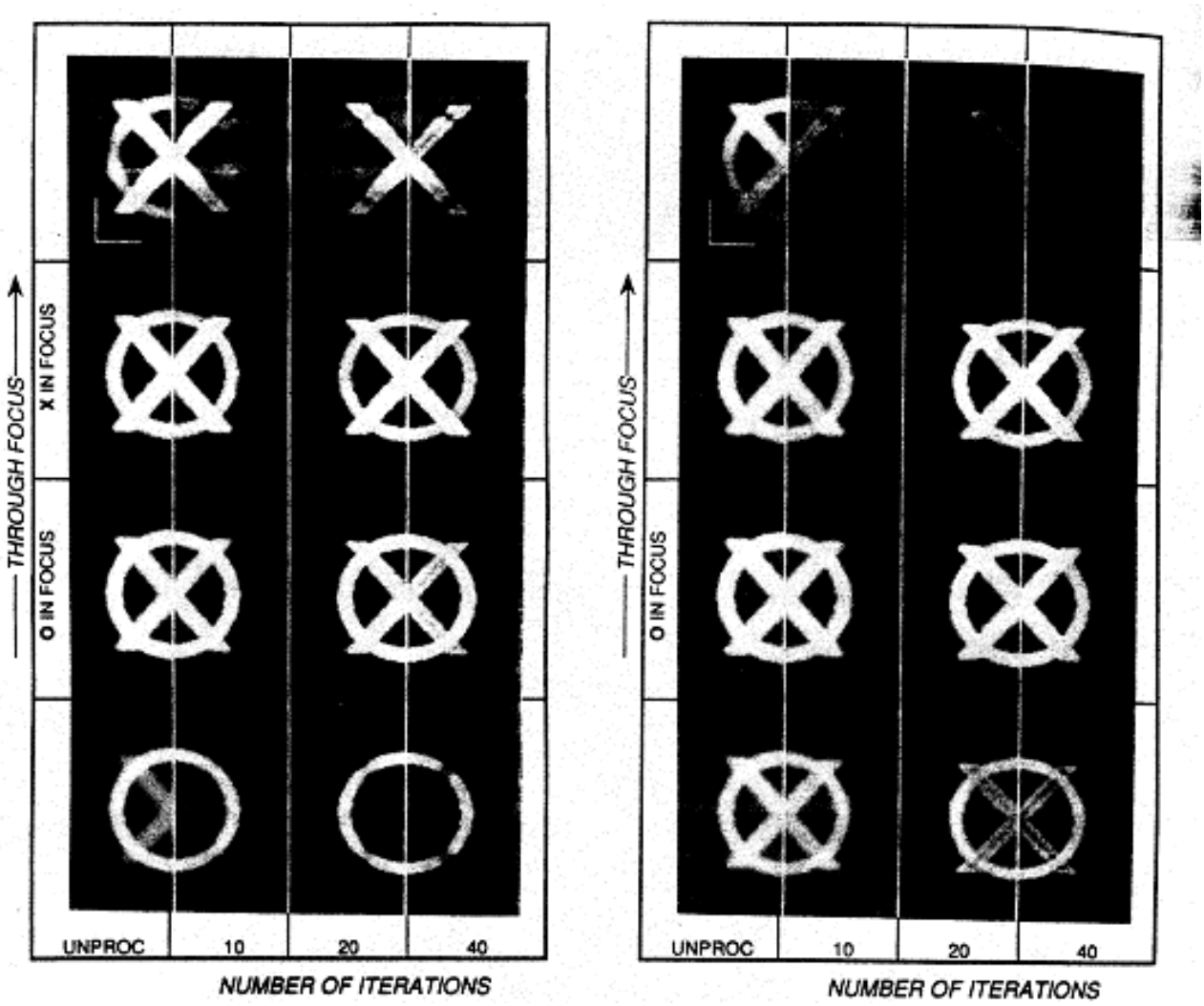


Figure 12: Additive algorithm results - without noise. Left: simulated cross and circle of equal intensity 200 nm apart, the optical slices where the cross and circle lays marked X and O respectively. the columns from left to right are the recorded image and the reconstruction after 10,20,40 iterations. Right: simulated cross and circle of equal intensity 50 nm apart, the optical slice where the circle lays marked O. the columns from left to right are the recorded image and the reconstruction after 10,20,40 iterations

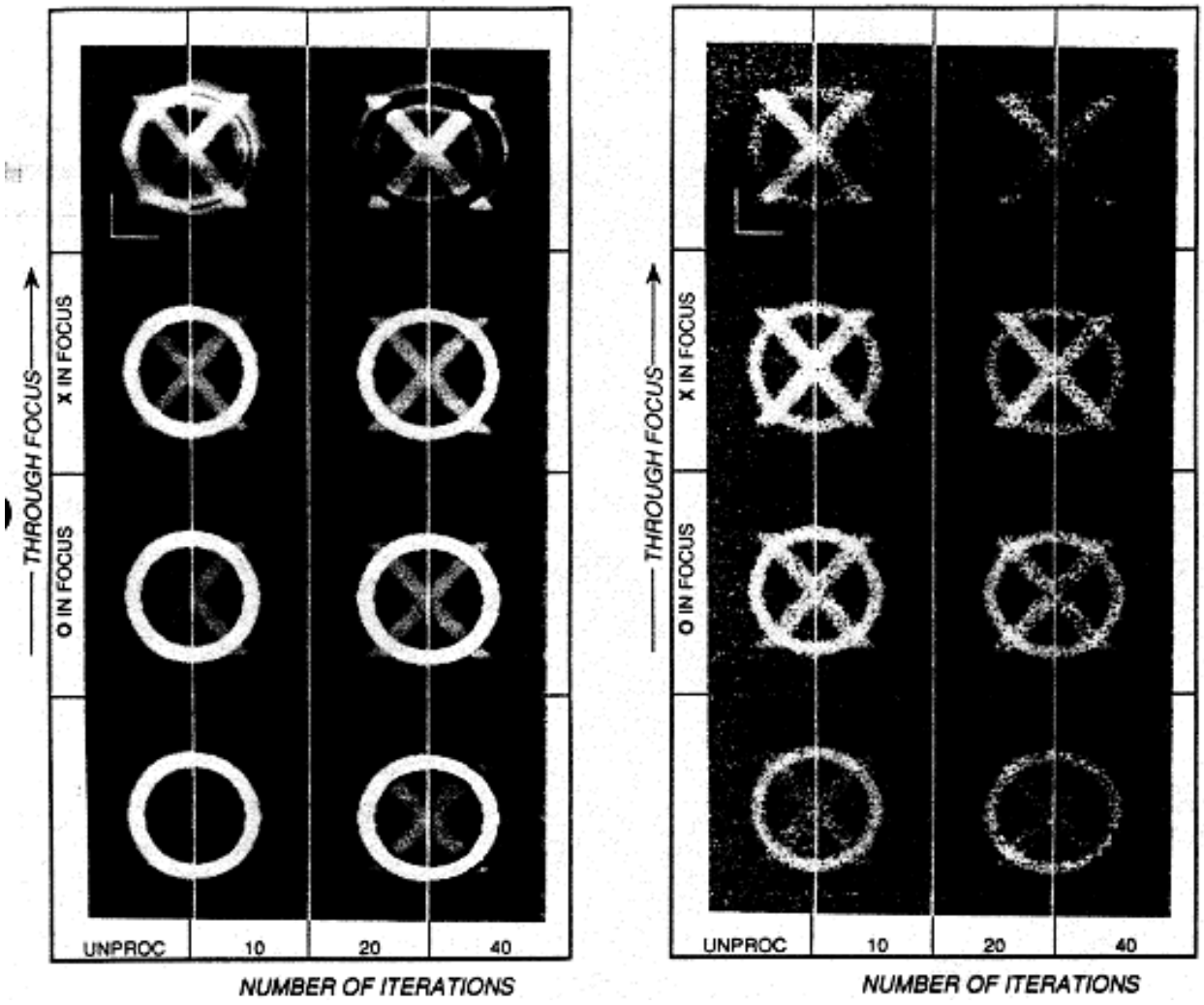


Figure 13: Additive algorithm results - with and without noise. Left: without noise - simulated cross and circle of 200 nm apart the cross intensity is 1/10 the circle intensity, the optical slices where the cross and circle lays marked X and O respectively. the columns from left to right are the recorded image and the reconstruction after 10,20,40 iterations. Right: with poisson noise max. SNR=3.16 simulated cross and circle of equal intensity 200 nm apart, the optical slice where the circle lays marked O. the columns from left to right are the recorded image and the reconstruction after 10,20,40 iterations

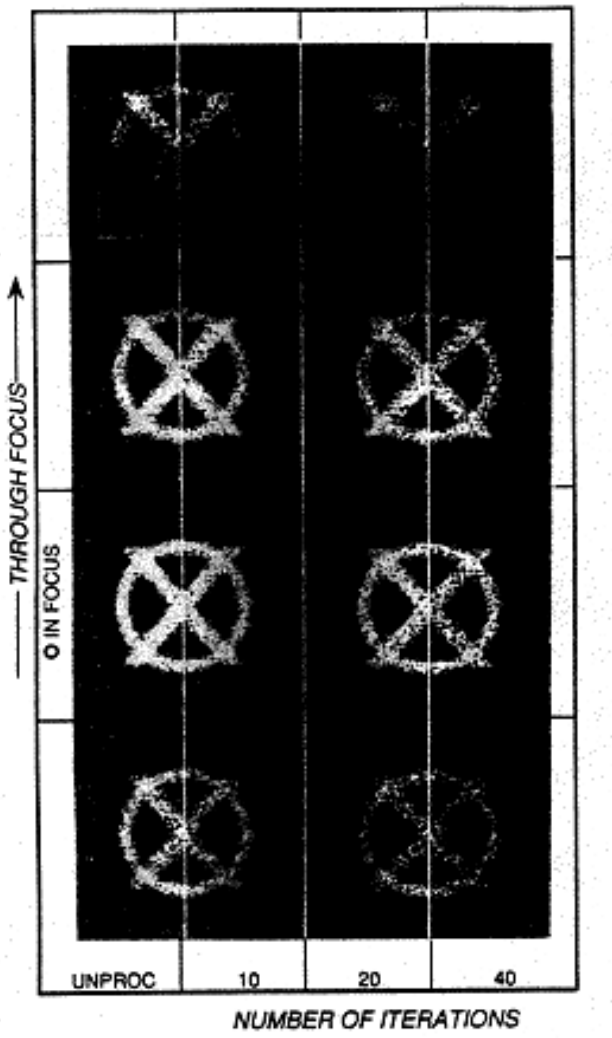


Figure 14: Additive algorithm results - with and poisson noise max. SNR=3.16. Left: simulated cross and circle of equal intensity 50 nm apart, the optical slice where the circle lays marked O. the columns from left to right are the recorded image and the reconstruction after 10,20,40 iterations.

4 Likelihood approach

4.1 Likelihood function

We assume that the specimen density is a poisson process with intensity $o(x, y, z)$ because of the fluorescence emission. The output image is also poisson process with intensity $o(x, y, z) \otimes h(x, y, z)$. The problem is to estimate the intensity of the specimen density given the output image.

A poisson density defined as

$$p(n, \lambda(x, y, z)) = \frac{\lambda^n(x, y, z)e^{\lambda(x,y,z)}}{n!} \tag{15}$$

By dividing the specimen and image to voxels l=1..N with λ_l as the poisson process density in voxel l. The probability of observing $i(1), i(2), ..., i(N)$ in the image is

$$p(i(1), i(2), .., i(N) \,|\, o(1), o(2), ..., o(N)) = \prod_{l=1}^{N} \left[\frac{1}{i(l)!} (o \otimes h)_{(l)}^{i(l)} e^{-(o \otimes h)_{(l)}} \right] \tag{16}$$

The a posteriori data likelihood probability $ln[p(o(1), o(2), ...|i(1), i(2), ...)]$ is by Bayes rule $\mathcal{L}(o(1), o(2), ...) = ln[p(i(1), i(2), ...|o(1), o(2), ...)] + ln[p(o(1), o(2), ...)] - ln[p(i(1), i(2), ...)]$. by omitting the terms that don't depend on $o(1), o(2), ...$ we get

$$\mathcal{L}(o(1), o(2), ...) = \Sigma_{l=1}^{N} i(l) ln[(o \otimes h)_{(l)}] - \Sigma_{l=1}^{N} (o \otimes h)_{(l)} + ln[p(o(1), o(2), ...)] \tag{17}$$

$P_{o(1),o(2),...}(o(1), o(2), ..., o(N))$ the probability density of the object intensity is unknown but this term enable us to insert constraints to the algorithm. this article don't specify the constraints they used but a previous article [5] used approximation for $p_{o(1),o(2),...}(o(1), o(2), ...)$ base on neighboring pixel correlation.

If the source element strengths $o(l)$ are quantized hypothetically into indistinguishable strength units (photons) and the total number $M = \Sigma_l o(l)$ of the strength units is assumed to be approximately fixed, the source (object) distribution may be characterized as a random process in which M indistinguishable strength units distribute randomly over the N

voxels, and each voxel has a probability $p_l(o(1), o(2), ...)$ of having $o(l)$ strength units. That means the a priori source information function can be expressed as

$$p(o(1), o(2), ...) = \frac{M!}{\prod_l o(l)!} \prod_{l=1}^{N} [p_l(o(1), o(2), ...)]^{o(l)} \tag{18}$$

$p_l(o(1), o(2), ...)$ is assumed to be a function of $o(l)$ and the neighboring source element correlation. This local correlation reflects the source strength continuity behavior and the physical interaction correlation from an element to its neighboring elements.

Let $\mathcal{X}_{lt}$ be the parameters characterizing the correlation of neighboring source elements around element l. The local correlation function $p_l(o(1), o(2), ...)$ can be expressed as

$$p_l(o(1), o(2), ...) = \prod_t \exp\left[-\frac{1}{2}\frac{\mathcal{X}_{lt}}{\sigma_l \sigma_t}(o(l) - \bar{o}(l))(o(t) - \bar{o}(t))\right] \tag{19}$$

where $t \in \{l \pm \Delta\}$ and Δ covers the neighboring elements of element l. σ_l and σ_t are the standard deviation of source elements l and t respectively, and $\bar{o}(l)$ is the mean strength value of voxel l.

The gaussian local correlation function in Eq. 19 incorporated the nearby source element continuity information and the physical interaction correlation information into the parameters $\{\mathcal{X}_{lt}\}$ and the mean values $\{\bar{o}(l)\}$. The gaussian parameters $\mathcal{X}_{lt}$,$\bar{o}(l)$ and σ_l need to be chosen arbitrary, a possible choice can be $\sigma_l^2 = \bar{o}(l)$, $\mathcal{X}_{l,l-2} = 0.1$, $\mathcal{X}_{l,l-1} = 0.5$, $\mathcal{X}_{l,l} = 1$, $\mathcal{X}_{l,l+1} = 0.5$, $\mathcal{X}_{l,l+2} = 0.1$ and the mean value can be updated by $\bar{o}(t) = o^{(k)}(t) + 2\eta(o_t^{(k)} - o_t^{(k-1)})$ where η is a learning rate constant.

For the minimization of the log likelihood we need to calculate the derivative of $ln[p(o(1), o(2), ...)]$ so from substituting Eq. 19 to Eq. 18 and derivating it we get

$$\frac{\partial ln[p(o(1), o(2), ...)]}{\partial o(l)} =$$

$$\frac{(2o(l) - \bar{o}(l))}{2\sigma_l} \sum_t -\frac{1}{2}\frac{\mathcal{X}_{lt}}{\sigma_t}(o(t) - \bar{o}(t)) + \frac{\mathcal{X}_{ll}}{2\sigma_l^2}(o(l) - \bar{o}(l))(3o(l) - \bar{o}(l)) + ln[o(l)] + \frac{1}{2\sigma_l} + 1 \tag{20}$$

.

4.2 Expectation Maximization

The expectation maximization algorithm is based on the concept of an incomplete data space $\mathcal{Y}$, a complete data space $\mathcal{Z}$ and a mapping $f : \mathcal{Z} \rightarrow Y$ between them [6].

The incomplete data space is the space $\mathcal{Y}$ in which measured data takes its values. in this problem $\mathcal{Y}$ is the image i and the incomplete data log-likelihood is the likelihood function we calculated in the previous subsection.
The complete data space $\mathcal{Z}$ is a hypothetical space that is contrived to accomplish two goals: 1) make the expectation and maximization steps of the algorithm analytically tractable. and 2) make the resulting computations required for the EM algorithm feasible for numerically producing estimates. The complete data space for a given problem is not unique, and the EM algorithm can be more or less complicated depending on the choice made.
The complete data space is larger than the incomplete data space in the sense that complete data must determine incomplete data. There must be a known function $f(\cdot)$ that maps complete data to incomplete data. This mapping function and the incomplete data y place a constraint determined by $f(z) = y$, on the values that the complete data z may have.

denote the collection of parameters to be estimated by θ, and assume that $\theta \in \Theta \subseteq \mathbb{R}^n$. The parameters may be deterministic or random. We use maximum likelihood estimation for deterministic parameters and maximum a posteriori probability estimation for random parameters. for the optical sectioning problem the parameters to be estimated are the object intensity o. and according to the previous section they are random variable, so we use the maximum a posteriori estimation.

Two log likelihood functions are important for estimating the parameters θ from measured data $y \in \mathcal{Y}$ when the EM algorithm is used. The incomplete data log likelihood $\mathcal{L}_{id}(\theta) = ln[p_y(Y|\theta)] + ln[p(\theta)]$. The maximum a posteriori probability estimate $\hat{\theta} \equiv \hat{\theta}(y)$ is the vector of the incomplete data set $y \in \mathcal{Y}$ that maximizes the incomplete data log likelihood $\mathcal{L}_{id}(\theta)$. For the optical sectioning problem the incomplete likelihood function is the one that we calculated in the previous subsection.

The second log likelihood that is important is the complete data log likelihood and it

is defined by $\mathcal{L}_{cd}(\theta) = ln[p_z(Z|\theta)] + ln[p(\theta)]$. $\hat{\theta}$ is not a maximizer of the complete data log likelihood. Nevertheless, $\mathcal{L}_{cd}(\theta)$ is important in the EM algorithm for determining $\hat{\theta}$ numerically.

The EM algorithm is iterative. Starting from an initial estimate $\hat{\theta}^{(0)}$ of the parameters, a sequence of estimates $\hat{\theta}^{(0)}, \hat{\theta}^{(1)}, \hat{\theta}^{(2)}, ...$ is produced for which the corresponding sequence of incomplete data log likelihood is nondecreasing $\mathcal{L}_{id}(\hat{\theta}^{(0)}) \leq L_{id}(\hat{\theta}^{(1)}) \leq L_{id}(\hat{\theta}^{(2)})....$ Two steps are required at each stage of the iteration to reach the next stage, an expectation (E) step and a maximization (N) step.

E-step. determines the conditional expectation of the complete data log likelihood,

$$Q(\theta|\hat{\theta}^{(k)}) = E[\mathcal{L}_{cd}(\theta)|y, \theta|\hat{\theta}^{(k)}] \tag{21}$$

M-step. determines the stage $k+1$ parameter estimate as the maximizer of $Q(\theta|\hat{\theta}^{(k)})$,

$$\hat{\theta}^{(k+1)} = \arg\max_{\theta\in\Theta}[Q(\theta|\hat{\theta}^{(k)})] \tag{22}$$

The sequence $\{\hat{\theta}^{(k)} : k = 1, 2, ...\}$ that is defined by the E and M steps corresponding to the sequence of incomplete data log likelihood $\{\mathcal{L}_{id}(\hat{\theta}^{(k)}) : k = 1, 2, ...\}$ is nondecreasing. The proof can be found at the appendix.

Back to the optical sectioning problem. Let $\mathcal{L}_{id}(o(1), o(2), ..., o(N)) = \mathcal{L}(o(1), o(2), ..., o(N))$ and the complete data set be the concatenation of the object probability density and the image intensity. The only terms of the log likelihood are those who depends on $o(1), o(2), ..., o(N)$ because others don't affect the M-step. so we simply get the log likelihood of the object probability density.

$$\mathcal{L}_{cd} = -\Sigma_{l=1}^{N} o(l) + \Sigma_{l=1}^{N} ln[o(l)]n_l + ln[p(o(1), o(2), ...)] \tag{23}$$

where n_l is the true density of the object at pixel l.

$$E\{\mathcal{L}_{cd}\} = -\Sigma_{l=1}^{N} o(l) + \Sigma_{l=1}^{N} ln[o(l)]E\{n_l|\hat{o}^k(l), i(1), i(2), ..., i(N)\} + ln[p(o(1), o(2), ...)] \tag{24}$$

the expectation expression $E\{n_l|o^k(l), i(1), i(2), ..., i(N)\}$ is expressed by (see development in the appendix)

$$E\{n_l|\hat{o}^k(l), i(1), i(2), ..., i(N)\} = \Sigma_{m=1}^{N} \frac{h(m-l)\hat{o}^k(l)}{(h \otimes \hat{o}^k)_(m)} i(m) \tag{25}$$

and the E-step is

$$\begin{aligned} &Q(o(1), o(2), ...|\hat{o(1)}^{(k)}, \hat{o(2)}^{(k)}, ...) = \\ &-\Sigma_{l=1}^{N} o(l) + \Sigma_{l=1}^{N} ln[o(l)] E\{n_l|\hat{o}^k(l), i(1), i(2), ..., i(N)\} + ln[p(o(1), o(2), ...)] = \\ &-\Sigma_{l=1}^{N} o(l) + \Sigma_{l=1}^{N} ln[o(l)] \Sigma_{m=1}^{N} \frac{h(m-l)\hat{o}^k(l)}{(h \otimes \hat{o}^k)_(m)} i(m) + ln[p(o(1), o(2), ...)] \end{aligned} \tag{26}$$

The function $Q(o(1), o(2), ...|\hat{o(1)}^{(k)}, \hat{o(2)}^{(k)}, ...)$ must now be maximized over $o(1), o(2), ...$ for the M-step. Using $\partial Q/\partial o(l) = 0$ we get $0 = -1 + \frac{1}{o(l)} E\{n_l|\hat{o}^k(l), i(1), i(2), ..., i(N)\} + \frac{\partial ln[p(o(1), o(2), ...)]}{\partial o(l)}$. By isolating $o(l)$ and substituting it with $\hat{o}^{k+1}(l)$ we get the next step estimate

$$\hat{o}^{k+1}(l) = \frac{\hat{o}^k(l) \Sigma_{m=1}^{N} \frac{h(m-l)}{(h \otimes \hat{o}^k)_(m)} i(m)}{1 + \frac{\partial ln[p(o(1), o(2), ...)]}{\partial o(l)}} \tag{27}$$

where $\frac{\partial ln[p(o(1), o(2), ...)]}{\partial o(l)}$ is calculated at Eq. 20 in the previous subsection.

4.3 Algorithm

The a priori data term is an estimate of the true density and is not precise. more over it is estimated iteratively with the algorithm progress. That is way we are not interested to enforce this a priori data completely from the first iterations. For this purpose we use a weight function $\xi^{(k)}$ that changes with iteration number. This function enable us to enforce the a priori data gradually on the estimation.

1. $\xi^{(k)} = \frac{C1k^\tau}{C2+k^\tau}$
2. $o^\star(l) = \hat{o}^k(l) + \eta(\hat{o}^k(l) - \hat{o}^{k-1}(l))$

3. $\bar{o}(l) = o^{(k)}(l) + 2\eta(o^{(k)}(l) - o^{(k-1)}(l))$

4. $Z^{(k)}(l) = \frac{(2o^{\star}(l)-\bar{o}(l))}{2\sigma_l} \sum_t -\frac{1}{2}\frac{\mathcal{X}_{lt}}{\sigma_t}(o^{\star}(t)-\bar{o}(t)) + \frac{\mathcal{X}_{ll}}{2\sigma_l^2}(o^{\star}(l)-\bar{o}(l))(3o^{\star}(l)-\bar{o}(l)) + ln[o^{\star}(l)] + \frac{1}{2\sigma_l} + 1$

5. $\hat{o}^{(k+1)}(l) = \frac{\hat{o}^{(k)}(l)\Sigma_{m=1}^{N}\frac{h(m-l)}{(h\otimes\hat{o}^{k})(m)}i(m)}{1+\xi^{(k)}Z^{(k)}(l)}$

6. k=k+1.

7. Repeat 1–6 until $\frac{|\mathcal{L}^{(k+1)}(o(1),o(2),...)-\mathcal{L}^{(k)}(o(1),o(2),...)|}{\mathcal{L}^{(k+1)}(o(1),o(2),...)} < \epsilon$

where $o^{\star}(l)$ is the a priori data estimate at iteration k, $\eta \approx 1$ is the a priori data estimation learning rate constant. For $\xi^{(k)}$ C1, C2 and τ are the weight function parameters that need to be chosen arbitrarily. Optional choice could be $C1 = 0.1$, $c2 = 100$ and $\tau = 1$. $\bar{o}(l)$ is the estimated mean of the a priori data probability density at voxel l and iteration k, and $Z^{(k)}(l)$ is the a priori data derivative that was calculated at Eq. 20 for voxel l and iteration k.

4.4 1D and 2D Simulation Results

Several simulations of different objects (smooth,sharp bright and sharp dim) were done both for Maximum Likelihood estimation and for Maximum a posteriori estimation. The first simulation was done on 1D objects sharp and smooth that can be seen in Fig. 15 and 16. The Maximum likelihood algorithm results and the Maximum a posteriori algorithm results are presented in Fig. 17, 19, 18 and 20.from these figures we can clearly see that although the Maximum Likelihood estimation is sharper and more accurate for sharp specimens, it also introduces artifacts to the image. Since the difference in the slope angle is minor, the maximum a posteriori is better for such objects especially in medical applications where artifacts can cause an unnecessary surgery operation. For the smooth object there is no difference in the slope angle but there is more artifacts in the ML estimation, and the object no longer looks smooth. We can derive another interesting information from the error figures 18 and 20. We can see that the MAP estimator is biassed and the error is always positive while the ML estimator is not biassed. The reason for this bias in the MAP estimation is the fact that the a priori knowledge we used for the estimation is not truly correct and it is more smoothing constraint than true data. Because we really don't know what the object

intensity is (we want to find it), we don't know what it's probability density. We are claiming that it is smooth and estimate it from neighboring voxels.
In all the cases Both ML estimation and MAP estimation are much better than the additive deconvolution algorithm results. It is more sharp and although there is artifacts in the ML estimate they are not worse than the artifacts in the additive deconvolution estimate.
The same conclusion are derived from the second simulation that was done on 2D objects, both smooth, sharp dim and sharp bright. In Fig. 21, 22 and 23 the different objects and system parameters are shown, and in Fig. 24, 25 and 26 the ML and MAP algorithm results are shown. The artifacts in the ML estimate are very obvious in the smooth object Fig. 24, where a grid of dark and bright lines is seen all over the picture. For the sharp and bright object it is obvious from Fig. 25 that the ML estimate achieves better results than the MAP estimate. The object edges are clearer and sharper although the artifacts are still present in the picture. for the sharp and dim object Fig. 26, neither algorithm achieve good results. The MAP estimate is not sharp enough and the object is blurred into the background. While in the ML estimate the object is sharp but it is swallowed up in the artifacts at the background.

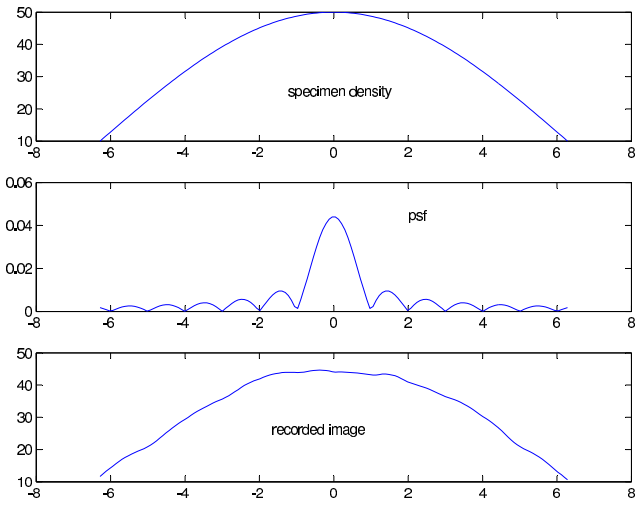


Figure 15: system parameters for smooth object

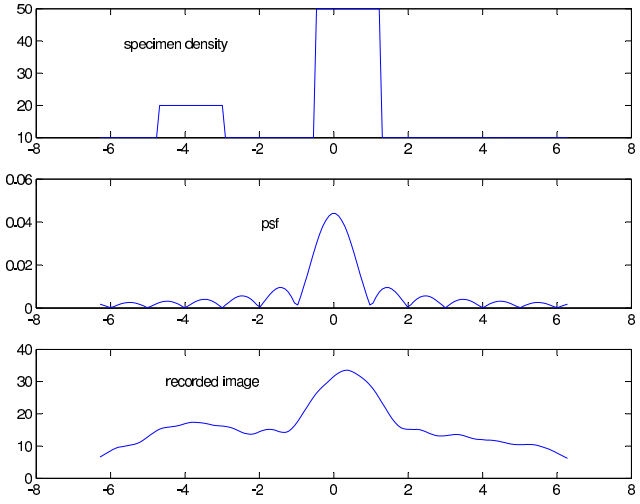


Figure 16: system parameters for sharp object

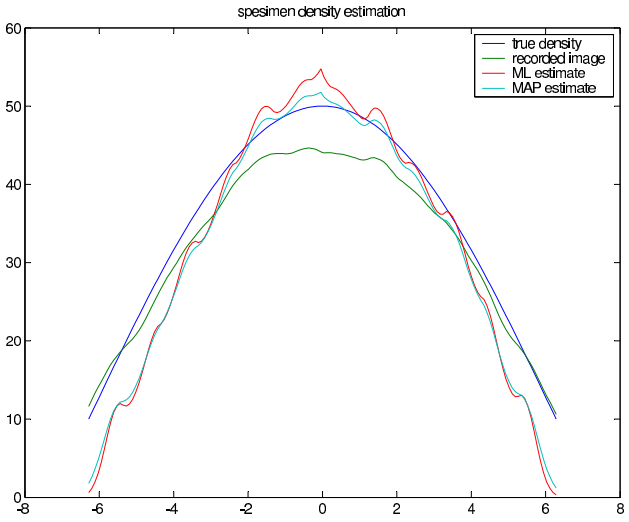


Figure 17: ML and MAP algorithm results for smooth object

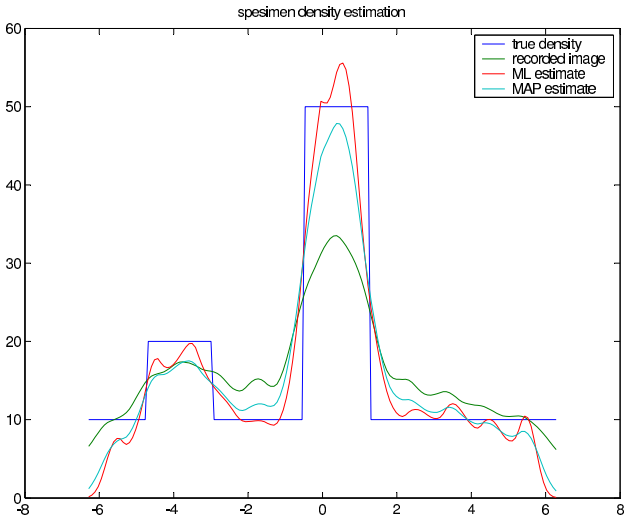


Figure 18: ML and MAP algorithm results for sharp object

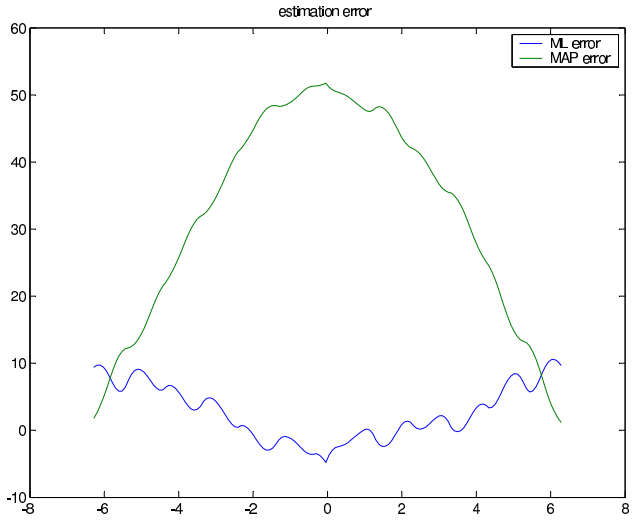


Figure 19: ML and MAP algorithm results for smooth object

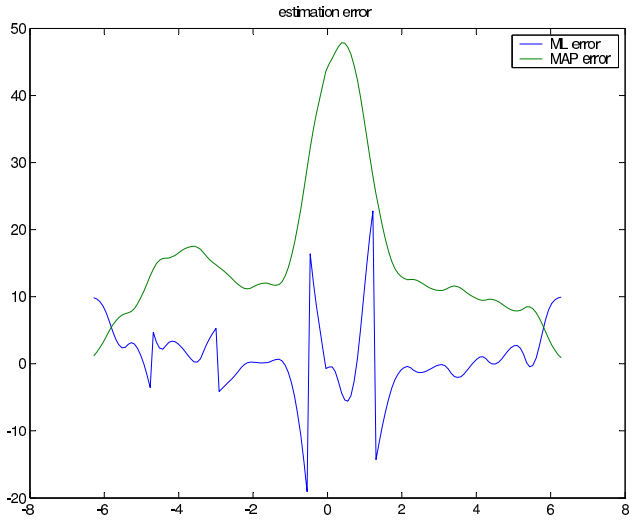


Figure 20: ML and MAP algorithm results for smooth object

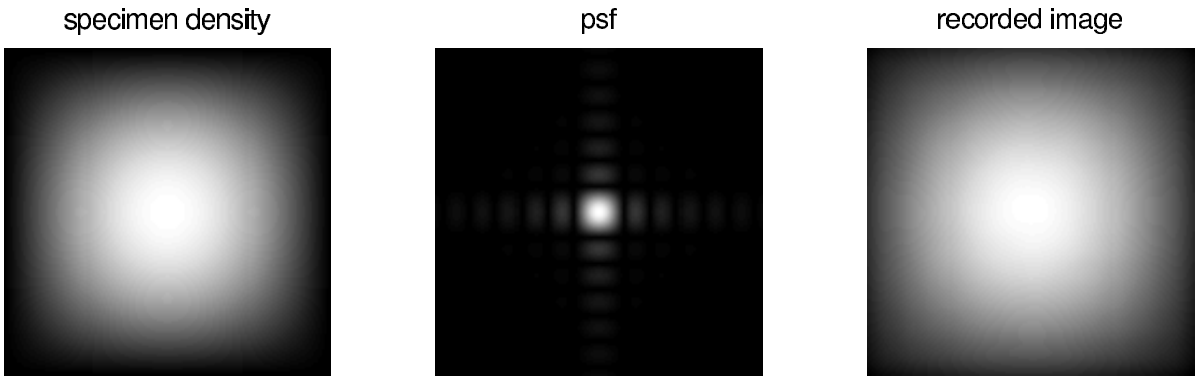


Figure 21: system parameters of smooth object

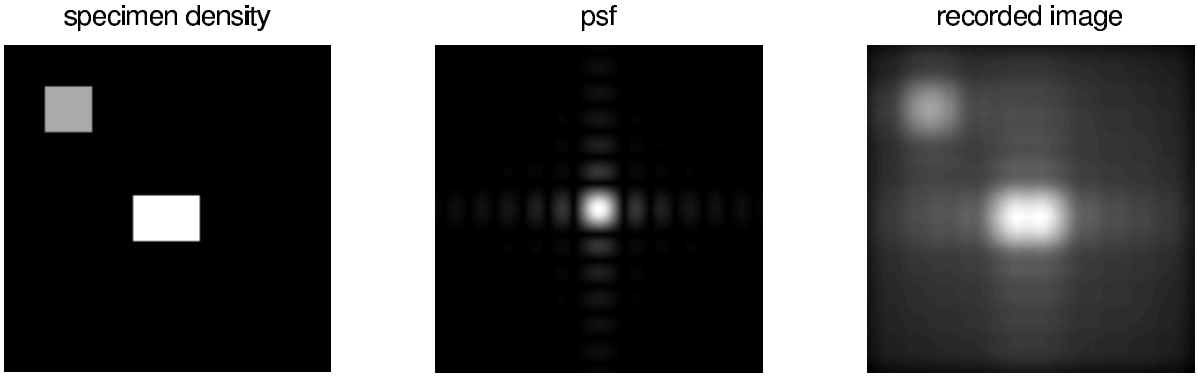


Figure 22: system parameters of sharp bright object

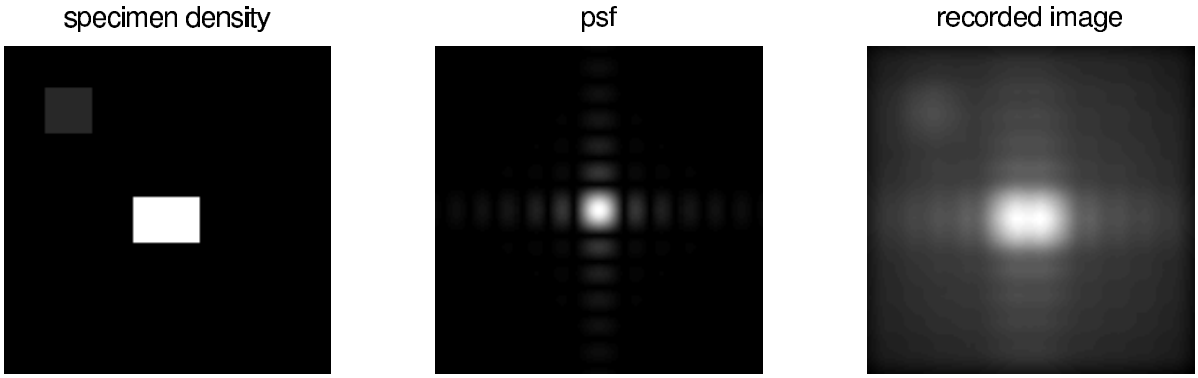


Figure 23: system parameters of dim bright object

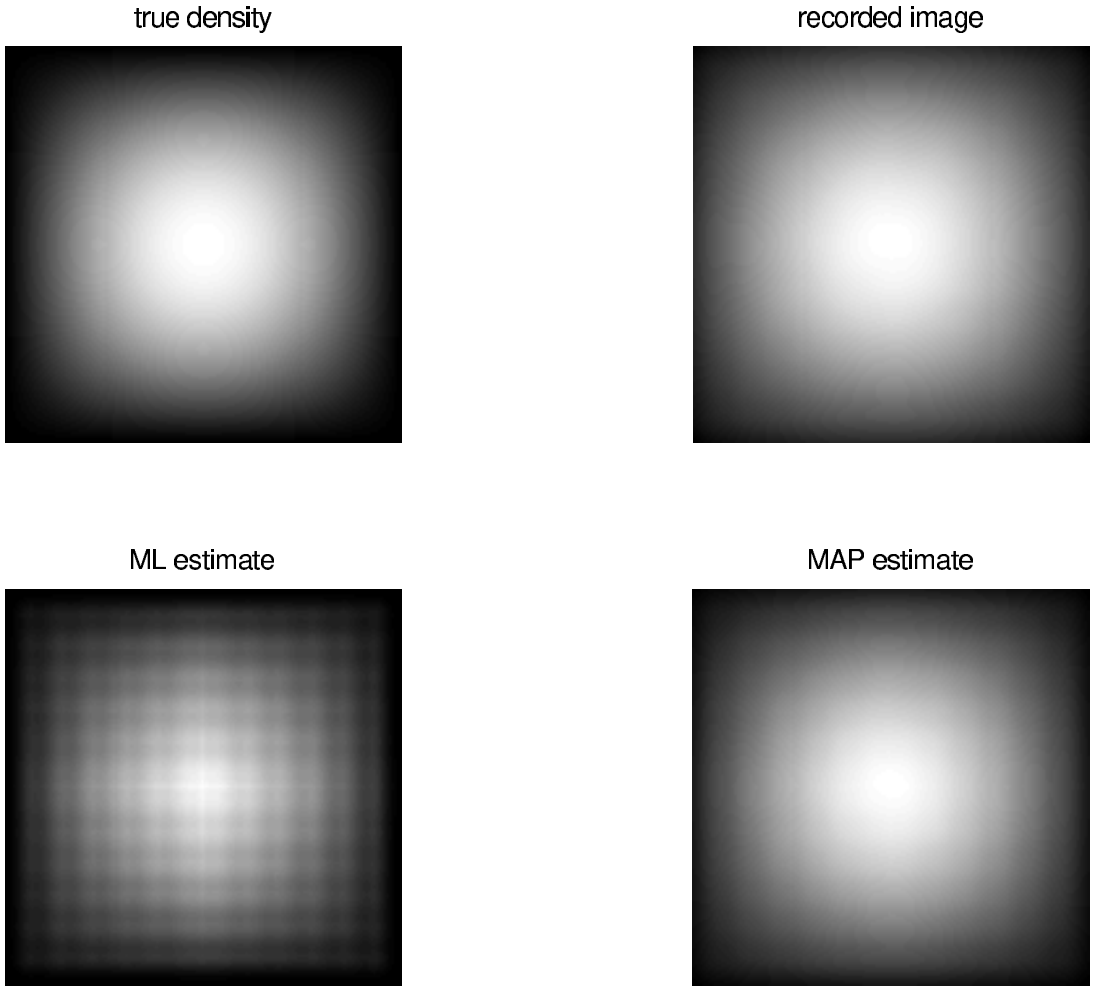


Figure 24: ML and MAP algorithm results for smooth object

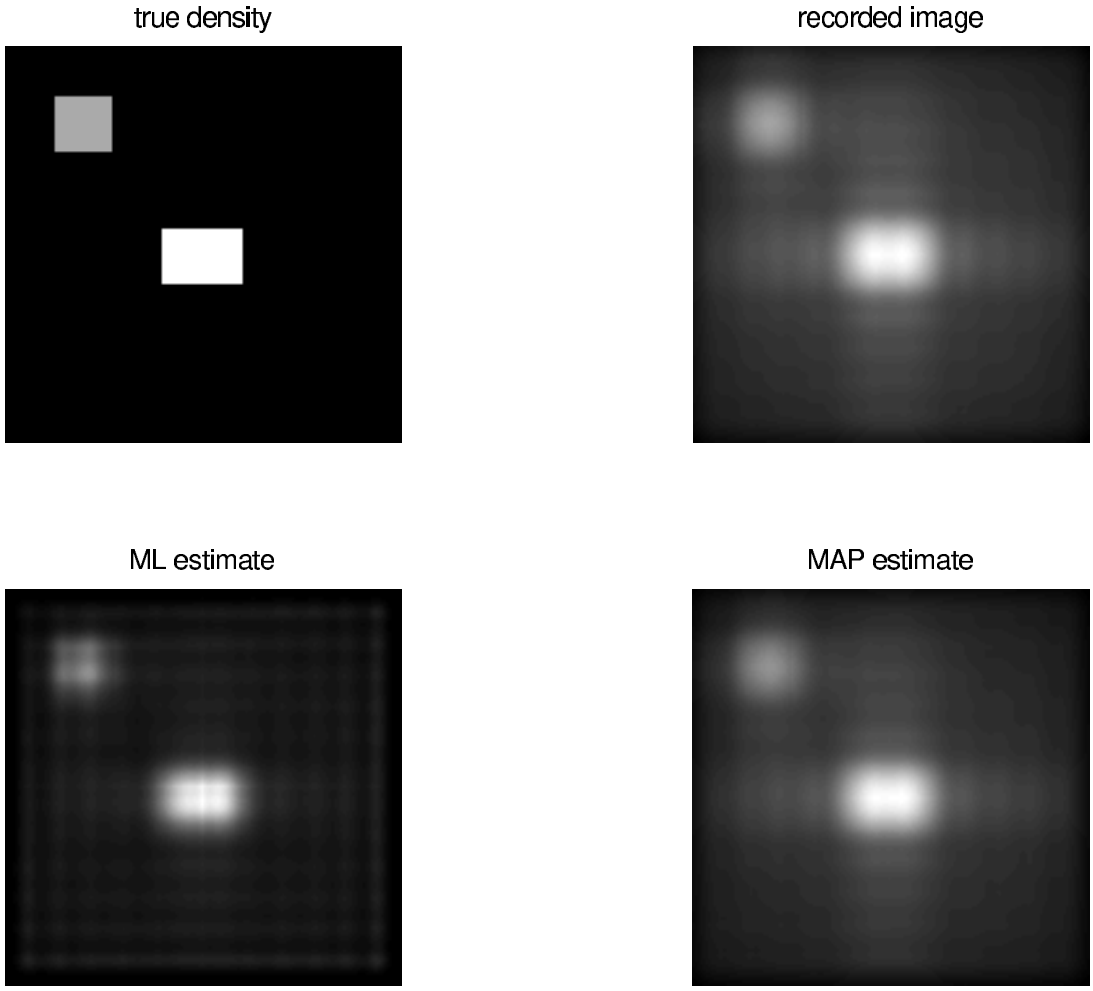


Figure 25: ML and MAP algorithm results for sharp bright object

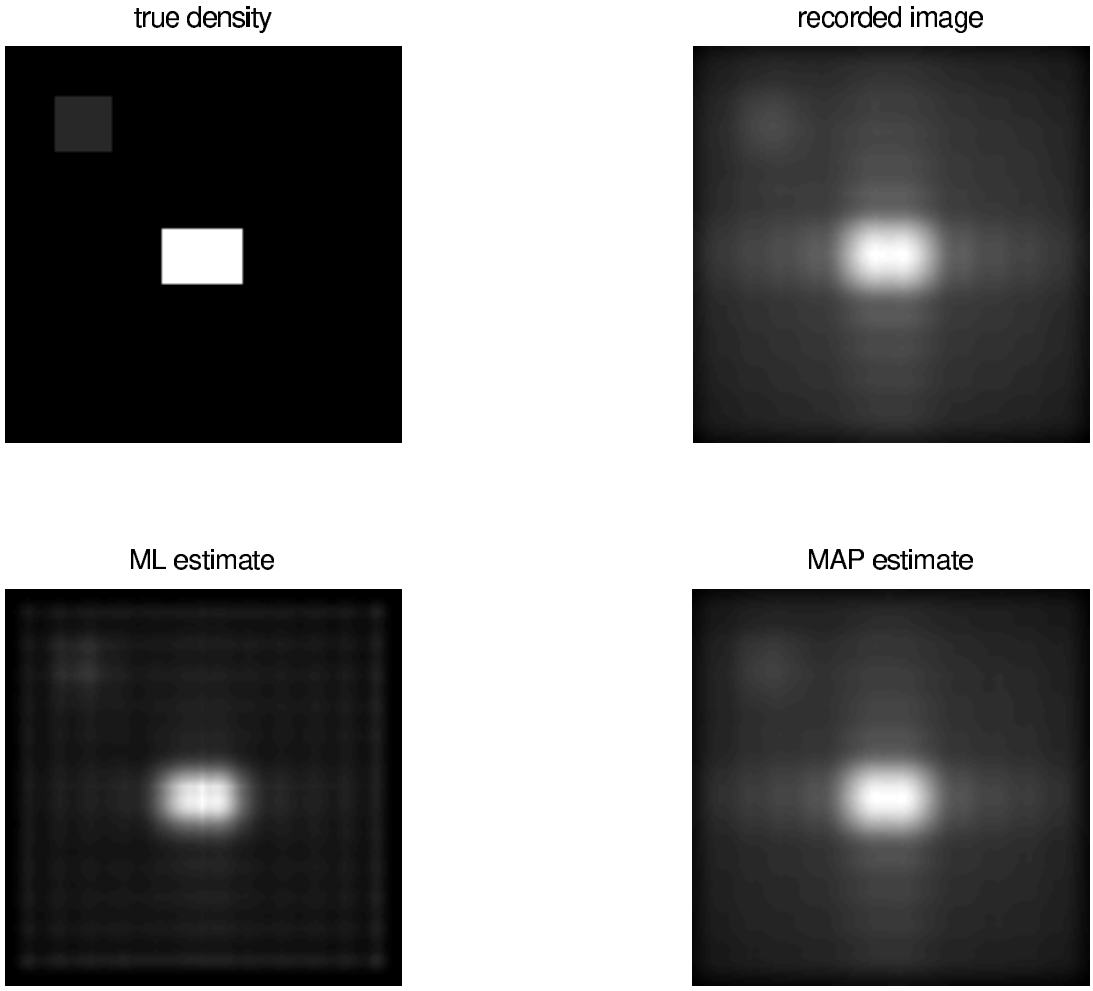


Figure 26: ML and MAP algorithm results for sharp dim object

4.5 Article Results

The likelihood algorithm was tested on both noiseless and noisy images. The first case was the cross and the circle with equal strength and 200 nm apart. The results for this case are shown in Fig. 27. As with the additive deconvolution, the likelihood algorithm reduces the out of focus contributions to the planes immediately above and below the plane where the object is. When compared with the additive deconvolution results, we can see that 20 iterations of the likelihood algorithm results in better reduction of the out of focus contributions than 40 iterations of the additive algorithm. We can also observe that the ringing in the out of focus planes is less marked for the likelihood algorithm.

The algorithm was then tested with the same image corrupted with a signal dependent poisson noise with SNR of 3.16 and 1. The results are shown in Fig. 28(left) and (right) respectively. When the SNR was 3.16, the likelihood gave better reconstruction than the additive deconvolution algorithm. For 20 iterations of the likelihood algorithm the out of focus contributions are dimmer than those obtained after 40 iterations of the additive algorithm. When the SNR was 1, the algorithm reduces the out of focus contributions but, the noise was amplified at the same time and after 20 iterations it is not clear that the reconstructed image is better than the recorded data.

The maximum likelihood was then applied to the case of a dim cross 200 nm above the circle. Two imaging conditions were simulated noiseless and with poisson noise with SNR=3.16. The results are shown in Fig. 29 (left) and (right) respectively. these results are showing us again that the likelihood is performing better than the additive algorithm and also that the artifacts are less severe.

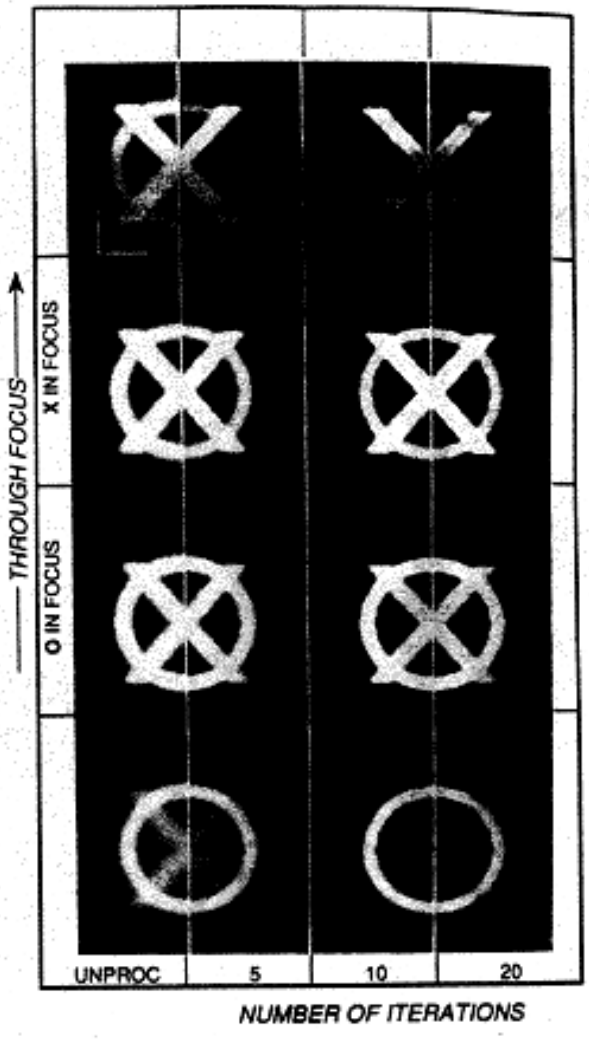


Figure 27: MAP algorithm results - without noise. simulated cross and circle of equal intensity 200 nm apart, the optical slices where the cross and circle lays marked X and O respectively. the columns from left to right are the recorded image and the reconstruction after 5,10,20 iterations.

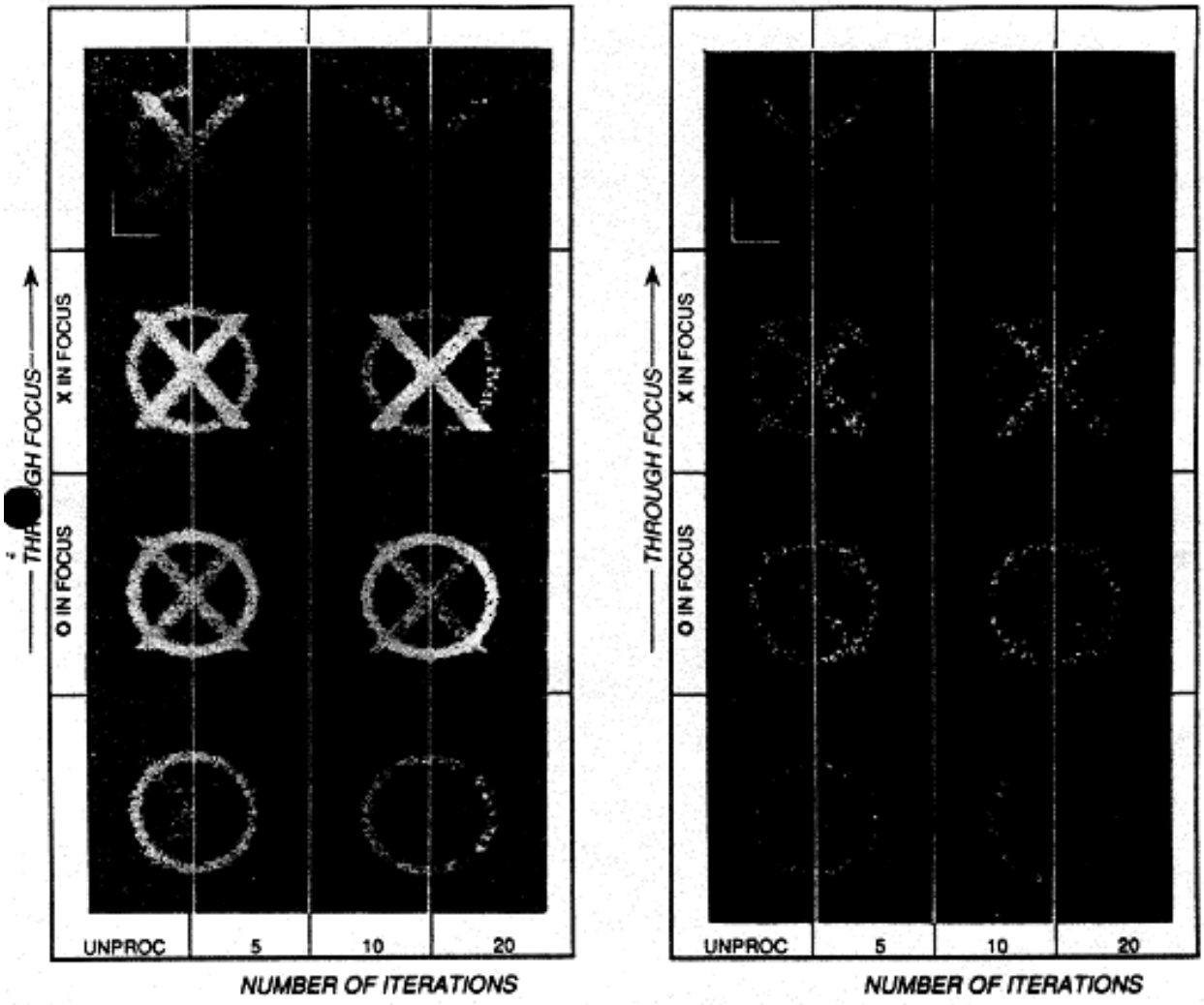


Figure 28: MAP algorithm results - with poisson noise. simulated cross and circle of equal intensity 200 nm apart, the optical slices where the cross and circle lays marked X and O respectively. the columns from left to right are the recorded image and the reconstruction after 5,10,20 iterations. Left: max SNR=3.16 Right: max SNR=1.

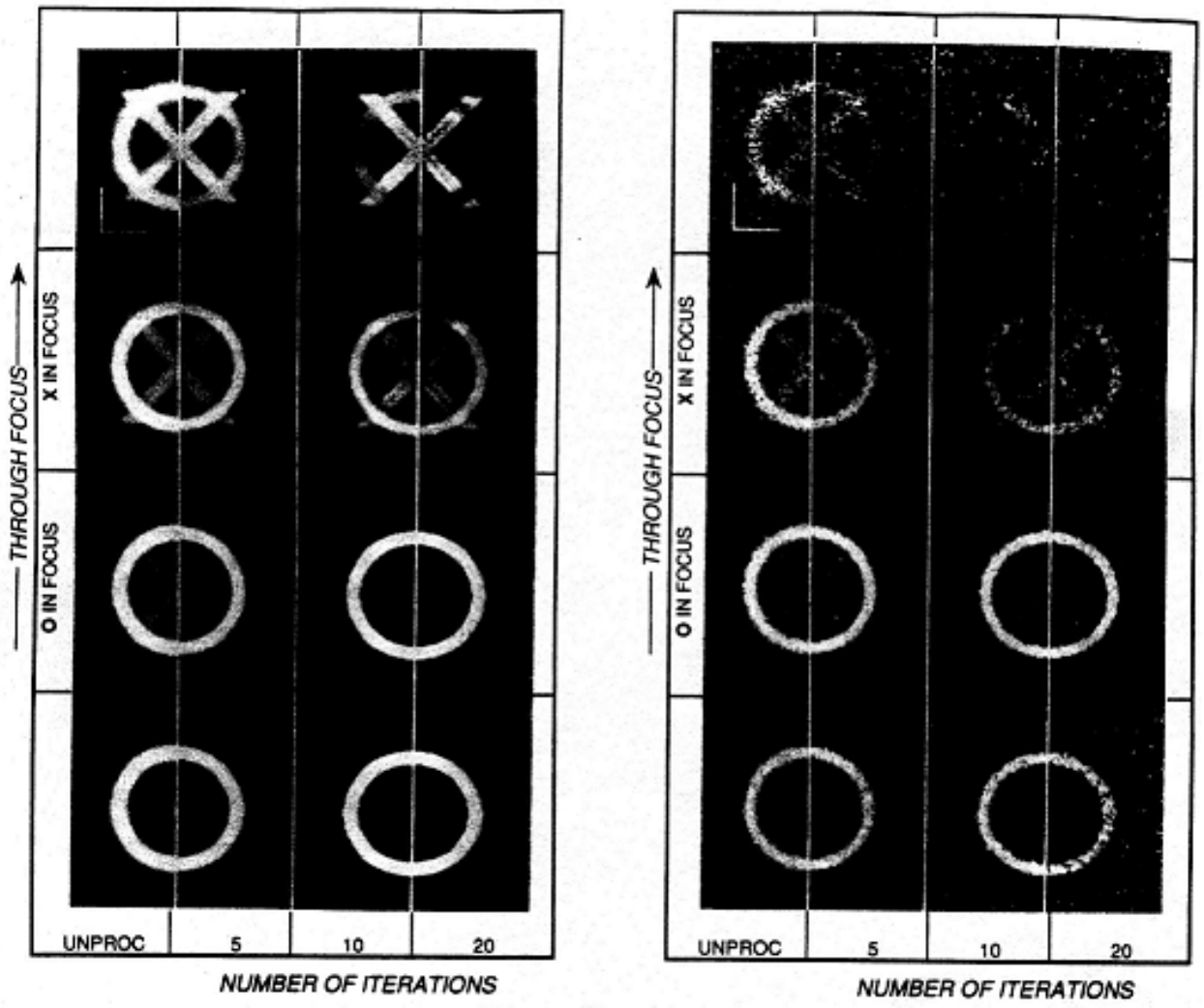


Figure 29: MAP algorithm results - with and without noise. simulated cross and circle 200 nm apart, the optical slices where the cross and circle lays marked X and O respectively the cross strength is 1/10 of the circle strength. the columns from left to right are the recorded image and the reconstruction after 5,10,20 iterations. Left: without noise. Right: max SNR=3.16

5 Discussion

The Maximum likelihood and Maximum a posteriori algorithms has several advantages over additive deconvolution. Even though each iteration of the formers takes about three times as much as the latter, fewer iterations are required to achieve about the same resolution. Both likelihood algorithms were observed to be more tolerant of noise and the MAP algorithm produces less severe artifacts.

Further more, the article has considered only the idealized case of an infinitesimal detector aperture. Such a configuration, while providing the highest prior resolution, also yields the lowest prior SNR in practical situations. While opening the detector aperture to admit more light improves the SNR, it is also admits more light from out of focus planes and degrades the resolution. The detector aperture may be adjusted to achieve an optimum balance, from the perspective of the post processor, thus tuning the microscope and the post processor as a system.

Finally, since all the methods discussed in this paper are nonlinear, they can potentially recover components of the specimen in the null space of the imaging operator.

6 Future work

The next step in reconstruction of volumetric data from optical sectioning by wide field or confocal microscope is blind deconvolution. that is simultaneous reconstruction of object and the point spread function. In this article it is assumed that the optical transfer function of the microscope is known. The conventional methods to acquire the information about the optical transfer function are: use a point source as the object and measure the point spread function, or use an analytical formula like the one we developed in section 2 and get the required lens parameters from the manufacturer. The limitation of the first method is that it is very hard to produce a small enough source of light for this problem, because the distances in this problem are very small to begin with. The limitation of the second method is the fact that not always the required parameters are accurate enough and there are variations between the manufacturer data and the actual lens. More over if you got the

volumetric data from an unknown source you don't have the information on the microscope at all.

There was some work done on blind deconvolution of microscopic data by several researchers. Holmes [7] used the EM algorithm for the estimation of both the point spread function and the object. Markham and Conchello [8] suggested an integrated algorithm based on parametric model for the point spread function and EM algorithm for the object. Carasso [9] suggested another method that don't assume smoothness of the object and summarized some other methods as well.

7 Appendix A - proof for the EM algorithm convergence

Theorem 1 *(EM algorithm convergence). The sequence* $\{\hat{\theta}^{(k)} : k = 1, 2, ...\}$ *that defined by the E and M steps corresponding to the sequence of incomplete data log likelihood* $\{\mathcal{L}_{id}(\hat{\theta}^{(k)}) : k = 1, 2, ...\}$ *is nondecreasing.*

Proof. For the proof of this problem not from ... and ... that for $Z \in \mathcal{Z}(Y)$ *there holds*

$$ln[p_y(Y|\theta)] = ln[p_z(Z|\theta)] - ln[p_{z|y}(Z|Y,\theta)] \tag{28}$$

The left size of this equation is the incomplete data log likelihood, and the first term on the right is the complete data log likelihood. Multiplying both sides by $p_{z|y}(Z|Y,\hat{\theta}^{(k)})$*, and integrating with respect to Z yields*

$$\mathcal{L}_{id} = Q(\theta|\hat{\theta}^{(k)}) - \int_{\mathcal{Z}(Y)} p_{z|y}(Z|Y,\hat{\theta}^{(k)}) ln[p_{z|y}(Z|Y,\theta)]\, dZ \tag{29}$$

Evaluating Eq. 30 for both $\theta = \hat{\theta}^{(k+1)}$ *and* $\theta = \hat{\theta}^{(k)}$*, and then subtracting, results in*

$$\begin{aligned} \mathcal{L}_{id}(\hat{\theta}^{(k+1)}) - \mathcal{L}_{id}(\hat{\theta}^{(k)}) = Q(\theta|\hat{\theta}^{(k+1)}) - Q(\theta|\hat{\theta}^{(k)}) - \\ \int_{\mathcal{Z}(Y)} p_{z|y}(Z|Y,\hat{\theta}^{(k)}) ln\left[\frac{p_{z|y}(Z|Y,\hat{\theta}^{(k+1)})}{p_{z|y}(Z|Y,\hat{\theta}^{(k)})}\right] dZ \end{aligned} \tag{30}$$

Application of the inequality $ln(x) \leq x - 1$, for which equality holds if and only if $x = 1$, then yields

$$\mathcal{L}_{id}(\hat{\theta}^{(k+1)}) - \mathcal{L}_{id}(\hat{\theta}^{(k)}) \geq Q(\theta|\hat{\theta}^{(k+1)}) - Q(\theta|\hat{\theta}^{(k)}) - \int_{\mathcal{Z}(Y)} p_{z|y}(Z|Y,\hat{\theta}^{(k)}) \left[\frac{p_{z|y}(Z|Y,\hat{\theta}^{(k+1)})}{p_{z|y}(Z|Y,\hat{\theta}^{(k)})} - 1 \right] dZ = Q(\theta|\hat{\theta}^{(k+1)}) - Q(\theta|\hat{\theta}^{(k)}) \tag{31}$$

with equality if and only if $p_{z|y}(Z|Y,\hat{\theta}^{(k+1)}) = p_{z|y}(Z|Y,\hat{\theta}^{(k)})$, which holds if $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)}$. The M-step ... implies that the right side of ... is nonnegative in general and equal to zero if $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)}$. Hence, $\mathcal{L}_{id}(\hat{\theta}^{(k+1)}) \geq \mathcal{L}_{id}(\hat{\theta}^{(k)})$. which establishes the theorem.
□

8 Appendix B - calculating $E\{n_l|o^k(l), i(1), i(2), ..., i(N)\}$

Theorem 2 *(conditional mean for histogram data). Let $\{N(A) : A \subseteq \mathcal{X}\}$ be a poisson process with an integrable intensity function $\{\lambda(x) : x \in \mathcal{X}\}$. Points of this process are translated to the output space $\mathcal{Y}$ to form the output point process $\{N(B) : B \subseteq \mathcal{Y}\}$, where each point is independently translated according to the transition density $p(y|x)$. Let $\{\mathcal{Y}_1, \mathcal{Y}_2, ...\}$ be disjoint sets in $\mathcal{Y}$ such that $\mathcal{Y} = \bigcup_{k=1}^{\infty} \mathcal{Y}_k$. Define $M(\mathcal{Y}_k)$ to be the number of points in $\mathcal{Y}_k$, and let $M = \{M(\mathcal{Y}_1), M(\mathcal{Y}_2), ...\}$ denote histogram data derived from points in the output space. Then if there are no insertions and deletions, the conditional expectation $E[N(A)|M]$ of the number of points in a subset A of the input space $\mathcal{X}$ given the histogram data M is*

$$E[N(A)|M] = \Sigma_{k=1}^{\infty} \left[\frac{\int_{\mathcal{Y}_k} \int_A p(y|x)\lambda(x)\, dxdy}{\int_{\mathcal{Y}_k} \int_{\mathcal{X}} p(y|x)\lambda(x)\, dxdy} \right] M(\mathcal{Y}_k) \tag{32}$$

Proof. Let $M(A; \mathcal{Y}_k)$ be the number of points in $\mathcal{Y}_k$ that are translated from A. points in the output space that are translated from the subset A of the input space form a Poisson process on $\mathcal{Y}$ with intensity $\int_A p(y|x)\lambda(x)\, dx$. $N(A) = \Sigma_{k=1}^{\infty} M(A; \mathcal{Y}_k)$, from which it follows that

$$E[N(A)|M(\mathcal{Y}_1), M(\mathcal{Y}_2)...] = \Sigma_{k=1}^{\infty} E[N(A, \mathcal{Y}_k)|M(\mathcal{Y}_1), M(\mathcal{Y}_2)...] \tag{33}$$

$$E[N(A,\mathcal{Y}_k)|M(\mathcal{Y}_1),M(\mathcal{Y}_2)...] = E[N(A,\mathcal{Y}_k)|M(\mathcal{Y}_k)] = \\ = \frac{\mu(A;\mathcal{Y}_k)}{\mu(A;\mathcal{Y}_k)+\mu(A^c;\mathcal{Y}_k)} M(\mathcal{Y}_k) \tag{34}$$

where $\mu(A;\mathcal{Y}_k) = E[M(A;\mathcal{Y}_k)] = \int_{\mathcal{Y}_k}\int_A p(y|x)\lambda(x)\,dxdy$ *is the average number of points in* $\mathcal{Y}_k$ *that were translated from* A*, and* $\mu(A^c;\mathcal{Y}_k) = E[M(A^c;\mathcal{Y}_k)] = \int_{\mathcal{Y}_k}\int_{A^c} p(y|x)\lambda(x)\,dxdy$ *is the average number of points in* $\mathcal{Y}_k$ *that were translated from the complement* $A^c = \mathcal{X} - A$ *of* A*. Substituting these expressions into Eq. 33 yields*

$$E[N(A)|M(\mathcal{Y}_1),M(\mathcal{Y}_2)...] = \Sigma_{k=1}^{\infty}\left[\frac{\mu(A;\mathcal{Y}_k)}{\mu(A;\mathcal{Y}_k)+\mu(A^c;\mathcal{Y}_k)}\right] M(\mathcal{Y}_k) \tag{35}$$

which is Eq. 32

□

for our reconstruction problem: substitute $N(A)$ with $n(l)$, $M(\mathcal{Y}_l)$ with $o^{(k)}(l)$, $\lambda(l)$ with $o(l)$, and $p(y|x)$ with $h(m-l)$. Then $\mu(A;\mathcal{Y}_k)$ equals $h(m-l)o(l)$ and Eq. 35 denominator is equal to $(h\otimes o)(m)$ so we get Eq. 27.

References

[1] S. Kimura and C. Munakata, "Calculation of three-dimensional optical transfer function for a confocal scanning fluorescent microscope," *J.Opt.Soc.Am.A*, vol. 6, pp. 1015–1019, July 1989.

[2] H. H. Hopkins, "The frequency response of a defocused optical system," *Proc.R.Soc.London Ser.A*, vol. 231, pp. 91–103, 1955.

[3] P. Stokseth, "Properties of a defocused optical system," *J.Opt.Soc.Am*, vol. 59, pp. 1314–1321, 1969.

[4] D. A. Agard, "Optical sectioning microscopy: cellular architecture in three dimensions," *Ann. Rev. Biophys. Bioeng.*, vol. 13, pp. 191–219, 1984.

[5] Z. Liang, "Statistical models of a priori information for image processing: neighboring correlation constraints," *J.Opt.Soc.Am.A*, vol. 5, pp. 2026–2031, December 1988.

[6] D. L. Snyder and M. I. Miller, *Random point processes in time and space.* New York: Springer-Verlag, 1991.

[7] H. Timothy J, "Blind deconvolution of quantum limited incoherent imagery: maximum-likelihood approach," *J.Opt.Soc.Am.A*, vol. 9, pp. 1052–1061, July 1992.

[8] J. Markham and J.-A. Conchello, "Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur," *J.Opt.Soc.Am.A*, vol. 16, pp. 2377–2391, October 1999.

[9] C. Alfred S, "Linear and nonlinear image deblurring a documented study," *SIAM J, Numer. Anal.*, vol. 36, no. 6, pp. 1659–1689, 1999.