

Danai Laksameethanasan and Sami S. Brandt. 2007. Generalised Skilling–Bryan minimisation for micro-rotation imaging in light microscopy. In: Kevin H. Knuth, Ariel Caticha, Julian L. Center Jr., Adom Giffin, and Carlos C. Rodríguez (editors). Proceedings of the 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2007). Saratoga Springs, New York, USA. 8-13 July 2007. AIP Conference Proceedings, volume 954, pages 354-361.

© 2007 American Institute of Physics

Reprinted with permission.

Generalised Skilling–Bryan Minimisation for Micro-Rotation Imaging in Light Microscopy

Danai Laksameethanasan and Sami S. Brandt

*Helsinki University of Technology, Laboratory of Computational Engineering,
P.O. Box 9203, FI-02015 TKK, Finland*

Abstract. The original Skilling–Bryan method, first introduced in the field of astronomy, minimises the statistical cost function assuming Gaussian noise and the entropy prior functional. In contrast to the conventional trust region methods, where the solution search in a high-dimensional space may be very expensive, the Skilling–Bryan scheme minimises the cost function in a subspace instead, yielding much more efficient computation. However, in low-photon-count processes, such as image formation in confocal microscopy, Poisson noise assumption is more suitable. In addition, it would be desirable to use the 2nd order, Skilling–Bryan optimisation framework in applications where general prior functionals with the positivity constraint is required. In this work, we hence generalise the Skilling–Bryan method to be used with the Poisson noise model as well as with various kinds of prior functions including the total variation prior. The proposed method is generic whereas here it is applied in the micro-rotation imaging, where we aim at 3D reconstruction of a rotating object from 2D projections taken by a light microscope. On the basis of our investigation, the extended Skilling–Bryan method is promising for computing estimates in high-dimensional problems efficiently.

Keywords: Maximum Entropy Method, Skilling–Bryan minimisation, Micro-Rotation imaging.

1. INTRODUCTION

A standard way for characterising a posterior distribution is computing the maximum a posteriori (MAP) estimate. Efficient numerical optimisation tools are essential for finding the MAP estimate, particularly in high-dimensional problems. Among numerous optimisation methods, line search and trust-region methods are conventional choices. The line search methods, in each iteration, search for a solution in a single direction whereas the trust-region methods compute the iterates in a high-dimensional space. In general, trust region methods tend to be more robust than line search methods, such as conjugate gradient [1]. However, the high-dimensional search may be expensive in the trust-region approach.

In contrast to the techniques above, Skilling and Bryan suggested to minimise a statistical cost function in a promising subspace, yielding more efficient computation [2, 3]. This approach is thus appropriate for solving nonlinear cost functions in high-dimensional spaces and, by their formulation, positivity of the solution is inherently enforced [2]. This method was first introduced in the field of astronomy and thereafter it has been widely used in other fields of science [4]. However, the original Skilling–Bryan method assumes the Gaussian noise model and entropy prior functional whereas in low-photon-count applications, such as image formation in confocal microscopy, Poisson noise assumption is more suitable. In addition, in many applications [5, 6], rather than the entropy prior, other priors may be needed.

Hence, we propose here a generalisation of the Skilling–Bryan method for minimising the statistical cost function with the Poisson noise model and general prior functions including the Gaussian and total variation (TV) priors. The method is generic and we apply it in solving a 3D reconstruction problem in the Micro-Rotation application in light microscopy [7, 8, 9]. Micro-Rotation imaging is an optical microscopic imaging technique which employs dielectric fields in rotating cells around a single axis parallel to the focal plane of the microscope [9, 10]. More precisely, the method involves manipulating dielectric fields to trap and control cells, aiming to achieve high resolution imaging of individual, intact live cells [10, 11]. Our ultimate goal is to obtain an efficient 3D reconstruction algorithm, based on Bayesian theory assuming Poisson noise with the general priors.

The paper begins by presenting the image formation model for micro-rotation imaging in Section 2 and introducing the statistical framework and the statistical cost functions in Section 3. The generalisation of the Skilling–Bryan minimisation of the cost functions is described in Section 4. The experimented results with simulated micro-rotation series are reported in Section 5. In Section 6, we discuss and conclude the paper.

2. IMAGE FORMATION MODEL

In this section, we describe the image formation model, used in the Micro-Rotation application. A typical microscope system can be characterised by its point spread function (PSF): if the microscope system is linear and shift invariant, the measurement image is

$$m_i(x, y) = h(x, y, z) * f_i(x, y, z) \Big|_{z=d}, \quad (1)$$

where h is the 3D PSF, $f_i = R_i f$ is the rotated object density for the projection i , R_i is the rotation operator and $*$ denotes the 3D convolution operator. The measurement image is thus recorded as the plane corresponding to the focal plane $z = d$ and the optical axis is assumed to be parallel to the z -direction.

In minimisation problem, the projection model (1) requires discretisation of the 3D space and the images onto finite grids. Since the imaging operator (1) is linear, the total discretised model for the M projections is described as

$$\mathbf{m} = \mathbf{A}\mathbf{f}, \quad (2)$$

where \mathbf{f} is a vector of object density values, $\mathbf{m} = [\mathbf{m}_1^T, \dots, \mathbf{m}_M^T]^T$ and $\mathbf{A} = [\mathbf{A}_1^T, \dots, \mathbf{A}_M^T]^T$ are the joint measurement vector and the joint projection matrix, respectively, combining all the projections. The implementation details of the matrix \mathbf{A} and its adjoint \mathbf{A}^T can be found from [11, 12].

3. BAYESIAN INVERSION THEORY

By the Bayesian inversion approach, prior knowledge can be used in a systematic way. The solution is described via the posterior distribution which, in our application, takes

the form

$$p(\mathbf{f}|\mathbf{m}) \propto p(\mathbf{f})p(\mathbf{m}|\mathbf{f}), \quad (3)$$

where $p(\mathbf{m}|\mathbf{f})$ is the likelihood density and $p(\mathbf{f})$ is the prior density. The computation of the MAP estimate

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} p(\mathbf{f}|\mathbf{m}) \quad (4)$$

corresponds to maximising the cost function

$$q(\mathbf{f}) = s(\mathbf{f}) - \lambda c(\mathbf{f}), \quad (5)$$

where $c(\mathbf{f}) = -\ln p(\mathbf{m}|\mathbf{f})$ is the likelihood term and $s(\mathbf{f}) = \ln p(\mathbf{f})$ is the prior term. The auxiliary regularisation parameter λ balances the likelihood and the prior terms. Over regularisation (too large λ) leads the algorithm to diverge where as under-regularisation amplifies noise. The likelihood density is constructed from the measurement noise model, as follows.

In many applications, such as those of fluorescence microscopy, the amount of photons reaching the detector is well described by the Poisson process, i.e., the pixel measurement m_j is a sample of a Poisson distributed random variable with the expected value $(\mathbf{A}\mathbf{f})_j$. Assuming independent measurements, the cost is

$$c_P(\mathbf{f}) = \mathbf{1}^T(\mathbf{A}\mathbf{f}) - \mathbf{m}^T \log(\mathbf{A}\mathbf{f}), \quad (6)$$

where $\mathbf{1} = [1, 1, \dots, 1]^T$. If the amount of photons arrived at the detector is high, the Poisson noise is well approximated by Gaussian noise [13]. In the case of isotropic Gaussian noise with unit variance, the cost function is

$$c_G(\mathbf{f}) = \frac{1}{2} \|\mathbf{m} - \mathbf{A}\mathbf{f}\|^2. \quad (7)$$

To control the fitting, it is reasonable to set the bound c_{aim} to which we adjust the likelihood c [2]. Typically, the constant c_{aim} is proportional to the noise level in the images. For the Gaussian noise model, c_{aim} is equal to $N\sigma^2$ where σ^2 is the noise variance and N is the number of measurement data. The Poisson case is more difficult, one possibility is to select the Csiszár I-divergence [14] as c_{aim} .

In this work, we have experimented Gaussian, Entropy, and the TV prior models. The Gaussian, white noise prior, with the cost

$$s_G(\mathbf{f}) = -\frac{1}{2} \|\mathbf{f}\|^2, \quad (8)$$

is a simple choice and is also known as Tikhonov regularisation. The entropy prior is obtained by setting

$$s_E(\mathbf{f}) = \mathbf{1}^T \mathbf{f} - \mathbf{f}^T \log\left(\frac{\mathbf{f}}{\mathbf{f}_0}\right), \quad (9)$$

TABLE 1. Likelihood and prior terms with their gradients and Hessians used in the extended Skilling–Bryan minimisation method. The operations between two vectors are performed element by element.

| | Functions | Gradient \mathbf{g} | Hessian \mathbf{H} |
|----------------------|--|--|--|
| Gaussian noise c_G | $\frac{1}{2}\ \mathbf{m} - \mathbf{A}\mathbf{f}\ ^2$ | $-\mathbf{A}^T(\mathbf{m} - \mathbf{A}\mathbf{f})$ | $\mathbf{A}^T\mathbf{A}$ |
| Poisson noise c_P | $\mathbf{1}^T(\mathbf{A}\mathbf{f}) - \mathbf{m}^T \log(\mathbf{A}\mathbf{f})$ | $-\mathbf{A}^T\left(\frac{\mathbf{m}}{\mathbf{A}\mathbf{f}} - \mathbf{1}\right)$ | $\mathbf{A}^T \text{diag}\left(\frac{\mathbf{m}}{(\mathbf{A}\mathbf{f})(\mathbf{A}\mathbf{f})}\right)\mathbf{A}$ |
| Gaussian prior s_G | $-\frac{1}{2}\ \mathbf{f}\ ^2$ | $-\mathbf{f}$ | $-\mathbf{I}$ |
| Entropy prior s_E | $\mathbf{1}^T\mathbf{f} - \mathbf{f}^T \log(\mathbf{f}/\mathbf{f}_0)$ | $-\log(\mathbf{f}/\mathbf{f}_0)$ | $-\text{diag}(\mathbf{1}/\mathbf{f})$ |
| TV prior s_T | $-\mathbf{1}^T\beta^{-1}\log(\cosh(\beta\mathbf{G}\mathbf{f}))$ | $-\mathbf{G}^T \tanh(\beta\mathbf{G}\mathbf{f})$ | $-\mathbf{G}^T \text{diag}(\beta \text{sech}^2(\beta\mathbf{G}\mathbf{f}))\mathbf{G}$ |

where \mathbf{f}_0 is the initial object, for which the ideal choice would be the true object. As the true object is unknown, \mathbf{f}_0 is often selected to have a constant value.

For the total variation prior, we use the definition

$$s_T(\mathbf{f}) = -\mathbf{1}^T|\mathbf{G}\mathbf{f}|. \quad (10)$$

where \mathbf{G} is the Toeplitz matrix representing the 3D convolution with the Laplacian of Gaussian kernel. Since the absolute function is not differentiable, we use the smooth approximation $|t| \approx \beta^{-1}\cosh(\beta t)$, as suggested in [15], where β is a parameter. The advantage of the TV is that it preserves the edges of the object while it smooths out homogeneous regions [16, 6].

In minimising the cost, we need to compute the gradient \mathbf{g} and Hessian \mathbf{H} of the likelihood and the prior terms. Table 1 summaries all the functions with their gradients and Hessians, which will be used in the following section.

4. SKILLING–BRYAN MINIMISATION

In this section, we generalise the Skilling–Bryan algorithm to be applied with the Poisson noise model with the three priors, described in the previous section. The method aims at maximising the prior term s subject to the constraint $c = c_{\text{aim}}$, by implicitly adjusting λ so that the constraint is satisfied. More precisely, in each iteration, s and c are projected onto a small dimensional subspace where s attains its maximum while the constraint is satisfied.

The cost function (5) can be approximated with the second-order Taylor expansion. Then the vector \mathbf{p} is selected so that the posterior cost function q is maximised, i.e.,

$$\max_{\mathbf{p}} q(\mathbf{f} + \mathbf{p}) = q(\mathbf{f}) + \mathbf{g}_q^T\mathbf{p} + \frac{1}{2}\mathbf{p}^T\mathbf{H}_q\mathbf{p}, \quad (11)$$

where $\mathbf{g}_q = \nabla q$ and $\mathbf{H}_q = \nabla\nabla q$. This implies that $\mathbf{g}_q = \mathbf{g}_s - \lambda\mathbf{g}_c$ and $\mathbf{H}_q = \mathbf{H}_s - \lambda\mathbf{H}_c$ where $\mathbf{g}_c = \nabla c$, $\mathbf{g}_s = \nabla s$, $\mathbf{H}_c = \nabla\nabla c$ and $\mathbf{H}_s = \nabla\nabla s$. Searching for the vector \mathbf{p} in a high dimensional space is costly. To obtain an efficient computation, Skilling and Bryan suggested maximising the cost function in the trust subspace, as follows.

The solution for (11) is

$$\mathbf{p} = -(\mathbf{H}_q + \gamma \mathbf{I})^{-1} \mathbf{g}_q, \quad (12)$$

$$\approx -(\mathbf{I} + \gamma^{-1} \mathbf{H}_q) \mathbf{g}_q, \quad (13)$$

$$\approx -\mathbf{g}_s + \lambda \mathbf{g}_c + \gamma^{-1} [(\mathbf{H}_s - \lambda \mathbf{H}_c)(\mathbf{g}_s - \lambda \mathbf{g}_c)] \quad (14)$$

where the inversion is approximated upto the second-order. Obviously, the search vector \mathbf{p} is just a linear combination of the six vectors, $\mathbf{g}_c, \mathbf{g}_s, \mathbf{H}_c \mathbf{g}_c, \mathbf{H}_s \mathbf{g}_s, \mathbf{H}_s \mathbf{g}_c$ and $\mathbf{H}_c \mathbf{g}_s$ in the approximation. In practice, the linear combination of the last four terms can be used as one search direction. Thus, we obtain the three search directions

$$\mathbf{e}_1 = \mathbf{f} \mathbf{g}_s, \quad (15)$$

$$\mathbf{e}_2 = \mathbf{f} \mathbf{g}_c, \quad (16)$$

$$\mathbf{e}_3 = \mathbf{f} \mathbf{H}_c \left(\frac{\mathbf{e}_1}{\|\mathbf{g}_s\|} - \frac{\mathbf{e}_2}{\|\mathbf{g}_c\|} \right) + \mathbf{f} \mathbf{H}_s \left(\frac{\mathbf{e}_1}{\|\mathbf{g}_s\|} - \frac{\mathbf{e}_2}{\|\mathbf{g}_c\|} \right), \quad (17)$$

which define a 3D subspace. The vector multiplications and divisions are performed element by element.

The gradient directions \mathbf{g}_s and \mathbf{g}_c are replaced by $\mathbf{f} \mathbf{g}_s$ and $\mathbf{f} \mathbf{g}_c$ to increase the weight for high values in order to achieve the positivity constraint. In the third direction, the normalisation by the length of gradient vector is performed before multiplying with the Hessian matrix. In the case of the entropy prior, the second term in (17) can be dropped since $\mathbf{f} \mathbf{H}_s$ is then equal to unity matrix. Note that solving the search vector \mathbf{p} from (12) corresponds to the Levenberg-Marquardt method whereas using the first-order Taylor approximation would coincide with optimisation with line search methods.

Within the 3D subspace, the search vector \mathbf{p} is

$$\mathbf{p} = \mathbf{E} \mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + x_3 \mathbf{e}_3, \quad (18)$$

where the matrix $\mathbf{E} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3]$ and the coefficient vector $\mathbf{x} = [x_1, x_2, x_3]^T$. Now, we determine \mathbf{x} in this subspace that gives the maximum of s subject to the constraint $c = \tilde{c}_{\text{aim}} < c_{\text{aim}}$. In order to estimate the maximum s in the 3D subspace, the three-element gradients and nine-element Hessian need to be computed. For details, see [2]. Finally, the current \mathbf{f} is moved to the new location by

$$\mathbf{f}_{\text{new}} = \mathbf{f} + \mathbf{E} \mathbf{x}, \quad (19)$$

while the updated \mathbf{f} needs to be protected against stray on negative values; if \mathbf{f}_{new} becomes negative, we scale the updated vector $\mathbf{E} \mathbf{x}$ down by the factors of two until it becomes positive. The iteration is repeated until the aim $c = c_{\text{aim}}$ is achieved or the maximum number of iterations is reached. Compared to line search methods, the optimisation scheme requires much less number of iterations but each iteration consists of six projection and back-projection ($\mathbf{A}^T \mathbf{A}$) pairs. So constructing the more sophisticated search directions is made on the expense of more frequent evaluation of the image formation model.

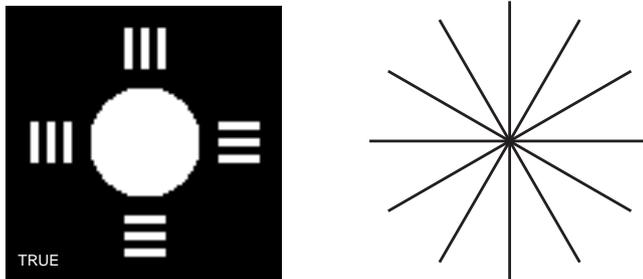


FIGURE 1. Left: The original 2D synthetic object. Right: Ideal geometries of the data acquisition of Micro-Rotation imaging where each line represents the focal plane in the object coordinate frame.

5. RESULTS

To evaluate the proposed method, we have experimented simulated wide-field microscope images. In the simulations, we created a synthetic 2D object in a volume of 100×100 samples. The object consists of a filled circle in the middle and three thin bars, arranged around the sphere in different directions, as illustrated in Figure 1. The object has only two intensity values, one for the background and the other for the object. Poisson noise was added to the object and it was then projected onto 50 one-dimensional views using the simulated micro-rotation imaging model.

In Figure 2, there are six reconstructions computed with different selections from the two noise models and the three prior functions. As the number of simulated photons was relatively large, the reconstructions with Poisson and Gaussian noise models have only a small difference when the same prior function is selected. Likewise, there is only a small difference between the Gaussian and entropy priors but the TV prior is the most competitive in smoothing the artifacts inside the homogeneous region inside the circle while the edges of the circle remain sharp. It can be additionally seen that the three bars oriented in the radial direction can not be reconstructed. We see that this is due to the rotation geometry and the elongation of the PSF function of the microscope which make the resolution weaker in the tangential direction of the rotation [11].

We also experimented with a 3D simulated cell. Figure 3 (left) shows some example images, obtained by the simulated image formation model with Poisson noise. The $100 \times 100 \times 100$ reconstruction using Poisson model and TV prior is illustrated in Figure 3 (right) which shows the rotated slice of the reconstruction corresponding to the original images. In the simulation, the algorithm converged in about 100 iterations for both Gaussian and Poisson noise models. In the early iterations, the algorithm typically has slow convergence but it becomes faster nearer the optimum. In addition, we found only slight differences in convergence curves between different prior models. However, the convergence of the algorithm highly depends on sharpness of the PSF; the narrower the PSF width, the faster the convergence seems to be.

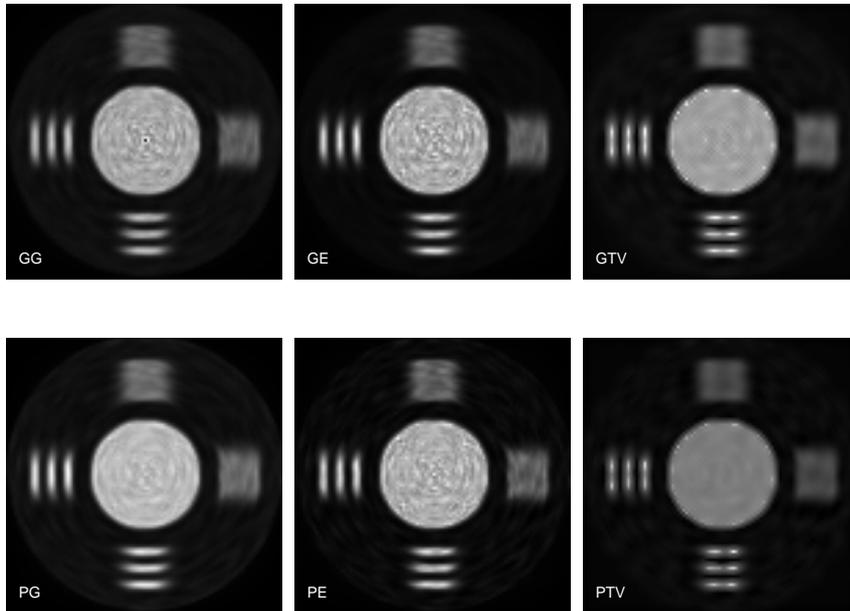


FIGURE 2. 2D reconstructions of simulated data using different priors and noise models. Upper row: Gaussian noise; lower row: Poisson noise; columns from left to right: Gaussian, Entropy, and Total Variation priors.

6. DISCUSSION

In this work, we have generalised the original Skilling–Bryan method to be applied with the Poisson noise model as well as the various kinds of prior structures, including the Gaussian, the Entropy and the TV priors. In general, our extended algorithm should be able to find the optimum solution for arbitrary convex, nonlinear cost functions with the positivity constraint. Only the gradients and Hessians of the cost functions is required for this method. The main features of the Skilling–Bryan method is that (1) a subspace of several search directions is used instead of a line search, (2) the algorithm directly controls the noise level constraint over the likelihood function so that the selection of regularisation parameters becomes implicit, and (3) the method inherently enforces the positivity constraint during each subspace search. These features make our extended algorithm attractive for solving high-dimensional, nonlinear problems. In addition, our experiments in the Micro-Rotation setting suggested that the extended algorithm is appropriate for solving the wide class of measurement noise and prior models within the Bayesian framework.

ACKNOWLEDGMENTS

This work is supported by the European Commission (NEST 2005 programme) in consortium AUTOMATION (Coordinator S. L. Shorte, Institut Pasteur, Paris). We additionally thank Scientific Volume Imaging, AUTOMATION project partner, for providing the 3D cell phantom and the PSF for testing.

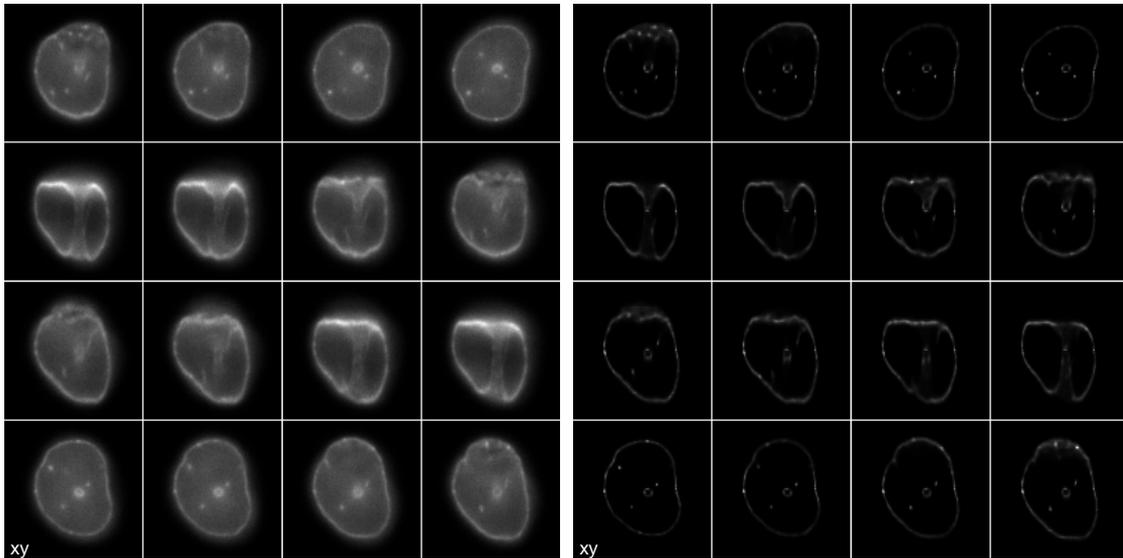


FIGURE 3. Left: Simulated, original wide-field micro-rotation images (SNR=20). Right: Deblurred images, i.e. the ideal slices of the reconstruction from the positions corresponding to the focal plane of the original images.

REFERENCES

1. J. Nocedal, and S. Wright, *Numerical Optimization*, Springer Verlag, 2006, 2nd edn.
2. J. Skilling, and R. K. Bryan, *Mon. Notices R. Astron. Soc.* **211**, 111–124 (1984).
3. J. Skilling, “Classic Maximum Entropy,” in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, Kluwer Academic, 1989.
4. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1992.
5. P. J. Verveer, M. J. Gemkow, and T. M. Jovin, *Journal of Microscopy* **193**, 50–61 (1999).
6. N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, P. Roux, J. Olivo-Marin, and J. Zerubia, *Microscopy Research Technique* **69**, 260–266 (2006).
7. T. Schnelle, R. Hagedorn, G. Fuhr, S. Fiedler, and T. Müller, *Biochim. Biophys. Acta.* **1157**, 127–140 (1993).
8. G. Fuhr, T. Müller, T. Schnelle, R. Hagedorn, A. Voigt, S. Fiedler, W. M. Arnold, U. Zimmermann, B. Wagner, and A. Heuberger, *Naturwissenschaften* **81**, 528–535 (1994).
9. R. Lizundia, L. Sengmanivong, J. Guergnon, T. Müller, T. Schnelle, G. Langsley, and S. L. Shorte, *Parasitology* **130**, 1–7 (2005).
10. O. Renaud, R. Heintzmann, A. Saez-Cirion, T. Schnelle, T. Müller, and S. L. Shorte, *Current Opinion in Microbiology* (2007).
11. D. Laksameethanasan, S. S. Brandt, P. Engelhardt, O. Renaud, and S. L. Shorte (2007), submitted for publication.
12. D. Laksameethanasan, S. S. Brandt, and P. Engelhardt, “A Three-Dimensional Bayesian Reconstruction Method with the Point Spread Function for Micro-Rotation Sequences in Wide-Field Microscopy,” in *Proc. IEEE International Symposium on Biomedical Imaging*, Arlington, VA, USA, 2006, pp. 1276–1279.
13. J. Kaipio, and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer Verlag, 2004.
14. I. Csizsár, *Ann. Stat.* **19**, 2032–2066 (1991).
15. V. Kolehmainen, S. Siltanen, S. Järvenpää, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo, *Phy. Med. Bio.* **48**, 1465–1490 (2003).
16. L. Rudin, S. Osher, and E. Fatemi, *Physica D* **60**, 259–268 (1992).