3D image deconvolution in light microscopy: theory and practice.

by Johan Philip and Kjell Carlsson
Royal Institute of Technology, 10044 Stockholm, Sweden.

Abstract: The objective of this study is to investigate the use of deconvolution for improving the three-dimensional resolution in both confocal and wide-field light microscopy. We start by assessing the theoretical limits of the deconvolution by defining and calculating a *characteristic function* which describes the imaging-deconvolution process. We continue by suggesting a variant of inverse filtering for deconvolving real noisy data and describe the application of the method to a confocal recording of a biological preparation. The result is comparable with that obtainable by Wiener filtering but we can avoid the problem of finding a best value of the parameter. We also describe a way of determining the optical transfer function of the microscope.

Keywords: Light microscopy, 3D imaging, deconvolution, characteristic function.

1. Introduction

Three-dimensional microscopy is by now an established technique for extracting information from a biological preparation. Both confocal and wide-field microscope systems for three-dimensional (3D) imaging are available commercially. The recording results in a data file holding the intensities of the voxels in a 3D grid. These data are usually presented as a sequence of ordinary images or as a stereo pair. Our objective is deconvolution of such data and reduction of noise. The resolution obtainable by deconvolution is not better than the voxel size used in the recording. This means that deconvolution is motivated only if the voxel side is smaller than the resolution of the microscope. Interesting voxel sides are of the order of 100 nm, which is approximately one fifth of the light wavelength.

We shall introduce and calculate the *characteristic function* of the imaging-deconvolution process. This function, which can be defined for any linear process, describes the theoretical limits of deconvolution and will be used to visualize these limits. The deconvolution technique that we are going to describe is based on inverse filtering and resembles Wiener filtering. The novelty of our method consists of taking advantage of the fact that the Fourier transform of the 3D point spread function has a support occupying only a small fraction of the whole 3D array. Using this, we can combine suppression of noise in the data with only a minor loss of information about the object. In two-dimensional microscopy there is no such small support and our method is not appropriate in 2D because the noise will ruin the deconvolution. A further advantage with our method is that one doesn't have to choose any parameters or test several values of a parameter

as is the case with regularization methods like Wiener filtering. We use one parameter which is calculated from a measurement of the point spread function of the microscope. Our method doesn't require any assumptions about probability distributions for the object and noise.
We demonstrate the performance of the method by applying the characteristic function to two artificial objects with and without noise. We suggest a method for exploiting the fact that the intensity of the object is nonnegative and apply it to an artificial object and continue by presenting a simple way of determining the optical transfer function of the microscope. In the last section, we describe the result of applying our method to a confocal recording of a biological specimen and compare the result with that obtained by Wiener filtering. The results turn out to be similar provided the Wiener filtering is performed with the "right" value of the signal-to-noise ratio. Finding this right value may require time consuming tests with several values.
Before the advent of the computer, neutron scattering data were deconvolved by paper, pencil, and scissors. Computerized deconvolution of power spectra in one dimension has with growing power of the computer been generalized to two and three dimensions. Erhardt et. al. (1985) [4] describe a deconvolution method for 3D microscopy which is similar to ours in using inverse filtering. Other early methods are given by Garza et. al. (1987) [5], Agard et. al. (1989) [1], and Shaw and Rawlins (1991) [10]. A recent comparison of various methods is presented by Verveer et. al. (1998) [13]. A good account of the mathematical-statistical background for handling measurements of weak intensities is given by Vardi et. al. (1985) [12]. They consider positron emission tomography but the theory is applicable to 3D microscopy.

2. The mathematical model and method.

We assume that the microscope is a linear instrument, meaning that it can be characterized by a point-spread-function (psf) which is the same for all points of the object. As the name indicates this function gives the pattern recorded when the object is a radiating point. Linearity is at hand when (i) the recording of a point source is proportional to its intensity and (ii) the recording from several points is the sum of the individual recordings. Mathematically, this implies that the microscopic recording can be described by the following equation

$$\iiint h(x-\xi, y-\eta, z-\zeta) f(\xi, \eta, \zeta) d\xi d\eta d\zeta \;=\; d(x, y, z) \qquad , \tag{2.1}$$

where h is the point spread function, f is the intensity distribution radiated from the studied object and d is the recorded data. The combination of h and f in (2.1) is called convolution and is denoted by a $*$, so that (2.1) can be abbreviated $h*f = d$.

We are fully aware that this model is only an approximately correct description of the recording process, but it is good enough for giving interesting results.

In the calculations, h, f, and d are not functions of the continuous variables x, y, andz ranging from $-\infty$ to ∞ but are discretizations of these functions over a finite three-dimensional grid. We let n_x and n_y denote the number of meshes in the transversal x- and y-directions and let n_z denote the number in the axial z-direction. The actual form of (2.1) is

$$\sum_{i=0}^{n_x-1}\sum_{j=0}^{n_y-1}\sum_{k=0}^{n_z-1} h(r-i, s-j, t-k)f(i, j, k) = d(r, s, t) \quad , \tag{2.2}$$

where $0 \le r \le n_x - 1$, $0 \le s \le n_y - 1$, and $0 \le t \le n_z - 1$. Since negative arguments occur in h in (2.2), the domain of definition must be extended cyclically in all three directions

$$h(i, j, k) = h(i + n_x, j, k) = h(i, j + n_y, k) = h(i, j, k + n_z) \quad . \tag{2.3}$$

The same cyclic extension of the domain shall be applied to f and d.

The discrete Fourier transform (DFT) is defined for cyclically defined functions and has the form

$$H(I, J, K) = \sum_{i=0}^{n_x-1}\sum_{j=0}^{n_y-1}\sum_{k=0}^{n_z-1} h(i, j, k)\, e^{2\pi i\left(\frac{iI}{n_x}+\frac{jJ}{n_y}+\frac{kK}{n_z}\right)} \quad . \tag{2.4}$$

When h is the point-spread-function of a microscope its DFT H is called the optical transfer function (OTF) of the microscope. We shall use the symbol F for the DFT operation writing $H = Fh$. The linear transformation F has a linear inverse F^{-1} being so similar to F that the same computer code can be used for calculating both F and F^{-1}.

The DFT is of interest here because it transforms cyclic convolution to multiplication

$$F(f*h) = Ff \cdot Fh \quad . \tag{2.5}$$

Explicitely, (2.2) is transformed to

$$H(I, J, K) \cdot F(I, J, K) = D(I, J, K) \quad . \tag{2.6}$$

This equation can easily be solved for the DFT of the unknown object f

$$F(I, J, K) = \frac{D(I, J, K)}{H(I, J, K)} \quad . \tag{2.7}$$

Applying the inverse DFT will produce the unknown $f = F^{-1}F$. This calculation is called *inverse filtering*. The simplicity of this formalism makes it very attractive and this is amplified by the fact that there exists a fast Fourier transform (FFT) algorithm for doing the calculations. Trying to use (2.7) in practice reveals several problems due to noisy data, that $H(I, J, K)$ has zeroes, and that the convolution performed by the microscope is not cyclic.

As we shall demonstrate, these problems can at least partly be mastered when H is the OTF of

a confocal microscope.

3. Example and figure standards.

All numerical examples with artificial data are carried out on a grid with $64 \times 64 \times 64$ voxels. The grid size is chosen to give a suitable fine structure of the figures. The mesh, that is the length of the voxel side, is $0{,}1\mu m$. The mesh size has been chosen equal to the mesh size of the real data studied in section 9. The figures are calculated for a microscope having an oil immersion objective with a magnification of 100 and a numerical aperture 1.3, ($\sin\alpha = 0{,}8$). The microscope is assumed to be ideal but for the detector aperture which isn't a point but has a diameter of 25 μm.

For any function g (which can be a psf, OTF, etc.) , the figures show the two-dimensional section $g(0, j, k)$ for the functions and $|G(0, J, K)|$ for the DFT:s. All plotted functions are normalized to have the maximal value one. For functions taking negative values the plots are for values in the range $(-0{,}2, 1)$.

4. The optical transfer function.

Let h_{wf} denote the wide-field psf and define the 'disk function'

$$a(x, y, z) = \begin{cases} 1, & x^2 + y^2 \le \rho^2, z = 0 \\ 0, & \text{otherwise} \end{cases}, \tag{4.1}$$

where $\rho = (12{,}5/100)\mu m$ is the back projected radius of the detector aperture. The confocal psf is $(h_{wf} * a)h_{wf}$. In the sequel, we will let h stand for either of these two psfs. A radius ρ of the size used here has only a minor influnce on h, Kimura and Munakata (1990) [6] and Wilson (1995) [14].

Since diffraction essentially performs a Fourier transformation, the wide-field OTF can be found directly from the geometry of the microscope, Sheppard et. al (1994) [11], and Philip (1999) [8]. These papers show that the support of the OTF, i.e. the set of points where it is different from zero, is a well defined set occupying only a small fraction of the whole discrete DFT array. This fact is crucial for the successful deconvolution of noisy data.

Assuming that the psf is rotationally symmetric around the optical axis (z-axis), we also have an OTF that is rotationally symmetric

$$H(I, J, K) = H(0, \sqrt{I^2 + J^2}, K) \ . \tag{4.2}$$

We show in Figure 1 the wide-field and the confocal psf and OTF.

As is seen in Figure 1, the confocal OTF support extends in the lateral direction beyond the Nyquist frequency, which is at the boundary of the diagram. This means that in theory, under perfect conditions, the mesh size $.1\ \mu m$ is too coarse for the used numerical aperture and detector aperture.

5. The characteristic function.

The object of this section is showing that the whole recording and deconvolution process can be described by a single function, which we shall call a characteristic function. This function describes the possibilities and limitations of the whole imaging process.

Let us denote the imaging process of the microscope (I) followed by the deconvolution process (D) by DI, writing $f_{deconv} = DIf$. Ideally, we would like DI to be the identity operator so that applying it to the object f would make f_{deconv} a true image of f. Since we are assuming that the microscope recording is a linear process and the following deconvolution is a linear process, it follows that DI is a linear operator.

Let δ be a point object having unit intensity and sitting at the origin of an object coordinate system. Define the *characteristic function* c by $c = DI\delta$. We shall determine c and show that it can be used for computing f_{deconv} for any object f.

The three-dimensional image of δ equals $d = \delta * h = h$ i. e. the psf sitting at the origin of the image space. Having $D = H$ and using (2.7), the DFT of the deconvolution becomes $F_{deconv} = \frac{D}{H} = \frac{H}{H}$. Wherever $H \neq 0$, this ratio is one. Where $H = 0$, the ratio is unknown and there is no reason to give it any other value than zero. We define

$$U(I, J, K) = \begin{cases} 1 & (I, J, K) \in \text{ support of} H \\ 0 & \text{elsewhere} \end{cases} \tag{5.1}$$

and put $F_{deconv} = U$ for the point object. Completing the deconvolution process by taking the inverse Fourier transform of U we obtain $c = F^{-1}U$.

Having a rotationally symmetric H, we get rotationally symmetric U and c. We display $c(0, j, k)$ for a wide-field and a confocal microscope in Figure 2. The figures show that the point object results in an image that is extended, in particular in the axial direction implying that not even in theory is it possible to obtain a true image of an object. The central peak of the characteristic function c is narrower than the psf but much wider than the one-point function δ as is

seen in the sections of Figure 2. The oscillations of c result in degradations of the deconvolution that are hard to describe. The imperfection of the imaging-deconvolution process stems from the limited support of the OTF.

Considering an arbitrary object f, we have the data $d = h*f$ with the DFT $D = H \cdot F$ implying that $D = 0$ in all points having $H = 0$. According to (2.7), we get the DFT of the deconvolution $F_{deconv} = D/H = F$ wherever $H \neq 0$. In other points, we put $F_{deconv} = 0$ implying that $F_{deconv} = F \cdot U$. We get

$$f_{deconv} = DIf = F^{-1}(F_{deconv}) = F^{-1}(F \cdot U) = F^{-1}(F)*F^{-1}(U) = f*c \quad . \qquad (5.2)$$

The formula $f_{deconv} = f * c$ implies that c completely describes the DI process.

6. Computer experiments with artificial objects and with noise.

Even if the characteristic function describes the imaging-deconvolution process completely it is clarifying to study more complicated objects than a point. The figures displayed in this section show the theoretical results obtainable under the best conditions. Here, the convolution and deconvolution kernels are identical and both are cyclical.

Judging the quality of a deconvolution is to a great extent subjective and is depending on the studied object and on the purpose of the process.

We have chosen to report the results for two artificial objects. The first consists of four cubes, three having the constant intensity = 1 and one having the intensity .25. The cube sides are 1 μm and the distance between them is $0{,}4\mu m$. We have chosen this object with sharp corners and avoided a round object to be able to study the degradation of the resolution. The second object consists of two parallel square slabs with side 2 μm, distance 1 μm, and thickness .2 μm. With our objects, we can study the possibility of separating objects having equal or different intensity and having different extension in radial and axial direction. The objects and their DFTs are shown in the top Figures 3 and 4.

The result of applying first the I (convolution with the psf) and then the D (deconvolution) operator to the objects are given in the bottom of Figures 3 and 4.

For the object in Figure 3, the wide-field deconvolution is almost as good as the confocal one because the cubes have the same extent in the radial and axial directions. For the wider and thinner slabs in Figure 4, there is a great difference in depth discrimination between the two microscopes.

We continue the described experiments by assuming that the data, i.e. the convolution values, are obtained from a photon counting detector. First, the data is normalized so that its maximal inten-

sity becomes 1000 photons/pixel in the wide-field case and 100 photons/pixel in the confocal case. These numbers shall reflect that the wide-field image is brighter and we have chosen it to be ten times brighter than the confocal image. Then, for every pixel, the function value, say v, is replaced by a variate drawn from a poisson distribution with parameter v .

For short, we shall call the so obtained data noisy. The noisy data are shown in the top of Figures 5 and 6, where one can see that the confocal data are noisier than the wide-field data.

Deconvolution of noisy data is always a compromise between getting a good resolution and suppressing noise. Here, it means that we cannot use (2.7) in the whole support of H since that would mean dividing the DFT of the noise with very small values of H near the boundary of the support and so amplifying the noise very much. A common way of handling this problem is using Wiener filtering. Here, we shall use a method with cut levels and shall compare it with Wiener filtering in section 10. The cut level method is implemented by dividing by H only when its modulus exceeds the cut level and put the remaining $F_{deconv} = 0$. With artificial data and well controlled noise, one can experiment with the cutlevel. In the deconvolutions displayed in Figures 5 and 6, we have used a cut level of .03 in the wide-field and .1 in the confocal case. For real data, we shall suggest a way of determining the cut level.

The relatively small influence of the noise is due to the 'low-pass filtering' of the data taking place when the DFT of the deconvolution is zeroed outside the used part of the support of H. While most of the information of the object sits inside the used support, the noise, being white, has its spectrum evenly distributed over the whole DFT cube. The noise supression is better in the wide-field case, because the H-support is smaller than in the confocal case. We define the error amplification of the deconvolution process as

$$error\ amplification = \sqrt{\frac{\text{noise energy of deconvolved data}}{\text{noise energy of input data}}} \ . \qquad (6.1)$$

Figure7 shows how the error amplification varies with the cut level. This level is the only parameter we are using and we call it $cutlevel$.

The abrupt decrease of F_{deconv} at the boundary of the support tends to introduce oscillations in the deconvolution. The remedy is smoothing of F_{deconv}. When working with real data, we implement the smoothing by putting

$$F_{deconv} = \begin{cases} \dfrac{D\bar{H}}{|H|^2} & |H| \geq cutlevel \\ \dfrac{D\bar{H}}{cutlevel^2} & cutlevel > |H| \geq cut2 \\ 0 & cut2 > |H| \end{cases}, \qquad (6.2)$$

where $\bar{H}$ denotes the complex conjugate of H and $cut2$ is some second cut level. This means that we smooth by stopping the decrease of the denominator $|H|^2$ when it reaches $cutlevel^2$. Since both D and $\bar{H}$ decrease when we depart from the used support, F_{deconv} will rapidly decrease towards zero. With real data, we put $cut2 = cutlevel/3$. In our examples, h is symmetric implying $\bar{H} = H$.

In the confocal case, the support of H has a great extension in the lateral direction allowing much noise to influence the deconvolution. In order to reduce the noise, one can put $F_{deconv} = 0$ for large lateral frequencies. This 'lateral low pass filtering' will reduce the lateral resolution but, because the noise is reduced, it may increase the possibility of discerning changes of intensity in the axial direction.

7. Exploiting positivity.

Several attempts have been made to use the information that the radiation from the object is non-negative. The described methods involve lenghty iterative procedures which don't converge but are interrupted when somebody thinks the result looks good, see e.g. [1] and [9].

We shall describe a convergent iterative method. It is seldom worth while running it more than five iterations. For the artificial objects, it produces a slight improvement of the deconvolution. For real images obtained from biological specimens, the method is hardly worth using because any improvement is hard to assess.

The principle of our method is that the calculated estimate of F inside the used support of H is the best one we can get. Outside the support, we have no information about F and have put it equal to zero but are free to give it any other values if that can make f nonnegative.

Our method starts with the obtained deconvolution f and goes as follows.

1. Extract the negative part f_- of f .

2. Calculate its DFT F_- .

3. Add $-\gamma F_-$ to the values of F outside the used support of H, where γ is an overrelaxation

factor. (The values of F inside the used support of H are not changed.)

4. Apply inverse DFT to F .

5. Return to 1. or stop.

Let N be the operator taking the negative part of a function so that $f_- = Nf$ and let P stand for the projection on the used support of H. Denoting the kth iterate by f^k, we can write the iterative step as

$$f^{k+1} = (I - \gamma F^{-1}(I-P)FN)f^k \quad . \tag{7.1}$$

Here, $F^{-1}(I-P)F$ is a linear operator having L_2-norm one. N is nonlinear and we have $\|Nf\| \le \|f\|$ implying $\|F^{-1}(I-P)FNf\| \le \|f\|$. If (7.1) had been a linear process, the theory of overrelaxation would have required $0 < \gamma < 2$ for convergence, see e.g. [3]. Since we have no theory for our nonlinear method, we have tested the convergence for various γ and found that $\gamma = 1{,}8$ gives good results.

The convergence can be checked by following the most negative value of f . In the first confocal example above, the input to the iteration has $f_{min} = -0{,}278$. After five iterations, $f_{min} = -0{,}028$ and after another five iterations, $f_{min} = -0{,}018$. Also the norm of the difference between successive iterates is decreasing.

8. Experimental determination of the OTF.

For determining the psf or the OTF one would like to have a one voxel object with high intensity. Realizing that such an object is hard to construct, we have instead used many (about 20) small fluorescent beads with diameter = .3 μm , dispersed in a transparent gel. Our interest is focused on the support of the OTF. We shall describe two methods for obtaining the OTF or at least its support.

The first method, included for comparison, is to select a subvolume of the data containing the image of one bead and take the DFT. This can be repeated for several beads for obtaining several estimates to average over. In practice, only few beads come out well separated and strong enough for this process. The result, which is the DFT of the convolution of the psf by a bead, is shown at the top of Figure 8.

The second method, which we advocate because it is so simple, is to take the DFT of the whole volume. The idea is that the image is the sum of say N equal images of the same bead b placed in the positions $\boldsymbol{x}_k$, $(1 \le k \le N)$. Here, $\boldsymbol{x}_k$ is a three-dimensional vector and the image is

$$B(\boldsymbol{x}) \;=\; \sum_{k\,=\,1}^{N} w_k b(\boldsymbol{x}-\boldsymbol{x}_k) \quad , \tag{8.1}$$

where w_k is a factor describing the intensity of bead k. The DFT is

$$\hat{B}(\xi) \;=\; \hat{b}(\xi) \sum_{k\,=\,1}^{N} w_k e^{-i(\xi\,\cdot\,x_k)} \quad . \tag{8.2}$$

For $\xi \neq 0$, the modulus of the sum in (8.2) is a random variable generated from the random positions $\boldsymbol{x}_k$. For N reasonably large the modulus of this random variable is fairly stable. For $\xi \;=\; 0$, the sum equals $\left| \sum_{k\,=\,1}^{N} w_k \right|$, which is much bigger than the sum of the random variables. The value at the origin is put to zero in the DFT:s shown in the middle of Figure 8. In both methods, we are assuming that the psf has circular symmetry around the microscope axis and this transfers to the DFT, so we have formed averages over circles around the z-axis. The results obtained by our second method (middle) are very similar to those obtained by the cumbersome first method (top). The DFT of a bead of the used size is non-zero in the studied area.

The measured DFT:s should be compared with the theoretically computed DFT of the convolution of a bead with radius $.3\ \mu m$ and the psf, being displayed in the bottom of Figure 8. The difference between theory and practice indicates that the real microscope isn't as good as the theory predicts.

9. Deconvolution of real data.

We shall describe the deconvolution of a $512 \times 512 \times 128$ voxel biological specimen of lung tissue recorded by a confocal microscope. The voxel side $= 0{,}1\mu m$.

When doing deconvolution by Fourier methods, one must remember that the DFT corresponds to *cyclic* convolution. This means that the volume under consideration is assumed to be repeated indefinitely in the x-, y-, and z-directions. The opposite boundary sides must be considered as adjacent. This will be visible in the deconvolution in particular making the structure in the top of the volume appear in the bottom of the deconvolution and vice versa. Constructing artificial objects we avoided any abrupt change of intensity between the top and bottom layers but for real data such changes are present causing oscillations in the deconvolution. Such oscillations, which go through the whole volume, can be reduced by smoothing the data across the boundary between the top and bottom layer. In the test described below, we have smoothed the data over five layers at the top and the bottom of the volume. The same smoothing is undertaken in the x-, and y-

directions.

A key problem is whether to use the theoretical or the measured OTF in the deconvolution. In Figure 8, the two OTF:s have the same shape but the theoretical one seems to have been computed for too big a mesh size. First, we tried deconvolving the real data using the measured OTF. Encountering various artifacts in the result we started studying the DFT of the deconvolution by plotting a section of it as in Figure 9. In this figure, we have plotted the modulus of the DFTs of the input data and the deconvolution as functions of K for $I = J = 0$. The values for $K = 0$ and $K = 1$ are omitted in the figure. For the data, the moduli have the typical decreasing trend of an image without special patterns. For the deconvolution, the moduli increase before they drop to zero outside the used support at $K = 18$. Obviously, the convolution frequencies close to the boundary of the used support have been amplified too much.

Our practical deconvolution is based on the following two principles:

1. Determine the support of the measured OTF using the second method of section 8.
2. Use the theoretical OTF, inside the measured support.

We will use the theoretical rather than the measured OTF for the deconvolution because we dont't know the origin of their difference. It may be aberrations, mis-alignment of lenses, aperture, or light source. The psf is probably different in different parts of the object. We treat all these errors as random errors and will not try to compensate for them.

Instead, we use the conservative approach of compensating only for the well specified action of the theoretical psf. As noted above, over-compensation is detrimental to the result.

We shall use, however, the knowledge that the spectrum outside the measured support of the OTF does not hold any information about the object but only noise.

Practically, this is implemented by determining the $cutlevel$ so that the used support of the theoretical OTF approximates the support of the measured OTF. This $cutlevel$ exists because the theoretical support is bigger than the measured one.

In the presented example, the calculated value of $cutlevel$ turned out to be .3. (The OTF is normalized to one at the origin.) This value and $cut2 = cutlevel/3 = 0{,}1$ was used in (6.2). Experimenting with other values of $cutlevel$ in the range .2 - .4 resulted in small changes of the result. The value of $cut2$ isn't very critical because $|D|$ can be expected to drop faster than H so that there will be a very small jump at $cut2$. We always check that the chosen cuts are big enough for preventing F_{deconv} from increasing like in Figure 9.

The supports defined by $cutlevel = 0{,}3$ and $cut2 = 0{,}1$ constitute respectively the fractions .0384 and .1145 of the whole DFT volume. This means that 11.45 % of the noise is amplified and 88.55 % (outside $cut2$) is extinguished. This combines to an error amplification of .76, thus

in fact an attenuation.

Figure 10 presents a yz-section of a subset of the data containing values above the background. We show also the 'filtered' data resulting from the removal all frequencies outside the used support of H. The improved resolution of the deconvolution is best seen in comparison with the filtered data. The structures have become thinner particularly in the axial (= vertical) direction. Because only a small fraction of the deconvolution becomes negative, the five iterations with the positivity algorithm have little effect.

The effect of performing cyclic deconvolution of the noncyclically recorded data is seen in the bottom of the deconvolution around 12 on the scale. It is the intensity at the top of the data that has wrapped around. By the smoothing of the data across five layers in top and bottom, the results in these layers are anyhow not reliable.

To illustrate the importance of the smoothing of F_{deconv} with $cut2$, we show in the top left corner of Figure 11 the deconvolution without smoothing. The oscillations in the axial direction are clearly visible.

The FFT of $512 \times 512 \times 128$ data takes 5 minutes on our SUN4u sparc SUNW, Ultra-Enterprise, so a whole deconvolution takes a little more than 10 minutes. Any further iteration with the positivity algorithm takes another 10 minutes.

10. Comparison with Wiener filtering

Using the Fourier Wiener filter implies replacing (2.7) by

$$F_{deconv} = \frac{D\overline{H}}{|H|^2 + \alpha} \tag{10.1}$$

where $\alpha = P_n/P_f$ and P_n and P_f are respectively the power spectra of the noise and the object, see [2]. All the quantities in (10.1) vary with the not displayed index set (I, J, K). Generally, P_n and P_f are not known in detail and one puts $\alpha = (SNR)^{-2}$, where SNR is the signal-to-noise ratio. Notice that $F_{deconv} = 0$ where $\overline{H} = 0$ in (10.1) so that the obtained values are similar to those obtained by (6.2). We have compared the results from using (6.2) and (10.1) on the artificial objects. The reasons for preferring the cut level method to Wiener filtering are

1) To limit the noise in Wiener filtering one should not use (10.1) as it stands but use a cut level like $cut2$ and put $F_{deconv} = 0$ when $|H| < cut2$.

2) For our artificial data with noise, we had $SNR = 7{,}7$. Using Wiener filtering with $\alpha = 7{,}7^{-2}$ resulted in three times more noise in the deconvolution than with the cut level method. Increasing

α to $0{,}3^2$, which is the value used by the cut level method, resulted in about the same noise as in that method. Generally, the SNR is not known. We prefer using the cut level method with $cutlevel$ determined from the measured OTF to experimenting with several values of α. For real data, there is no good criteria for selecting the best α. The result of Wiener filtering with three different values of α is shown in Figure 11. The input data is the same as in Figure 10. The sensitivity to the value of α is clearly visible and it is hard to determine which convolution to choose as the most reliable.

11. Discussion.

We have deduced the characteristic function for the recording process of a microscope followed by a deconvolution process for improving the resolution. The characteristic function describes the theoretical limits for what can be achieved by deconvolution. The application of a wide-field and a confocal characteristic function to two artificial objects is demonstrated both with and without noise.

Continuing with real data, we start by suggesting a simple method for determining the OTF of the microscope. In the latter part of the paper, we describe how the theories can be used for improving the resolution of a biological specimen. The success of the method relies heavily on the fact that the 3D OTF has such a small support so that most of the noise can be removed. By actively preventing the method from producing oscillations, very few deconvolved values will violate the positivity condition. As a consequence, the iterative positivity algorithm suggested produces very little improvement. Roughly speaking, the deconvolution process boils down to limiting noise and amplifying some frequencies, though using some ingenuity.

Statistical inference is about estimating a few parameters from many data. Here, we have as many parameters as we have data so it doesn't matter if we assume gaussian noise or poisson data, use maximum likelihood, maximal entropy, least squares, or a baysian approach. As long as there is only one observation per variable, the observation is the best estimate we have. There is no ground for assuming any particular prior probability distribution of the object or any correlation between the noise contributions to different voxels.

Attempting to determine more parameters than there are data, as in blind deconvolution, is no good statistical practice. The supplementing of some inequalities is inadequate for determining many more paramaters than there are data. Even if the number of sought parameters is reduced, Markham and Conchello (1996) [7], an optimization algorithm will produce a solution on the boundary, which by convexity theory is known to consist of extreme points manifesting themselves as bright spots in the image. Trying to remove such spots by smoothing, regularization, or generalized Wiener filtering is equivalent to partly undoing the deconvolution. Regularized

solutions, like Wiener filtered solutions, are known to be very sensitive to the choice of the regularization parameter.
The reason why blind deconvolution produces something like a solution is that it doesn't matter so much what kernel is deconvolved as long as it isn't too wide and that the iteration is started with the data and is interrupted before it starts oscillating.

12. Acknowledgement.

The Swedish Research Council for Engineering Sciences provided financial support, The lung tissue specimen was provided by Eben Oldmixon, Brown University.

13. References

[1] D.A. Agard, Y. Hiraoka, P. Shaw, and J.W. Sedat, “Fluorescence Microscopy in Three Dimansions”, Methods in Cell Biology, vol. 30, Academic Press, 1989, pp. 353-377.

[2] H.C. Andrews and B.R. Hunt, “Digital Image Restoration”, Prentice Hall, NJ, 1972.

[3] G. Dahlquist, Å. Björck, and N. Anderson, “Numerical Methods”, Prentice-Hall, 1974.

[4] A. Erhardt, G. Zinser, D. Komitowski, and J. Bille, “Reconstructing 3-D light-microscopic images by digital image processing”, Applied Optics, vol. 24, no. 2, 15 January 1985, pp. 194-200.

[5] F.M. Garza, A.C. Bovik, K.R. Diller, J.K. Aggarwal, S.J. Aggarwal, “Optical 3D Reconstruction of Sectionally Sectioned Microscopic Images by Computer”, Proc. IASTED Intern. Symposium on Siganal Processing and its Applications, Brisbane, Australia, Aug. 24-28, 1987.

[6] S. Kimura and C. Munakata, “dependence of 3-D optical transfer functions on the pinhole radius in a fluorescent confocal optical microscope”, Applied Optics, vol. 29, no. 20, July 1990, pp. 3007-3010.

[7] J. Markham and J-A Conchello, “Parametric blind deconvolution of microscopic images: Further results”, Proc. SPIE-96, vol. 3261, pp. 38-49.

[8] J. Philip, “Optical transfer function in 3D for a large numerical aperture”, J, Modern Optics, vol. 46, no. 6, 1999, pp.1031-1042..

[9] P.J. Shaw and D.J. Rawlins, “Measurements of point spread function and its use in deconvolution of confocal microscope images”, Proc. Micro 90, London 2-6 July 1990, Chapter 8, pp. 227-230.

[10] P.J. Shaw and D.J. Rawlins, “Three-Dimensional Fluorescence Microscopy”, Prog. Biophys. molec. Biol., 1991, vol. 56, pp. 187-213.

[11] C.J.R. Sheppard, Min Gu, Y. Kawata, and S. Kawata, “Three-dimensional transfer functions for high-aperture systems”, J. Opt. Soc. Am. A, vol. 11, No. 2, February 1994, pp. 593-598.

[12] Y. Vardi, L.A. Shepp, and L. Kaufman, “A Statistical Model for Positron Emission Tomography”, Journal of the American Statistical Association vol. 80, no 389, March 1985, pp. 8-37.

[13] P.J. Verveer, M.J. Gemkow, & T.M. Jovin, “A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy”, J. Microscopy, vol. 193, Pt 1, January 1999, pp. 50-61.

[14] T. Wilson, “The Role of the Pinhole in Confocal Imaging System”, Handbook of Biological Confocal Microscopy, 2nd ed., ed. J. B. Pawley, Plenum Press, NY and London, 1995, pp. 167-182.

Figures

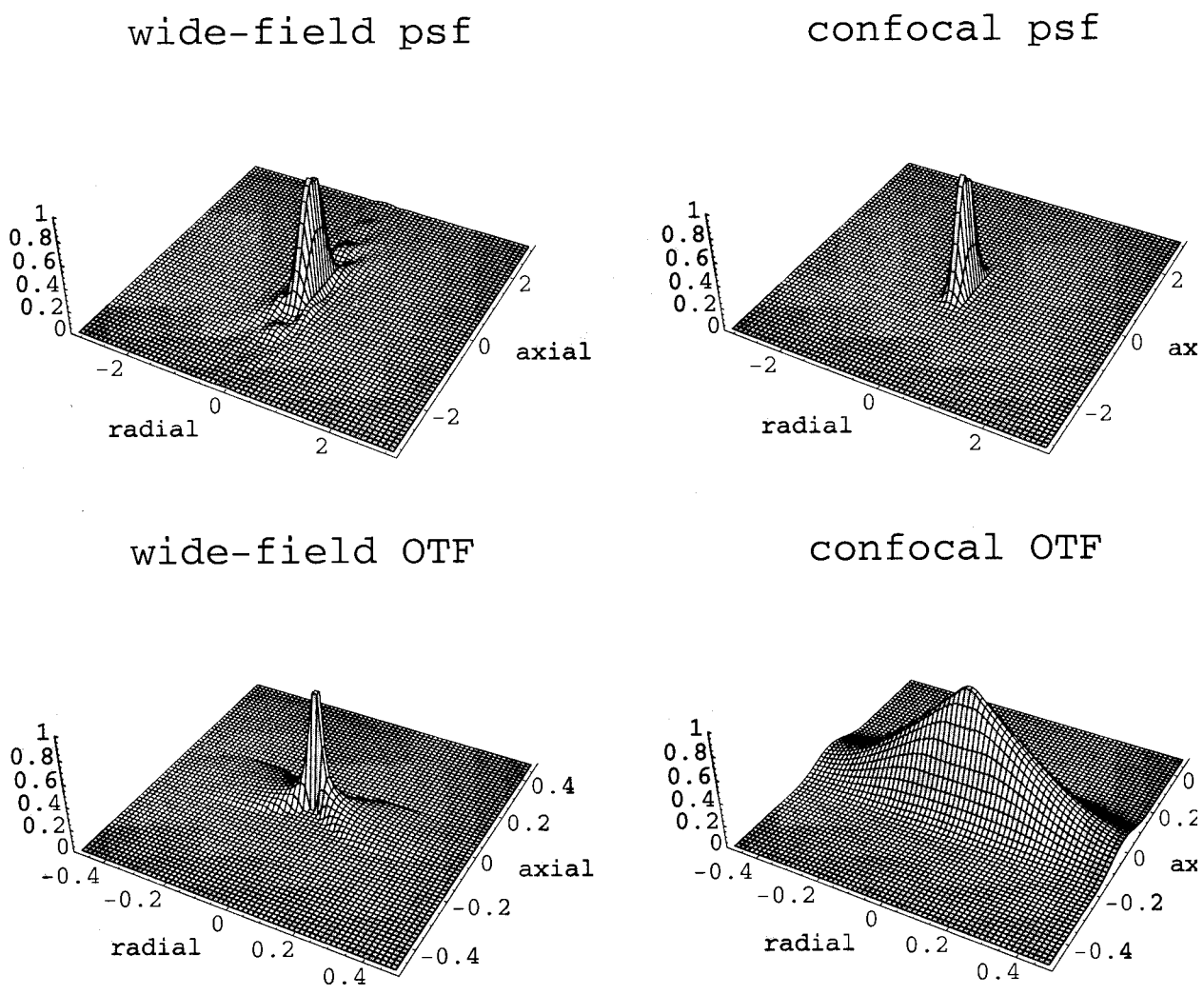

Figure 1. Wide-field and confocal psf and OTF. Psf axis scale is in μm . Mesh size = .1 μm . $\sin\alpha$ = 0,8. Numerical aperture = 1.3. Detector pinhole diameter = 25 μm . Objective magnification = 100.

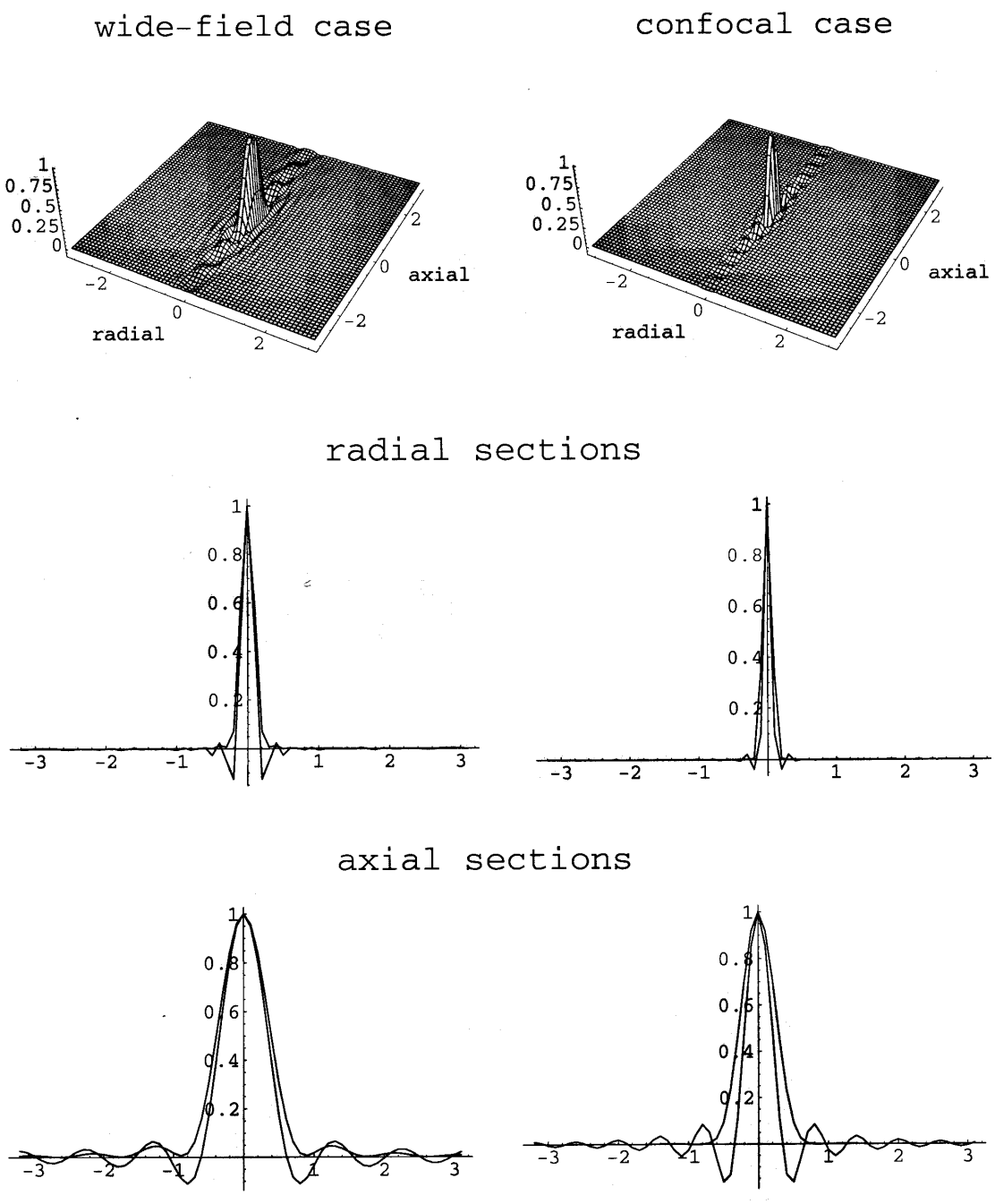

Figure 2. Characteristic function in wide-field and confocal case. Scale in μm . Mesh size = .1 μm . $\sin\alpha$ = 0,8. Numerical aperture = 1.3. Detector pinhole diameter 25 μm . Objective magnification = 100.

The sections show the characteristic function (narrower oscillating curve) and the psf (wider curve).

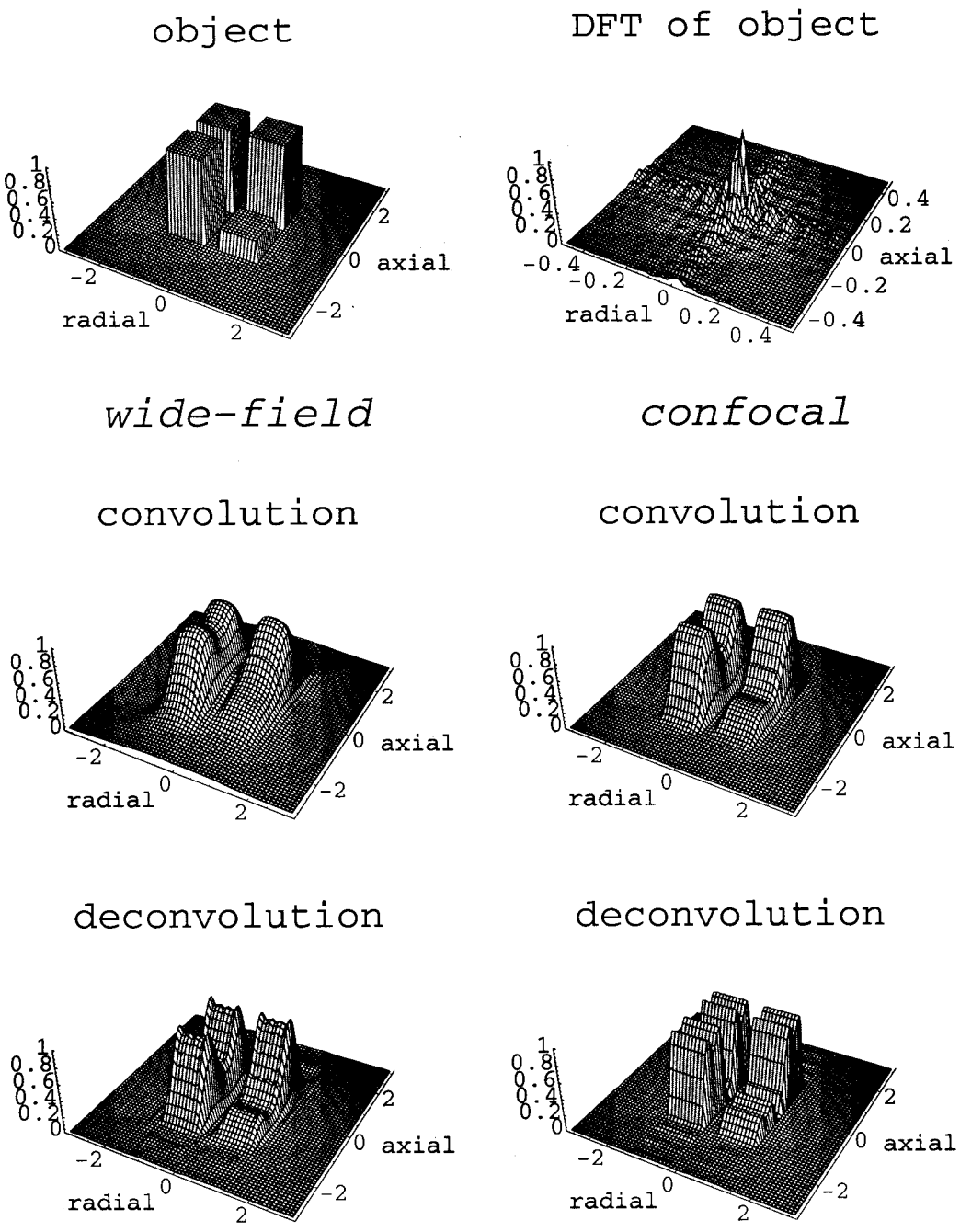

Figure 3. Top: Object and DFT of object. The object consists of four cubes, three with intensity = 1 and one with intensity = .25, cube side = 1 μm , separation = .4 μm . Below: Convolution and deconvolution of object in wide-field and confocal case.

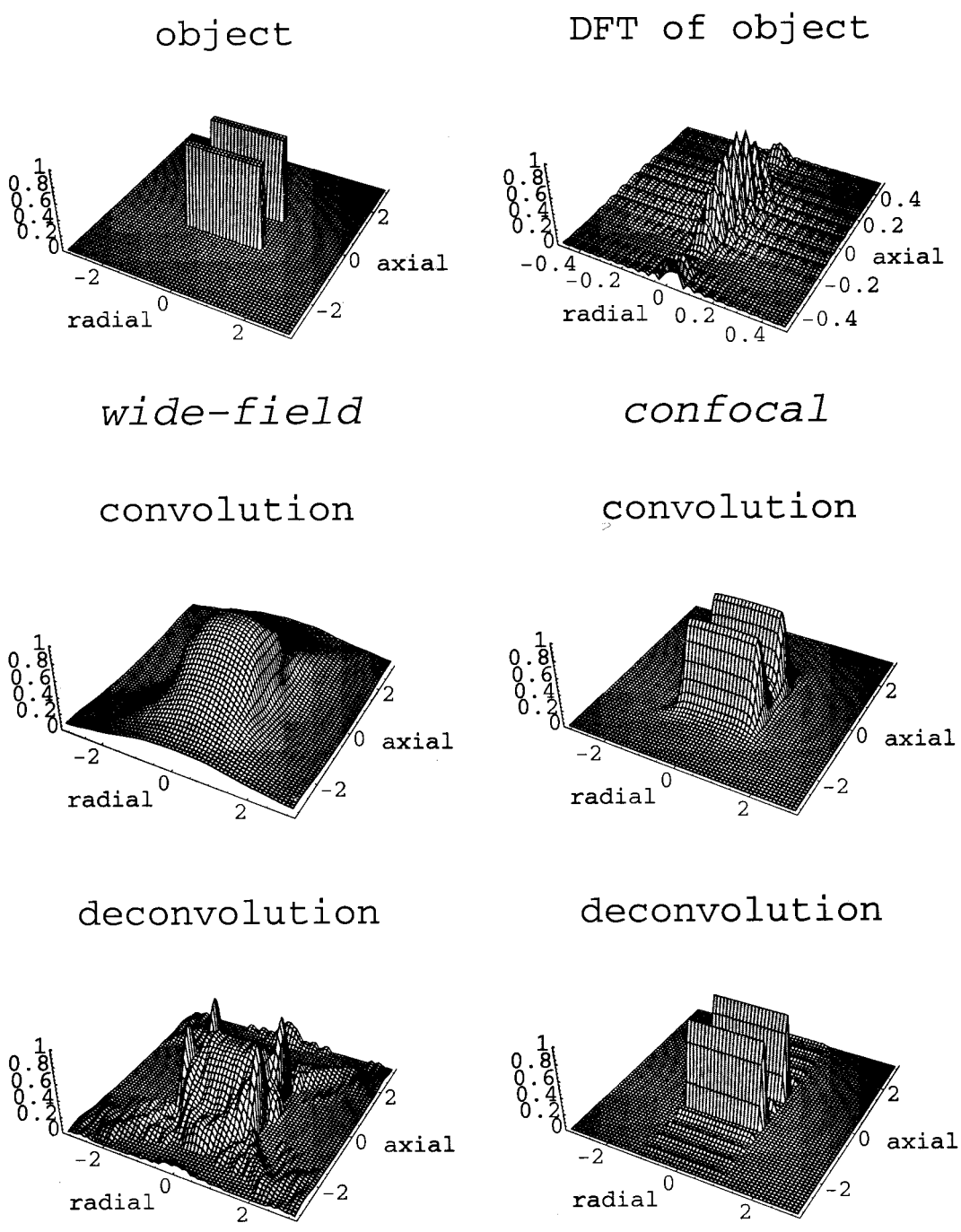

Figure 4. Top: Object and DFT of object. The object consists of two square slabs with intensity = 1, side = 2 μm , separation = 1 μm , thickness = .2 μm . Below: Convolution and deconvolution of object in wide-field and confocal case.

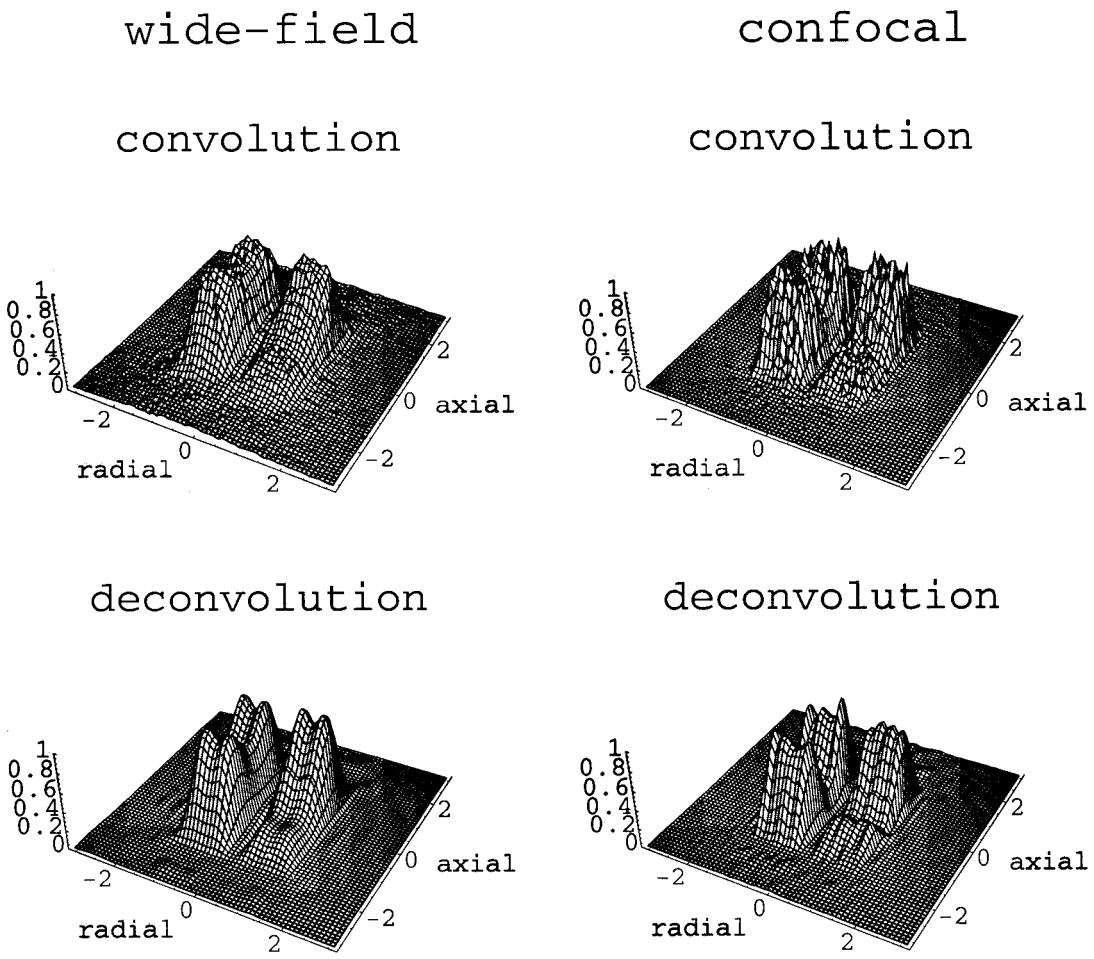

Figure 5. Noisy convolution and deconvolution of four cubes. The maximal intensity of the convolution corresponds to 1000 photons/voxel in the wide-field and to 100 photons/voxel in the confocal case. Wide-field cutlevel = .03. Confocal cutlevel = .1. These cutlevels result in the error amplifications 2.7 and 1.8, respectively.

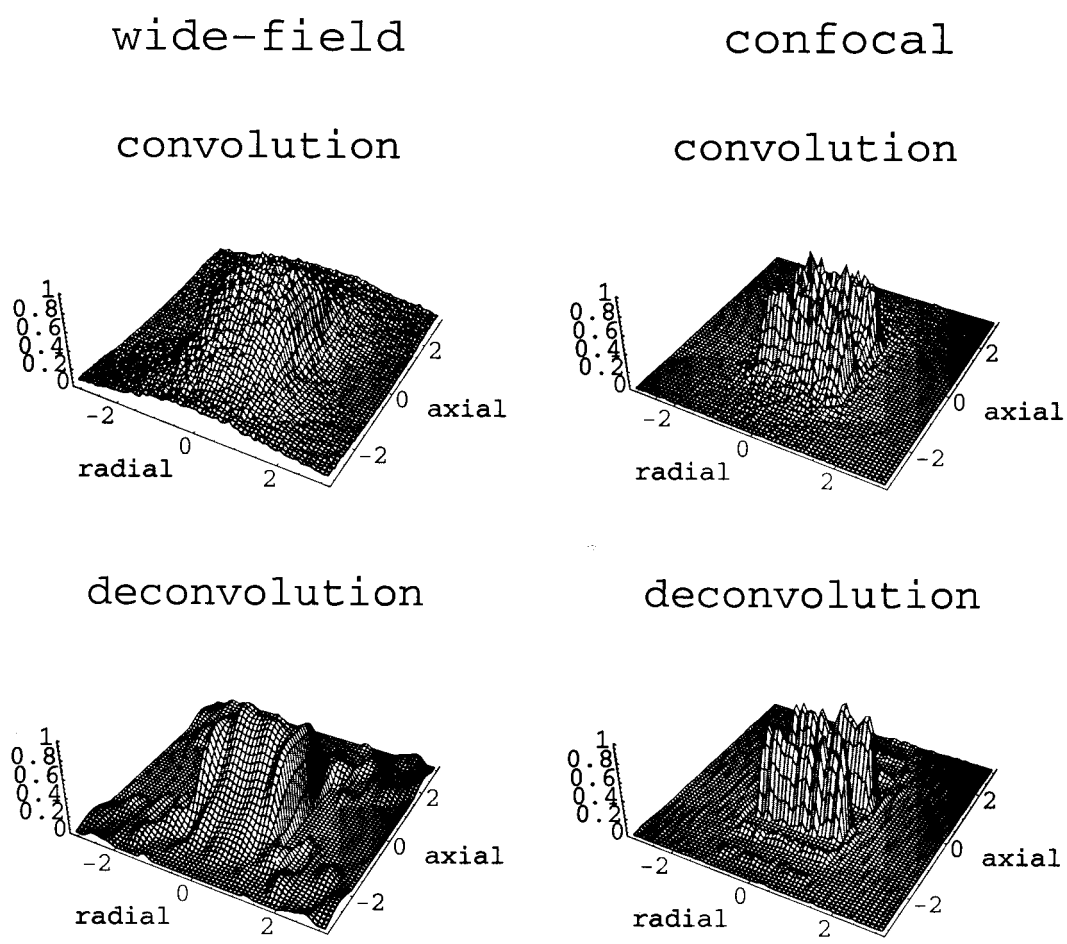

Figure 6. Noisy convolution and deconvolution of two slabs. The maximal intensity of the convolution corresponds to 1000 photons/voxel in the wide-field and to 100 photons/voxel in the confocal case. Wide-field cutlevel = .03. Confocal cutlevel = .1. These cutlevels result in the error amplifications 2.7 and 1.8, respectively

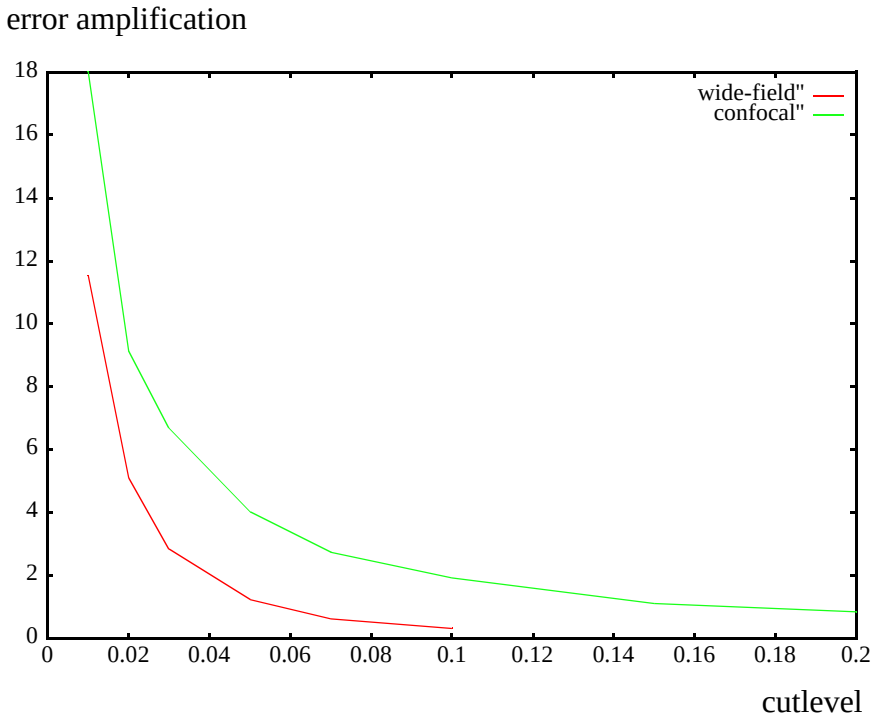

Figure 7. Error amplification of deconvolution process as function of *cutlevel* for *meshsize* = 0,1µm in the wide-field and confocal case. The error amplification is essentially independent of the number of voxels in the data cube, the noise level, and the shape of the object. It is roughly proportional to the *meshsize*.

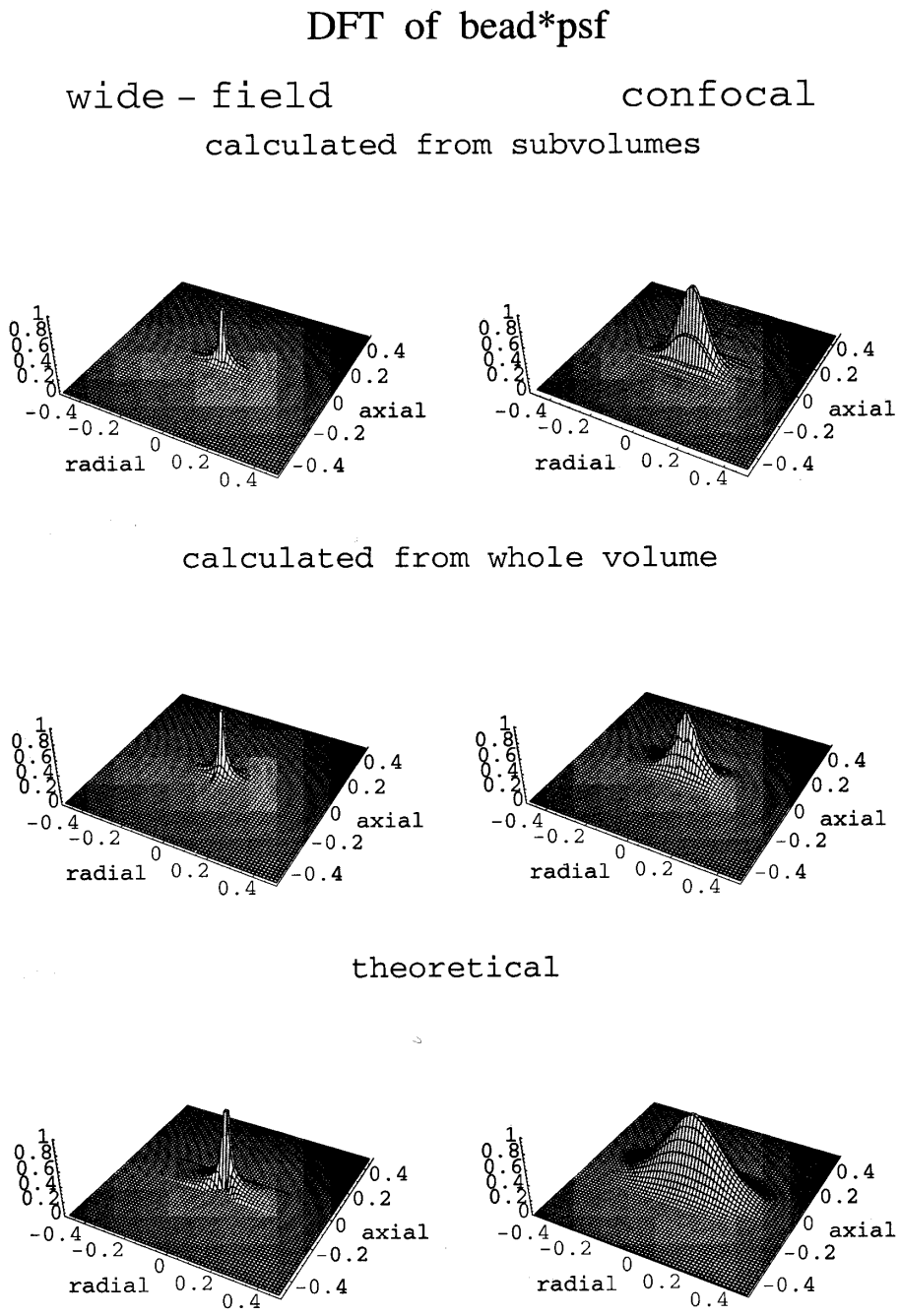

Figure 8. Estimates of DFT of bead*psf. Top row: averages of subvolumes each containing one bead. Middle: DFT of the whole volume. Bottom: Theoretically expected result obtained as DFT of (bead with radius = .3 μm convolved with psf).

There is a clear difference between theory (bottom) and practice (top and middle). The measured OTF is narrower, meaning that the psf is wider than the theory predicts.

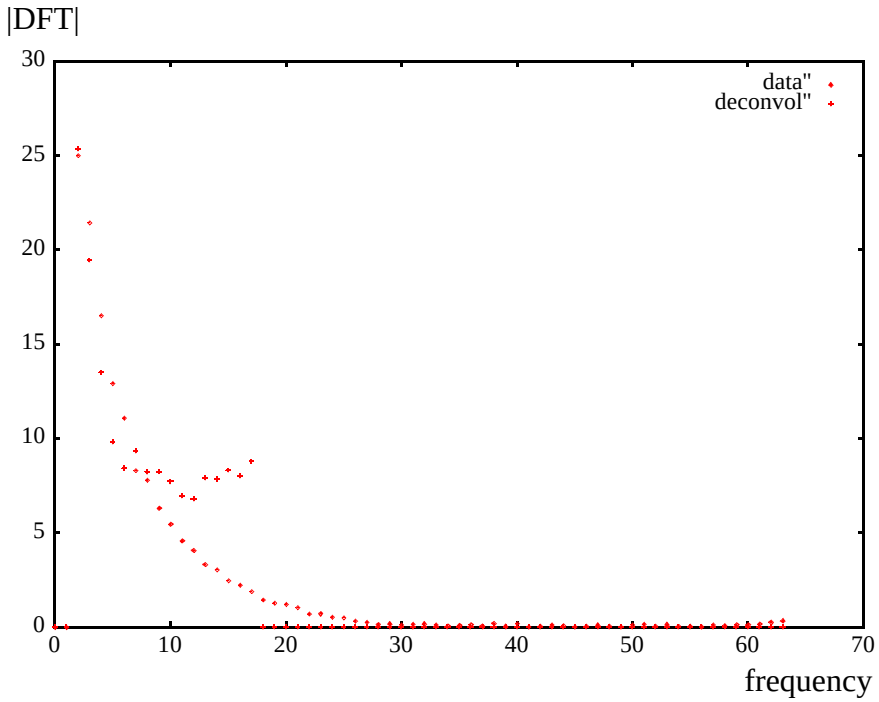

Figure 9. Moduli of DFT of data and of deconvolution as functions of K for $I = J = 0$.

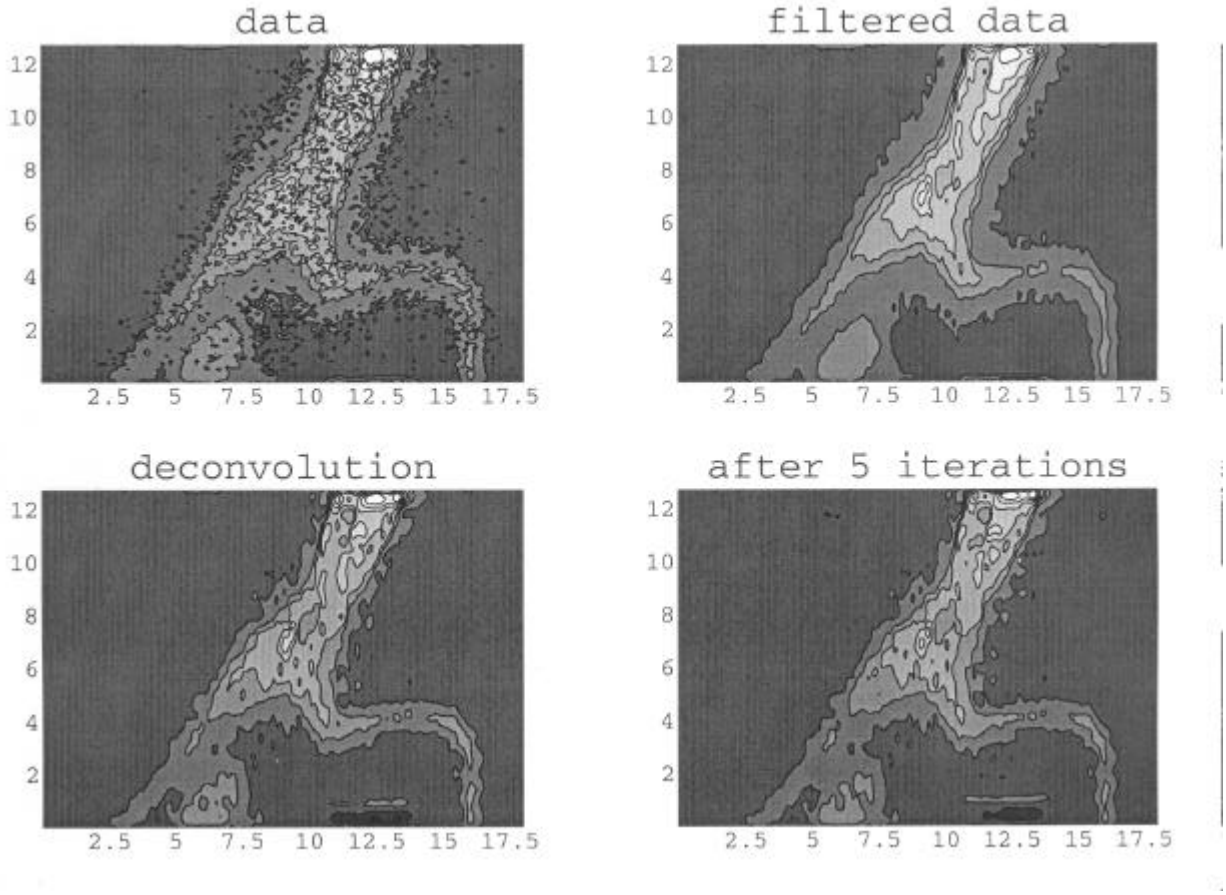

Figure 10. yz-sections of real data, filtered data, deconvolution with cut levels = .3 and .1, and after 5 iterations with the positivity algorithm. One gets the filtered data by Fourier-transforming the data, restricting it to the "used support", and then inverse transform. Usually, the filtered data is not computed, but one goes directly from the data to the deconvolution. Scale in μm . Mesh size = .1 μm . Vertical axis = z-axis = axial direction. Horisontal axis = y-axis = radial direction.

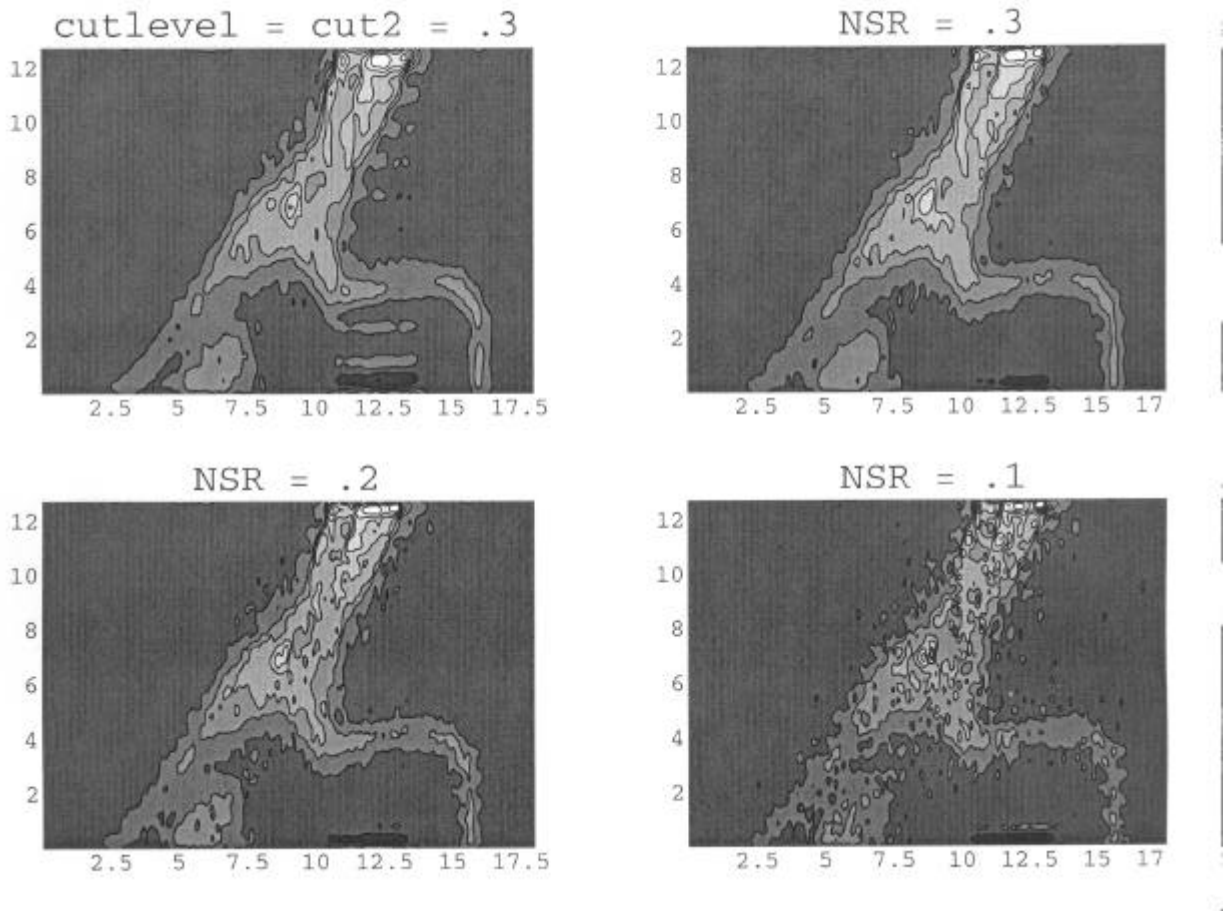

Figure 11. Deconvolution of the same data as in Figure 10 but with other parameters and by Wiener filtering. In the upper left corner, with cut level method using cutlevel = cut2 = .3 so that there is no smoothing at the boundary of the used support. The axial oscillations are clearly visible. The other three deconvolutions are by Wiener filtering using three different values of the Noise-to-Signal Ratio (NSR). Notice how sensitive the deconvolution is to the NSR.