

Application of the split-gradient method to 3D image deconvolution in fluorescence microscopy

G. VICIDOMINI^{*,†,‡,§}, P. BOCCACCI^{†,‡}, A. DIASPRO^{‡,§}
& M. BERTERO[†]

^{*}Department of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry,
Am Fassberg 11, Goettingen, Germany

[†]DISI, University of Genoa, Via Dodecaneso 35, Genoa, Italy

[‡]LAMBS-MicroSCoBiO, University of Genoa, Via Dodecaneso 33, Genoa, Italy

[§]IFOM, Foundation FIRC Institute for Molecular Oncology, Via Adamello 16, Milan, Italy

Key words. Deconvolution, edge-preserving, fluorescence microscopy, Markov random field, split-gradient method.

Summary

The methods of image deconvolution are important for improving the quality of the detected images in the different modalities of fluorescence microscopy such as wide-field, confocal, two-photon excitation and 4Pi. Because deconvolution is an ill-posed problem, it is, in general, reformulated in a statistical framework such as maximum likelihood or Bayes and reduced to the minimization of a suitable functional, more precisely, to a constrained minimization, because non-negativity of the solution is an important requirement. Next, iterative methods are designed for approximating such a solution.

In this paper, we consider the Bayesian approach based on the assumption that the noise is dominated by photon counting, so the likelihood is of the Poisson-type, and that the prior is edge-preserving, as derived from a simple Markov random field model. By considering the negative logarithm of the *a posteriori* probability distribution, the computation of the *maximum a posteriori* (MAP) estimate is reduced to the constrained minimization of a functional that is the sum of the Csiszár I-divergence and a regularization term. For the solution of this problem, we propose an iterative algorithm derived from a general approach known as *split-gradient method* (SGM) and based on a suitable decomposition of the gradient of the functional into a negative and positive part. The result is a simple modification of the standard Richardson–Lucy algorithm, very easily implementable and assuring automatically the non-negativity of the iterates. Next, we apply this method to the particular case of confocal microscopy for investigating the effect of several edge-preserving priors proposed in the literature using both synthetic and real

confocal images. The quality of the restoration is estimated both by computation of the Kullback–Leibler divergence of the restored image from the detected one and by visual inspection. It is observed that the noise artefacts are considerably reduced and desired characteristics (edges and minute features as islets) are retained in the restored images. The algorithm is stable, robust and tolerant at various noise (Poisson) levels. Finally, by remarking that the proposed method is essentially a scaled gradient method, a possible modification of the algorithm is briefly discussed in view of obtaining fast convergence and reduction in computational time.

Introduction

Several approaches and algorithms have been proposed for image deconvolution, an important topic in imaging science. The purpose is to improve the quality of images degraded by blurring and noise. The ill-posedness of the problem is the basic reason of the profusion of methods: many different formulations are introduced, based on different statistical models of the noise affecting the data, and many different priors are considered for regularizing the problem, for instance, in a Bayesian approach. As a result, deconvolution is usually formulated in one of several possible variational forms, and for a given formulation, several different minimization algorithms, in general iterative, are proposed.

For the specific application to microscopy, we mention a few methods, without pretending to be exhaustive. Under the assumption of a Gaussian white noise, the approach based on the Tikhonov regularization theory (Tikhonov & Arsenin, 1977; Engl *et al.*, 1996) and the methods proposed for computing non-negative minimizers of the corresponding functional, such as the method of Carrington (1990) or the iterative method proposed in van der Voort & Strasters (1995),

are worth mentioning. On the other hand, in the case of Poisson statistics (noise dominated by photon counting), the classical Richardson–Lucy (RL) algorithm (Richardson, 1972; Lucy, 1974), derived from a maximum likelihood approach (Shepp & Vardi, 1982), can be routinely used. A quantitative comparison of these methods is given in van Kempen *et al.* (1997) and van Kempen & van Vliet (2000b). They were applied mainly to wide-field and confocal microscopy, but the cases of the 4Pi and two-photon excitation microscopy were also investigated (Schraeder *et al.*, 1998; Mondal *et al.*, 2008). We also remark that in some applications, the images are affected by both Poisson and Gaussian read-out noise. For treating this case, a more refined model was proposed by Snyder *et al.* (1993), with a related expectation maximization (EM) algorithm. However, as confirmed by a recent analysis (Benvenuto *et al.*, 2008), this model improvement does not produce a significant improvement in the image restoration, so the assumption of Poisson statistics is, in general, satisfactory.

As is known, early stopping of the RL method (RLM) provides a regularization effect (Bertero & Boccacci, 1998), even if in some cases, the restoration is not satisfactory. To this purpose, a regularization of the Poisson likelihood by means of the Tikhonov functional is proposed in Conchello & McNally (1996). This approach is known as the RL–Conchello algorithm (van Kempen & van Vliet, 2000a). In a more recent paper, Dey *et al.* (2006) investigate the regularization based on the total variation (TV) functional, introduced by Rudin *et al.* (1992) for image denoising. They also propose an iterative algorithm derived from the one-step-late (OSL) method of Green (1990). OSL is introduced as a modified EM algorithm, but its structure is that of a scaled gradient method, with a scaling that is not automatically positive. For this reason, convergence can be proved only in the case of a sufficiently small value of the regularization parameter (Lange, 1990).

In this paper, we consider the regularization of the Poisson likelihood in the framework of a Bayesian approach. In general, regularization can be obtained by imposing *a priori* information about properties of the object to be restored, and the Bayesian approach enables integration of this available prior knowledge with the likelihood using Bayes' law; then, the estimate of the desired object is obtained by maximizing the resulting posterior function. This approach is called the *maximum a posteriori* (MAP) method and can be reduced to the solution of a minimization problem by taking the negative logarithm of the posterior function.

In general, the prior information can be introduced by regarding the object as a realization of a Markov random field (MRF) (Geman & Geman, 1984); then, the probability distribution of the object is obtained using the equivalence of MRF and Gibbs random fields (GRF) (Besag, 1974). In particular, the potential function of the Gibbs distribution can be appropriately chosen to bring out desired statistical properties of the object. The combination of MRF and MAP estimation is extensively studied in single

photon emission computed tomography (SPECT) and positron emission tomography (PET) image reconstruction (Geman & McClure, 1985; Green, 1990). Recently, it was demonstrated that such a combination provides a powerful framework also for three-dimensional (3D) image restoration in fluorescence microscopy (Vicidomini *et al.*, 2006; Mondal *et al.*, 2007).

A simple and well-known regularization is based on the assumption that objects are made of smooth regions, separated by sharp edges. This is called edge-preserving regularization and requires non-quadratic potential functions. Therefore, in this paper, we consider the regularization of the negative logarithm of the Poisson likelihood by means of different edge-preserving potential functions proposed by different authors (see, for instance, Geman & Geman, 1984; Charbonnier *et al.*, 1997).

In view of the application to fluorescence microscopy, we must consider the minimization of this functional on the convex set of the non-negative images (the non-negative orthant) and, in order to overcome the difficulties of the OSL algorithm, we investigate the applicability of the split-gradient method (SGM) to this problem. SGM, proposed by Lantéri *et al.* (2001, 2002), is a general approach that allows designing of iterative algorithms for the constrained minimization of regularized functionals both in the case of Gaussian and in the case of Poisson noise as well as in the case in which both noises are present (Lantéri & Theys, 2005). The general structure is that of a scaled gradient method, with a scaling that is always strictly positive. Therefore, from this point of view, SGM is superior to the OSL method.

Finally, we point out that, thanks to SGM, the edge-preserving regularizations investigated in this paper can also be easily applied to the case of the least-square problem (i.e. additive Gaussian noise); hence, we provide an approach to image deconvolution both for microscopy techniques in which Poisson noise is dominant, such as confocal microscopy, and for techniques in which a Gaussian noise can be more appropriate, such as wide-field microscopy.

The paper is organized as follows. In 'The edge-preserving approach', we briefly recall the maximum likelihood approach in the case of Poisson noise, its reduction to the minimization of Csiszár I-divergence (also called Kullback–Leibler divergence) and the RL algorithm; moreover, we introduce the edge-preserving potentials considered in this paper. In 'Split-gradient method (SGM)', we give SGM in the simple case of step length 1. This form can be easily obtained (Bertero *et al.*, 2008) as an application of the method of successive approximations to a fixed-point equation derived from the Karush–Kuhn–Tucker (KKT) conditions for the constrained minimizers of the functional and based on a suitable splitting of the gradient into a positive and negative part. The convergence of this simplified version of SGM is not proved, even if it has always been verified in numerical experiments. However, convergence can be obtained by a suitable search of the step length (Lantéri *et al.*, 2002) or by applying a recently proposed

scaled gradient projection (SGP) method (Bonettini *et al.*, 2007). The relationship between the simplified version of SGM and OSL is also shown. Moreover, we determine the splitting of the gradient for the edge-preserving potentials introduced in the previous section. In ‘Numerical results’, we present the results of our numerical experiments in the case of confocal microscopy in which photon counting noise is dominant, and therefore, a Poisson noise assumption is more appropriate. We compare the effect of the different potential functions using a simple 3D phantom consisting of spheres with different intensities and conclude that high-quality restorations can be obtained with the hyper-surface potential (Charbonnier *et al.*, 1994) and the Geman–McClure potential (Geman & McClure, 1985), both being superior to the quadratic potential (the Tikhonov regularization). The estimates of the parameters derived from these simulations are used for a deconvolution of real images. In particular, the improvement provided by the deconvolution method is proved by comparing an image obtained with a high numerical aperture (NA) and the restored image obtained from a low NA image of the same object. In ‘Concluding remarks’, we give conclusions.

The edge-preserving approach

In this section, we briefly summarize the maximum likelihood approach in the Poisson noise case and the Bayesian approach and introduce the edge-preserving regularization functionals investigated in this paper.

Maximum likelihood approach

We use bold letters for denoting $N_1 \times N_2 \times N_3$ cubes, whose voxels are labelled by a multi-index $\mathbf{n} = (n_1, n_2, n_3)$. If $\mathbf{g}(\mathbf{n})$ is the value of the image at voxel \mathbf{n} , then we can write

$$\mathbf{g}(\mathbf{n}) = \mathbf{g}_{obj}(\mathbf{n}) + \mathbf{g}_{back}(\mathbf{n}), \quad (1)$$

where $\mathbf{g}_{obj}(\mathbf{n})$ is the number of photoelectrons due to radiation from the object (specimen), and $\mathbf{g}_{back}(\mathbf{n})$ is the number of photoelectrons due to background.

If we assume that noise is dominated by photon counting, these two terms are realizations of independent Poisson processes $\mathbf{G}_{obj}(\mathbf{n})$ and $\mathbf{G}_{back}(\mathbf{n})$ so that their sum $\mathbf{G}(\mathbf{n}) = \mathbf{G}_{obj}(\mathbf{n}) + \mathbf{G}_{back}(\mathbf{n})$ is also a Poisson process; its expected value is given by

$$E\{\mathbf{G}(\mathbf{n})\} = E\{\mathbf{G}_{obj}(\mathbf{n}) + \mathbf{G}_{back}(\mathbf{n})\} = (\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n}), \quad (2)$$

where \mathbf{f} is the object array, \mathbf{b} is the expected value of the background (usually constant over the image domain) and \mathbf{A} is the imaging matrix, and under the linear and space-invariant assumption, is the convolution matrix related to the point spread function (PSF) \mathbf{k} by (Bertero & Boccacci, 1998)

$$\mathbf{A}\mathbf{f} = \mathbf{k} * \mathbf{f}. \quad (3)$$

Because the Poisson processes related to different voxels are statistically independent, the probability distribution of

the multi-valued random variable $\mathbf{G} = \{\mathbf{G}(\mathbf{n})\}$, $P_G(\mathbf{g}|\mathbf{f})$ is the product of the probability distributions of the random variables $\mathbf{G}(\mathbf{n})$, and therefore, the likelihood function is given by

$$\mathcal{L}_g^G(\mathbf{f}) = P_G(\mathbf{g}|\mathbf{f}) = \prod_{\mathbf{n}} e^{-[(\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n})]} \frac{((\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n}))^{\mathbf{g}(\mathbf{n})}}{\mathbf{g}(\mathbf{n})!}, \quad (4)$$

where \mathbf{g} is the detected image.

An estimate of the unknown object \mathbf{f} corresponding to the image \mathbf{g} is defined as any maximum point \mathbf{f}^* of the likelihood function. It is easy to demonstrate that the maximization of the likelihood (ML) function is equivalent to the minimization of the Csiszár I-divergence (Csiszár, 1991) (also called Kullback–Leibler divergence (Kullback & Leibler, 1951)) of the computed image $\mathbf{A}\mathbf{f} + \mathbf{b}$ (associated with \mathbf{f}) from the detected image \mathbf{g} ,

$$\begin{aligned} J_0(\mathbf{f}; \mathbf{g}) &= D_{KL}(\mathbf{g}, \mathbf{A}\mathbf{f} + \mathbf{b}) \\ &= \sum_{\mathbf{n}} \left\{ \mathbf{g}(\mathbf{n}) \ln \frac{\mathbf{g}(\mathbf{n})}{(\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n})} \right. \\ &\quad \left. + [(\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) - \mathbf{g}(\mathbf{n})] \right\}, \end{aligned} \quad (5)$$

which is a convex and non-negative function.

An iterative method for the minimization of $J_0(\mathbf{f}; \mathbf{g})$ was proposed by several authors, in particular, by Richardson (1972) and Lucy (1974) for spectra and image deconvolution and by Shepp & Vardi (1982) for emission tomography. As shown in (Shepp & Vardi, 1982), this method is related to a general approach for the solution of ML problems, known as EM. For these reasons, the algorithm is known as both RL and EM method. The very same method was adopted by Holmes & Liu (1989) and Holmes (1988) for 3D fluorescence microscopy. In this paper, we refer to it as the RLM. For the convenience of the reader, we recall that it is as follows:

- choose an initial guess $\mathbf{f}^{(0)}$ such that

$$\mathbf{f}^{(0)}(\mathbf{n}) > 0 \text{ and} \quad (6)$$

- given $\mathbf{f}^{(i)}$, compute

$$\mathbf{f}^{(i+1)} = \mathbf{f}^{(i)} \left(\mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}} \right). \quad (7)$$

It is evident that each approximation $\mathbf{f}^{(i)}$ is non-negative. For $\mathbf{b} = \mathbf{0}$, several proofs of the convergence of the algorithm to an ML solution are available. Moreover, if the PSF \mathbf{k} is normalized in such a way that the sum of all its voxel values is 1, then for each iteration, the following property is true:

$$\sum_{\mathbf{n}} \mathbf{f}^{(i)}(\mathbf{n}) = \sum_{\mathbf{n}} \mathbf{g}(\mathbf{n}). \quad (8)$$

This property is also called flux conservation because it guarantees that the total number of counts of the reconstructed object coincides with the total number of counts of the detected image. It is not satisfied in the case $\mathbf{b} \neq \mathbf{0}$. Moreover, the convergence of the algorithm seems not to be proved in such a case.

Edge-preserving Bayesian approach

The ML estimate is affected by large noise artefacts that become evident at large number of RLM iterations. In general, early stopping is used for regularizing the solution. However, ML does not incorporate prior knowledge about the unknown object. This can be done in a Bayesian approach.

Indeed, it is assumed that for each \mathbf{n} , $\mathbf{f}(\mathbf{n})$ is a realization of a random variable $\mathbf{F}(\mathbf{n})$. We denote by \mathbf{F} the set of the $\mathbf{F}(\mathbf{n})$ and by $P_{\mathbf{F}}(\mathbf{f})$ its probability distribution, the so-called *prior*. If $P_{\mathbf{G}}(\mathbf{g} | \mathbf{f})$ is interpreted as the conditional probability of \mathbf{G} for a given value \mathbf{f} of \mathbf{F} , then Bayes' formula allows computing of $P_{\mathbf{F}}(\mathbf{f} | \mathbf{g})$, that is, the conditional probability of \mathbf{F} for a given value \mathbf{g} of \mathbf{G} . Finally, the *posterior* probability distribution of \mathbf{F} is obtained from $P_{\mathbf{F}}(\mathbf{f} | \mathbf{g})$ when \mathbf{G} is the detected image. The final result is given by

$$\mathcal{P}_{\mathbf{g}}^{\mathbf{F}}(\mathbf{f}) = \frac{\mathcal{L}_{\mathbf{g}}^{\mathbf{F}}(\mathbf{f}) P_{\mathbf{F}}(\mathbf{f})}{P_{\mathbf{G}}(\mathbf{g})}, \quad (9)$$

where $P_{\mathbf{G}}(\mathbf{g})$ is the probability distribution of \mathbf{G} . The posterior function $\mathcal{P}_{\mathbf{g}}^{\mathbf{F}}(\mathbf{f})$ gives the probability for an object \mathbf{f} , given that the observed data is \mathbf{g} .

Maximization of the posterior function leads to minimizations of functionals of the following form:

$$J_{\mu}(\mathbf{f}; \mathbf{g}) = J_0(\mathbf{f}; \mathbf{g}) + \mu J_R(\mathbf{f}), \quad (10)$$

where, in the case of Poisson noise, $J_0(\mathbf{f}; \mathbf{g})$ is given by Eq. (5), $J_R(\mathbf{f})$ is a regularization functional (called also regularization term) coming from the negative logarithm of the prior and μ is a hyper-parameter that usually is called the *regularization parameter*. A minimum of this functional is generally termed as the MAP estimate of the unknown object. Note that in the case $\mu = 0$, the MAP problem becomes the ML problem.

We conclude by remarking that it is not obvious that a minimum point \mathbf{f}^* of $J_{\mu}(\mathbf{f}; \mathbf{g})$ is a sensible estimate of the unknown object \mathbf{f} . In fact, in this formulation, we have a free parameter μ . In the classical regularization theory (Tikhonov & Arsenin, 1977), a wide literature exists on the problem of the optimal choice of this parameter (Engl *et al.*, 1996; van Kempen & van Vliet, 2000b), but as far as we know, this problem has not yet been thoroughly investigated in the more general framework provided by the Bayesian regularization.

A wide class of priors is provided by MRF, first introduced into image processing by Geman & Geman (1984). The key result concerning MRF is the Hammersley–Clifford theorem, stating that \mathbf{F} is an MRF if and only if it is a suitable GRF so that its probability distribution is given by (Besag, 1974)

$$P_{\mathbf{F}}(\mathbf{f}) = \frac{1}{Z} \exp \left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}} \varphi_c(\mathbf{f}) \right\}, \quad (11)$$

where Z is a normalization constant, β is the Gibbs hyperparameter, \mathcal{C} is the collection of all possible cliques associated to the MRF (we recall that a clique is either a single site or a set of sites such that each one of them is a neighbour of all the other sites in the set.) and φ_c is a *clique potential* that depends only on the values of \mathbf{f} in the sites of the clique c . With such a choice of the prior, the regularization term in Eq. (10) is given by

$$J_R(\mathbf{f}) = \sum_{c \in \mathcal{C}} \varphi_c(\mathbf{f}), \quad (12)$$

and the regularization parameter μ becomes equal to $1/\beta$.

In general, clique potentials are functions of a derivative of the object (Geman & Reynolds, 1992). The order of the derivative and the neighbourhood system are chosen depending on the kind of object that is sought. In this work, we assume that the object consists of piecewise smooth regions separated by sharp edges and use first-order differences between voxels belonging to the double-site 3D clique shown in Fig. 1 so that Eq. (12) becomes

$$J_R(\mathbf{f}) = \frac{1}{4} \sum_n \sum_{m \in \mathcal{N}_n} \varphi \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})\delta} \right), \quad (13)$$

where \mathcal{N}_n is the set of the first neighbours of the voxel \mathbf{n} as derived from Fig. 1; δ and $d(\mathbf{n}, \mathbf{m})$ are two scaling parameters, the first tuning the value of the gradient above which a discontinuity is detected, the second taking into account the different distances between the voxels of a clique. For example, in a real case in which sampling in the lateral direction (δ_{xy}) is different from sampling in the axial direction (δ_z), $d(\mathbf{n}, \mathbf{m})$

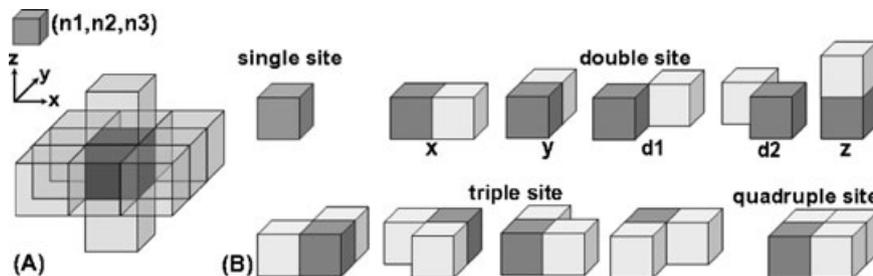


Fig. 1. (A) The set of neighbours of a voxel $\mathbf{n} = (n_1, n_2, n_3)$ used in this paper and (B) the corresponding double-site cliques.

Table 1. Potential functions considered in this paper: φ_Q = quadratic potential; φ_{GM} = Geman–McClure potential; φ_{HL} = Hebert–Leahy potential; φ_{HB} = Huber potential; φ_{HS} = hyper-surface potential.

Potential function	Expression of $\varphi(t)$	Expression of $\psi(t) = \varphi'(t)/2t$	Reference
φ_Q	t^2	1	(Charbonnier <i>et al.</i> , 1997)
φ_{GM}	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	(Geman & McClure, 1985)
φ_{HL}	$\log(1+t^2)$	$\frac{1}{1+t^2}$	(Hebert & Leahy, 1989)
φ_{HB}	$\begin{cases} t^2, & t \leq 1 \\ 2 t - 1, & t > 1 \end{cases}$	$\begin{cases} 1, & t \leq 1 \\ 1/ t , & t > 1 \end{cases}$	(Schultz & Stevenson, 1995)
φ_{HS}	$2\sqrt{1+t^2} - 2$	$\frac{1}{\sqrt{1+t^2}}$	(Charbonnier <i>et al.</i> , 1994)

are defined as follows:

$$d(\mathbf{n}, \mathbf{m}) = \begin{cases} 1, & \mathbf{m} \in \{(n_1 \pm 1, n_2, n_3), (n_1, n_2 \pm 1, n_3)\} \\ \sqrt{2}, & \mathbf{m} \in \{(n_1 \pm 1, n_2 \pm 1, n_3)\} \\ \delta_{z,xy}, & \mathbf{m} \in \{(n_1, n_2, n_3 \pm 1)\}. \end{cases} \quad (14)$$

Higher-order neighbourhood systems, combined with second-order derivatives, promote the formation of piecewise linear areas in the solution; combined with third-order derivatives, they promote piecewise planar areas (Geman & Reynolds, 1992).

In this work, we refer to φ as the *potential function*. Its form takes a key part in the estimation of the object to be restored.

Table 1 shows different potential functions proposed in the literature and their corresponding *weighting functions* $\psi(t) = \varphi'(t)/2t$, where φ' stands for the first derivative of φ . In the standard Tikhonov regularization (Tikhonov & Arsenin, 1977), the potential function takes the pure quadratic form φ_Q .

It is well known that quadratic regularization imposes smoothness constraint everywhere; because large gradients in the solution are penalized, the result is that edges are completely lost. To overcome this problem, a series of potential functions were proposed. In recent years, the problem such as ‘What properties must a potential function satisfy to ensure the preservation of edges?’ has provided abundant literature (see, for instance, Geman & McClure, 1985; Hebert & Leahy, 1989; Schultz & Stevenson, 1995). To unify all these approaches, Charbonnier *et al.* (1997) proposed a set of general conditions that the edge-preserving regularization potential function should satisfy, independently of the algorithm used to minimize Eq. (10). For the convenience of the reader, these conditions are summarized below:

- (a) $\varphi(t) \geq 0 \forall t$ and $\varphi(0) = 0$;
- (b) $\varphi(t) = \varphi(-t)$;

(c) φ continuously differentiable;

(d) $\varphi'(t) \geq 0 \forall t \geq 0$;

(e) $\psi(t) = \frac{\varphi'(t)}{2t}$ continuous and strictly decreasing on $[0, +\infty)$;

(f) $\lim_{t \rightarrow +\infty} \psi(t) = 0$ and

(g) $\lim_{t \rightarrow 0} \psi(t) = M, 0 < M < +\infty$.

All the potential functions tested in this work (see Table 1) are in complete agreement with these conditions, except φ_{HB} . In fact, the weighting function of φ_{HB} is not strictly decreasing on $[0, +\infty)$, as required by condition (e), but only for $t > 1$. However, tests on both synthetic and real data show that φ_{HB} also leads to well-preserved edge restorations. Therefore, we believe that condition (e) can be weakened and must be satisfied only for sufficiently large values of t .

Split-gradient method (SGM)

An iterative method for the minimization of Eqs. (10), (5) and (13), with the additional constraints of non-negativity and flux conservation,

$$\mathbf{f}(\mathbf{n}) \geq 0, \quad \sum_{\mathbf{n}} \mathbf{f}(\mathbf{n}) = \sum_{\mathbf{n}} \{\mathbf{g}(\mathbf{n}) - \mathbf{b}(\mathbf{n})\} \doteq c, \quad (15)$$

can be obtained by means of a general approach known as the SGM, proposed by Lantéri *et al.* (2001, 2002). The method provides iterative algorithms that, in general, are not very efficient (they require a large number of iterations) but have a very nice feature: for very broad classes of regularization terms, they can be obtained by means of a very simple modification of RLM and therefore can be very easily implemented. Therefore, it is possible to compare in a simple way the results obtainable by means of different regularization terms. In particular, we will apply the method to all the regularization terms derived from the potential functions of Table 1.

The SGM is based on the following decomposition of the gradient of the functional of Eq. (10):

$$\begin{aligned} -\nabla_{\mathbf{f}} J_0(\mathbf{f}; \mathbf{g}) &= \mathbf{U}_0(\mathbf{f}; \mathbf{g}) - \mathbf{V}_0(\mathbf{f}; \mathbf{g}), \\ -\nabla_{\mathbf{f}} J_R(\mathbf{f}) &= \mathbf{U}_R(\mathbf{f}) - \mathbf{V}_R(\mathbf{f}), \end{aligned} \quad (16)$$

where $\mathbf{U}_0(\mathbf{f}; \mathbf{g})$ and $\mathbf{V}_0(\mathbf{f}; \mathbf{g})$ are positive arrays depending on \mathbf{f} and \mathbf{g} , and $\mathbf{U}_R(\mathbf{f})$ and $\mathbf{V}_R(\mathbf{f})$ are non-negative arrays depending on \mathbf{f} . It is obvious that such a decomposition always exists but is not unique. The general structure of the iterative algorithm, as described in (Lantéri *et al.*, 2001), is as follows:

- choose an initial guess $\mathbf{f}^{(0)} > 0$, satisfying the constraints of Eq. (15);
- given $\mathbf{f}^{(i)}$, compute

$$\begin{aligned} \tilde{\mathbf{f}}^{(i+1)} &= \mathbf{f}^{(i)} \left(\frac{\mathbf{U}_0(\mathbf{f}^{(i)}; \mathbf{g}) + \mu \mathbf{U}_R(\mathbf{f}^{(i)})}{\mathbf{V}_0(\mathbf{f}^{(i)}; \mathbf{g}) + \mu \mathbf{V}_R(\mathbf{f}^{(i)})} \right), \\ \tilde{c}^{(i+1)} &= \sum_{\mathbf{n}} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n}) \text{ and} \end{aligned} \quad (17)$$

- set

$$\mathbf{f}^{(i+1)} = \frac{c}{\tilde{c}^{(i+1)}} \tilde{\mathbf{f}}^{(i+1)}.$$

It is important to remark some interesting features. The first is that all the iterates are automatically non-negative if the initial guess $\mathbf{f}^{(0)}$ is non-negative, as one can easily verify. The second is that the algorithm is a scaled gradient method, with step size 1, because the update step can be written in the following form:

$$\tilde{\mathbf{f}}^{(i+1)} = \mathbf{f}^{(i)} - S^i \left[\nabla_{\mathbf{f}} J_0(\mathbf{f}^{(i)}; \mathbf{g}) + \mu \nabla_{\mathbf{f}} J_R(\mathbf{f}^{(i)}) \right], \quad (18)$$

where

$$S^i = \text{diag} \left\{ \frac{\mathbf{f}^{(i)}(\mathbf{n})}{[\mathbf{V}_0(\mathbf{f}^{(i)}; \mathbf{g}) + \mu \mathbf{V}_R(\mathbf{f}^{(i)})](\mathbf{n})} \right\}. \quad (19)$$

The convergence of the algorithm has not been proved in general, but only in some particular cases (Lantéri *et al.*, 2001); nevertheless, it has been verified experimentally in all the applications that we have considered. Moreover, it is not difficult to prove that if the sequence of the iterates is convergent, then the limit is a constrained minimum point of the functional in Eq. (10) (Bertero *et al.*, 2008).

In the case with PSF \mathbf{k} normalized in such a way that the sum all its voxel values is 1, a possible SGM decomposition of $\nabla_{\mathbf{f}} J_0(\mathbf{f}; \mathbf{g})$, both for the Poisson and the additive Gaussian noise cases, is given by Bertero *et al.* (2008). Because we are mainly interested in the Poisson case, we report only the corresponding decomposition

$$\mathbf{U}_0(\mathbf{f}; \mathbf{g}) = \mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f} + \mathbf{b}}, \quad \mathbf{V}_0(\mathbf{f}; \mathbf{g}) = \mathbf{1}, \quad (20)$$

where $\mathbf{1}$ is the array whose entries are all equal to 1. Therefore, if we insert these expressions in Eq. (17) with $\mu = 0$, we obtain RLM, with the flux conservation constraint (Eq. (7)).

The gradient of the regularization term $J_R(\mathbf{f})$ of Eq. (13) is given by

$$(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(\mathbf{n}) = \frac{1}{\delta^2} \sum_{\mathbf{m} \in \mathcal{N}_n} \frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})^2} \psi \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})\delta} \right). \quad (21)$$

Using conditions (b) and (d), it is easy to demonstrate that $\psi(t) \geq 0$ for all t . Thus, by taking into account that we are considering the restriction of the regularization functional $J_R(\mathbf{f})$ to the closed and convex set of the non-negative arrays, we can choose the following decomposition of the gradient:

$$\begin{aligned} (\mathbf{U}_R(\mathbf{f}))(\mathbf{n}) &= \frac{1}{\delta^2} \sum_{\mathbf{m} \in \mathcal{N}_n} \frac{\mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})^2} \psi \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})\delta} \right), \\ (\mathbf{V}_R(\mathbf{f}))(\mathbf{n}) &= \frac{1}{\delta^2} \sum_{\mathbf{m} \in \mathcal{N}_n} \frac{\mathbf{f}(\mathbf{n})}{d(\mathbf{n}, \mathbf{m})^2} \psi \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})\delta} \right). \end{aligned} \quad (22)$$

In conclusion, the iterative step of Eq. (17) becomes

$$\tilde{\mathbf{f}}^{(i+1)} = \frac{\mathbf{f}^{(i)}}{\mathbf{1} + \mu \mathbf{V}_R(\mathbf{f}^{(i)})} \left(\mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f} + \mathbf{b}} + \mu \mathbf{U}_R(\mathbf{f}^{(i)}) \right). \quad (23)$$

By looking at Eq. (7), we see that this regularized algorithm is a simple modification of the RLM. As we have already remarked, even if we do not have a proof of convergence, convergence has always been observed in our numerical experiments. In any case, convergence can be obtained by a suitable line search (Lantéri *et al.*, 2002) because, as shown in Eq. (18), this algorithm is a scaled gradient method with step length 1.

It may be interesting to compare our algorithm with the OSL algorithm proposed by Green (1990) that, in our notations, is given by

$$\tilde{\mathbf{f}}^{(i+1)} = \frac{\mathbf{f}^{(i)}}{\mathbf{1} + \mu \nabla_{\mathbf{f}} J_R(\mathbf{f}^{(i)})} \mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f} + \mathbf{b}}. \quad (24)$$

We note that in our algorithm, the denominator is strictly positive (in fact ≥ 1) for any value of μ , whereas this is not true for OSL. Indeed, it is known that OSL may not converge (Pierro & Yamagishi, 2001) and may not produce images with non-negative intensities. This is mainly due to possible oscillations of the gradient of the regularization terms, which can be amplified by the regularization parameter. Because OSL is also a scaled gradient method with step length 1, in Lange (1990), a line search was added to the algorithm, proving convergence and non-negativity of the iterates in the case of sufficiently small values of μ . To solve this problem, the original algorithm was further modified (Gravier & Yang, 2005; Mair & Zahnen, 2006) to ensure positivity of the estimates and (numerical) convergence of the negative logarithm of the posterior function; however, these methods lead to cumbersome steps in the algorithm.

Numerical results

We apply the previous method to the particular case of confocal scanning laser microscopy (CSLM) in which Poisson noise assumption is more realistic.

Point spread function model

For CLSM, the 3D PSF is defined by

$$k_{CLSM}(x, y, z) = (R(x, y) * |\vec{h}_{\lambda_{em}}(x, y, z)|^2) \times |\vec{h}_{\lambda_{ex}}(x, y, z)|^2, \quad (25)$$

where R describes the geometry of the pinhole, $*$ is the convolution operator, $\vec{h}_{\lambda_{ex}}$ is the time-independent part of the electric field generated by the excitation laser light (at wavelength λ_{ex}) in the focus of the objective and $\vec{h}_{\lambda_{em}}$ is similar to $\vec{h}_{\lambda_{ex}}$, but for the emission light (at wavelength λ_{em}).

The electric field distribution \vec{h} in the focal region for an aberration-free lens is given in Richards & Wolf (1959). This distribution depends on the wavelength of the light, the NA of the lens, and the refractive index of the medium in which the lens works. Note that in this work, we use the well-developed vectorial theory of light, because for real data experiments, we use high NA objectives.

Simulated data

A first experiment is the restoration of a synthetic image with simulated degradation. Figure 2(A) represents a sphere containing five ellipsoids with different sizes and intensities ($128 \times 128 \times 64$ voxels; intensity of the larger sphere: 100 units, larger ellipsoid: 200 units, other structures: 255 units). To simulate CLSM image formation, the object is convolved

with a confocal PSF, a constant background of 5 units is added and finally the result is corrupted by Poisson noise (Fig. 2). Confocal PSF is computed using Eq. (25) with the following acquisition parameters: excitation wavelength $\lambda_{ex} = 488$ nm; emission wavelength $\lambda_{em} = 520$ nm; NA of the objective = 1.4; refractive index of the oil in which the objective is embedded = 1.518; diameter of the circular pinhole = 1 Airy unit, which corresponds to a backprojected radius on the sample of ~ 220 nm.

The Nyquist samplings for a confocal system under the acquisition parameters described above are ~ 150 nm along the optical axis and ~ 50 nm along the lateral axis. In this simulation, we oversample the object with respect to the Nyquist criteria, assuming $\delta_{ey} = 35$ nm in the lateral direction and $\delta_{ez} = 105$ nm in the axial direction. To change the signal-to-noise ratio (SNR) of the simulated image, we change τ , the reciprocal of the photon conversion factor. We assume that the noise is described by a Poisson process, implying that the SNR in fluorescence imaging depends solely on the total number of detected photons. For this reason, we choose the mean of the Poisson process equal to $\tau(Af + b)$, with fixed intensities of $Af + b$. Thus, by increasing τ , the average number of detected photons increases and hence the noise level decreases. In the real life, photon conversion factor is determined by several multiplicative factors such as integration time and quantum efficiency of the detector (Jovin

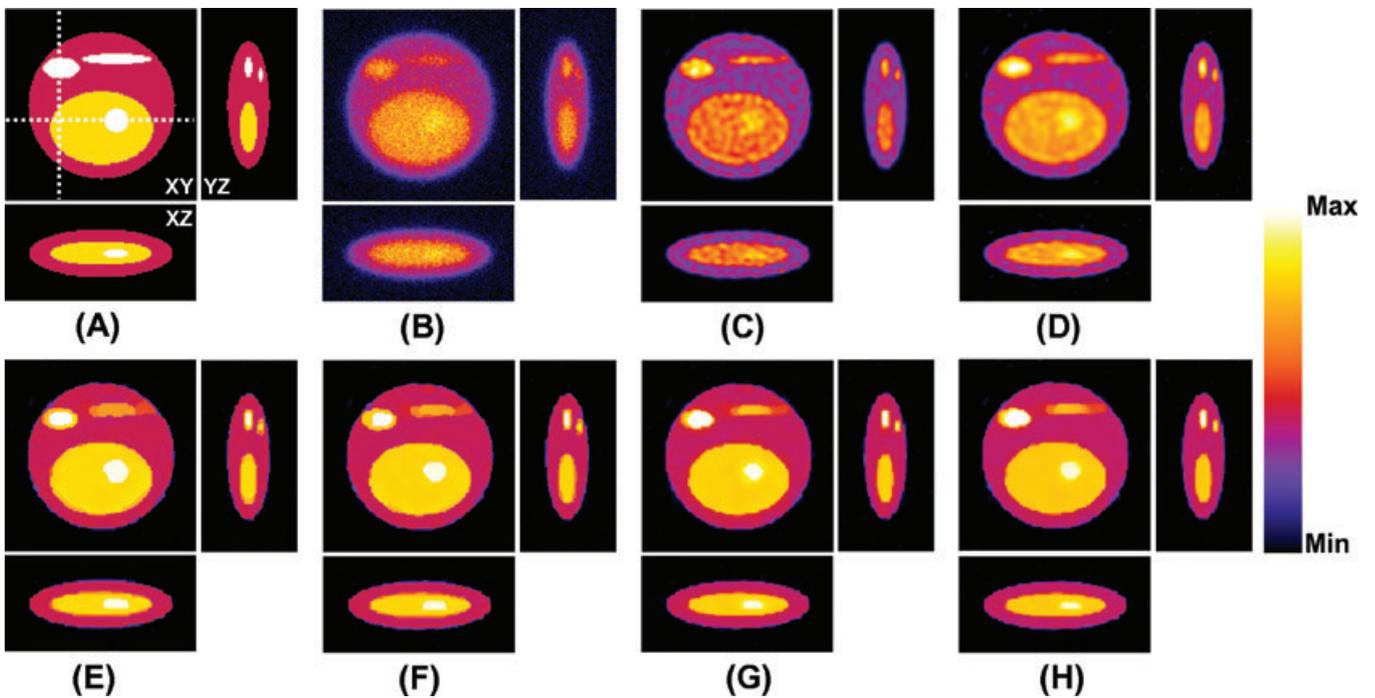


Fig. 2. Slice view (slice no. 33) and two axial views along the dotted lines of the original object (A), simulated confocal image (B) and restored objects using RLM (C), quadratic potential (D), Geman–McClure potential (E), Hebert–Leahy potential (F), Huber potential (G) and hyper-surface potential (H). The simulated confocal image is corrupted by Poisson noise, with a value $\tau = 0.46$ for the reciprocal of the photon conversion factor, corresponding to $SNR = 20$; this means that the data are normalized to a maximum of 100 counts before the Poisson noise filter is applied.

& Arndt-Jovin, 1989). The relation between SNR and τ is given by

$$SNR = 10 \log \left[\tau \max_{\mathbf{n}} (\mathbf{A}\mathbf{f} + \mathbf{b})(\mathbf{n}) \right]. \quad (26)$$

For a numerical comparison of the quality of the restored images obtained with the proposed SGM method, coupled with the different potential functions, we use the Kullback–Leibler divergence of the restored object from the true one, which is the best measure in the presence of a non-negativity constraint (Csiszár, 1991). We recall that the Kullback–Leibler divergence (KL_D) of a 3D image \mathbf{q} from a 3D image \mathbf{r} is defined by

$$KL_D(\mathbf{r}, \mathbf{q}) = \frac{1}{N} \sum_{\mathbf{n}} \left\{ \mathbf{r}(\mathbf{n}) \ln \frac{\mathbf{r}(\mathbf{n})}{\mathbf{q}(\mathbf{n})} + [\mathbf{q}(\mathbf{n}) - \mathbf{r}(\mathbf{n})] \right\}, \quad (27)$$

where $N = N_1 \times N_2 \times N_3$. Note that KL_D is non-symmetric, that is, $KL_D(\mathbf{r}, \mathbf{q}) \neq KL_D(\mathbf{q}, \mathbf{r})$.

In Fig. 3, for each proposed algorithm, we plot the behaviour of $KL_D(\mathbf{f}, \mathbf{f}^{(i)})$ as a function of the number of the iterations i (SNR = 20). Numerical instability of RLM is evident. Moreover, it is important to note that there is a minimum also in the case of GM and HL (semiconvergence of the method), whereas HB, HS and Q are convergent.

The Kullback–Leibler divergence can also be used for finding a suitable criterion for stopping the iterations, especially in the case of algorithms with a convergent behaviour. Indeed, we stop the iterations when the difference between the values of $KL_D(\mathbf{f}, \mathbf{f}^{(i)})$ corresponding to two consecutive iterations is smaller than a certain threshold T

$$KL_D(\mathbf{f}, \mathbf{f}^{(i)}) - KL_D(\mathbf{f}, \mathbf{f}^{(i+1)}) \leq T. \quad (28)$$

A simple rule for choosing T does not exist, and its choice follows from our numerical experience. In this work, we use $T = 10^{-6}$ as a reliable value for obtaining satisfactory

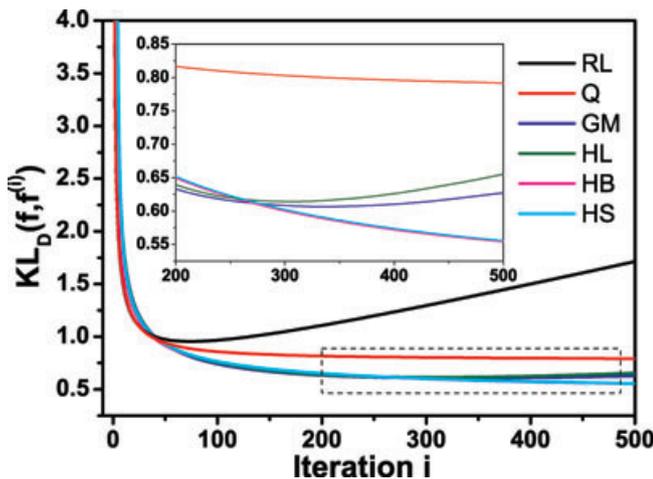


Fig. 3. Behaviour of the KL_D of the iterates $\mathbf{f}^{(i)}$ from the original object \mathbf{f} as a function of the number of iterations for RLM and for the algorithms regularized with the potential functions of Table 1 (SNR = 20).

reconstructions. In the case of convergence (Q, HB and HS), this means that the reconstructed solution $\hat{\mathbf{f}}$ and the associated $KL_D(\mathbf{f}, \hat{\mathbf{f}})$ value do not change much after the threshold has been reached.

As already discussed, all edge-preserving potential functions contain a scaling parameter δ . Thus, for an accurate quantitative evaluation of the proposed regularized algorithms, it is necessary to compare their results using appropriate values of the parameters β and δ . Again, the KL_D criterion is used for obtaining these values that we define as those providing the minimum of $KL_D(\mathbf{f}, \hat{\mathbf{f}})$. Figure 4 shows, for each potential function, the behaviour of the $KL_D(\mathbf{f}, \hat{\mathbf{f}})$ as a function of the regularization parameter β for different values of the scaling parameter δ in the case of the object of Fig. 2 with SNR = 20. In order to reduce the computational burden, the iterations, initialized with a uniform image, are stopped with a threshold $T = 10^{-4}$. In Table 2, we report the values obtained with this procedure.

The reconstructed images $\hat{\mathbf{f}}$ of the object of Fig. 2, with different SNR values, are computed using the optimal values of Table 2 and a threshold $T = 10^{-6}$. In order to measure the quality of the reconstructions, we introduce an *improvement factor* (IF) defined by

$$IF = \frac{KL_D(\mathbf{f}, \mathbf{g}) - KL_D(\mathbf{f}, \hat{\mathbf{f}})}{KL_D(\mathbf{f}, \mathbf{g})}. \quad (29)$$

IF measures the relative global improvement provided by the reconstructed image $\hat{\mathbf{f}}$ with respect to the initial raw image \mathbf{g} in terms of the KL_D divergence from the true object. It is equal to 1 in the case of a perfect reconstruction. Its values are reported in Table 3. As expected, they increase when the SNR increases. They indicate that a considerable improvement can be provided by the use of a reconstruction method; however, they do not discriminate significantly between the different methods. In Table 4, we also report the number of iterations and the computational cost per iteration in C++ software environment (Version 6.0, Microsoft Corporation, USA). An HP workstation XW4200 (Hewlett-Packard Company, Palo Alto, CA, USA) Pentium(R)4 CPU 3.40 GHz equipped with 2.00 GB of RAM is used for reconstruction. All computations, including the simulation of the image formation process (i.e. blurring and noise), are performed with double precision. The asterisk indicates the cases in which the threshold is not reached within the maximum number of iterations allowed by the code. We see that the computational costs per iteration for the different algorithms are comparable.

The parameter IF may be a good choice as a global measure of the quality of a restoration, but it may not be able to detect small artefacts that may be important for deciding the reliability of the reconstruction. Thus, the previous comparisons are only the first step in determining which algorithm is the most suitable for a given application.

In order to complete our analysis, we also considered visual inspection as direct evidence for quantifying the

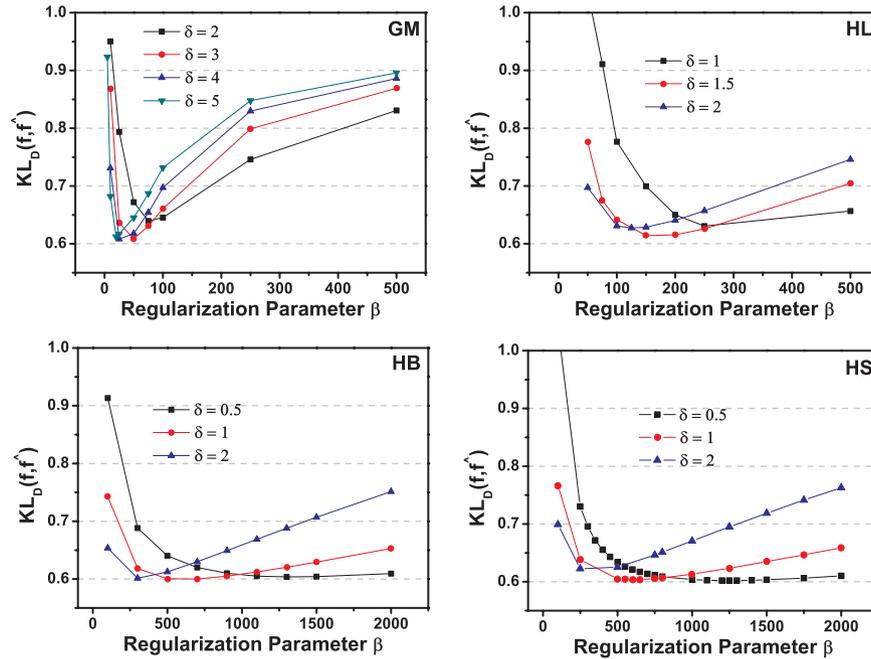


Fig. 4. Relationship between the parameters β and δ (SNR = 20).

Table 2. Optimal values of the parameters δ and β for the different regularization potentials and different values of SNR.

Noise	Q		GM		HL		HB		HS		
	δ	β									
15	0.145	2	2250	4	20	1.5	125	0.5	500	0.5	600
20	0.461	2	5000	4	25	1.5	150	1	700	0.5	1200
25	1.457	2	15000	5	20	2	125	1.5	600	0.5	1800

Table 3. Values of the improvement factor IF corresponding to the reconstructions provided by the different methods and for different SNR values. For each SNR, the table reports both the mean value (upper lines) and the standard deviation (lower lines) obtained with 10 different realizations of noise.

SNR	RLM	Q	GM	HL	HB	HS
15	0.7187	0.7727	0.8086	0.8017	0.8321	0.8315
	± 0.0014	± 0.0012	± 0.0017	± 0.0031	± 0.0021	± 0.0012
20	0.7409	0.7878	0.8381	0.8348	0.8572	0.8581
	± 0.0012	± 0.0008	± 0.0031	± 0.0021	± 0.0020	± 0.0029
25	0.7745	0.8140	0.8667	0.8622	0.8848	0.8857
	± 0.0011	± 0.0011	± 0.0014	± 0.0022	± 0.0009	± 0.0012

quality of the 3D reconstructions. It is evident that RLM-restored images (Fig. 2C) suffer severely from noise artefacts, known as the checkerboard effect (many components of the solution are too small, whereas others grow up), whereas the smoothing nature of the quadratic potential is evident in Q-restored images (Fig. 2D). All the restorations using edge-

preserving potential functions (Figs. 2E–H) exhibit excellent performance in both suppressing noise effect and preserving edges: the corresponding restored images are free from the unfavourable oversmoothing effect and the checkerboard effect. It is important to note that the GM- (Fig. 2E) and HL-restored (Fig. 2F) images are very similar. A similar behaviour can be observed by comparing HB- and HS-restored images (Fig. G and H, respectively). These results are in complete agreement with the values of IF . It is also interesting to remark that GM and HL restorations present very sharp edges in comparison to HB and HS restorations. This behaviour becomes more evident from line plot investigations. Figure 5 shows the intensity profiles along pre-defined lines through the original object and the restored images, both for lateral (Figs. 5A–C) and for axial views (Figs. 5D and E). It is evident that the profiles of the edge-preserving reconstructions have a much better agreement with the profiles of the synthetic object. The insets of Figs. 5(E) and (F) show in detail the edge-preserving capability of the different potential functions and reveal clearly the very sharp results obtained with GM and HL in comparison to HB and HS.

However, the capability of GM and HL to obtain very sharp edges represents also the limit of these potential functions. As it is possible to observe in the restored simulated images, GM (Fig. 2E) and HL (Fig. 2F) tend to produce blocky piecewise regions that are not present in the original object. This effect, also called the staircase effect, is thereby reduced in HB- (Fig. 2G) and HS- (Fig. 2H) restored images. In other words, HL and GM produce very sharp edges but also introduce ‘false’ edges in the restoration.

Table 4. Number of iterations used for the optimal reconstructions provided by the different methods. The asterisk indicates the cases in which the convergence was not reached before the maximum allowed number of iterations (1000). The table reports the mean value and the standard deviation obtained by means of ten different realizations of noise. For each method, we also give the computational time per iteration.

SNR	RLM	Q	GM	HL	HB	HS
15	37 ± 1	1000*	192 ± 6	178 ± 5	751 ± 48	783 ± 45
20	76 ± 1	1000*	355 ± 15	307 ± 10	1000*	1000*
25	165 ± 3	1000*	404 ± 13	388 ± 10	1000*	1000*
Iteration time (s)	1.82	2.35	2.57	2.61	2.65	3.19

To better demonstrate and quantify such a drawback, we apply an edge detection algorithm to the restored images, in particular, a Laplacian- and second derivative of the gradient direction (SDGD)-based zero-crossing procedure (Verbeek & van Vliet, 1994). A threshold step on the image resulting from this algorithm gives us a binary image representing the edges in the restored image. The very same algorithm is applied to the original object and the results are compared. The difference between the binary map associated to the original object and that associated to the restored image gives the distribution of what we call ‘bad voxels’. A voxel of the binary map associated to the restored image is defined ‘bad’ in two cases: (1) the voxel contains an edge in the map of the original object but not in the map of the restored image (negative bad voxel); (2) the voxel does not contain an edge in the map of the original object but contains an edge in the map of the restored image (false-positive bad voxel). Figure 6 compares the number of ‘bad voxels’ for the different potential functions at different noise levels (for simplicity, we consider only Q, GM and HS). As expected, the edge-preserving capability of HS and GM potential functions is better than that of the Q potential. Moreover, it is important to note that for all noise levels, the number of ‘bad voxels’ of GM is always greater than that of HS; in particular, although the number of ‘negative bad voxels’ is comparable for the two methods, the number of ‘false-positive bad voxel’ is larger for GM. This is in agreement with the drawback of GM due to the generation of ‘false’ edges during the restoration process. No appreciable difference is observed between GM and HL and between HS and HB.

In conclusion, the convex potentials HB and HS provide better values, both in terms of the parameter IF and in terms of the edge-preserving capability, than the non-convex ones, GM and HL. However, we believe that the difference between the edge-preserving capabilities of these methods tends to disappear for higher SNR, as can be observed for SNR = 25 (Fig. 6).

Real data

To test the validity of our methods, we compare the different algorithms on real confocal images. To this purpose,

we use both a well-known pattern such as fluorescence beads and a more intricate biological specimen such as bovine pulmonary artery endothelial (BPAE) cells. The 3D image of green Tetraspeck™ beads (Molecular Probes) of ~500-nm diameter is shown in Fig. 7(A). BPAE cells are stained with a combination of fluorescent dyes. F-actin was stained using green fluorescent BODYPY FL phalloidin (Molecular Probes, Invitrogen, Milan, Italy), as shown in Fig. 8(A), and mitochondria were labelled with red fluorescent MitoTracker Red CMXRos (Molecular Probes), as shown in Fig. 8(B). The fluorescence excitation wavelength for MitoTracker is $\lambda_{ex} = 543$ nm; the emission wavelength is $\lambda_{em} = 600$ nm; the fluorescence excitation wavelength for F-actin and green beads is $\lambda_{ex} = 488$ nm; the emission wavelength is $\lambda_{em} = 520$ nm; the pinhole size was set to 1 Airy unit (backprojected radius of 208 nm); sampling is 45 nm in the lateral direction and 135 nm in the axial direction for BPAE images, whereas for beads image, it is 30 nm in the lateral direction and 90 nm in the axial direction.

The imaging device is a Leica TCS (true confocal scanning system) SP5 (Leica Microsystems, Heidelberg, Germany) equipped with a variable NA objective (immersion oil HXC PL APO 63×/NA = 0.6–1.4) (Leica Microsystems).

For real-data experiment, it is not easy to carry out an unambiguous comparison of the different algorithms, particularly because the regularization parameter, the scaling parameter or the optimal number of iterations can not be determined in the same way as is used for synthetic images. To partially overcome this problem, we try several values of the regularization and scaling parameters and inspect the resulting image visually to obtain a restoration with optimal sharpness of the edges without introducing artefacts due to noise amplification. In particular, in the case of beads restoration, where we know the exact size, we pay attention to use parameters that do not lead to the underestimate dimension of the beads. Moreover, we use the following criterion for stopping the iterations of regularized and unregularized algorithms, based on the Kullback–Leibler divergence of $\mathbf{Af}^{(i)} + \mathbf{b}$ from \mathbf{g}

$$KL_D(\mathbf{g}, \mathbf{Af}^{(i)} + \mathbf{b}) - KL_D(\mathbf{g}, \mathbf{Af}^{(i+1)} + \mathbf{b}) \leq T. \quad (30)$$

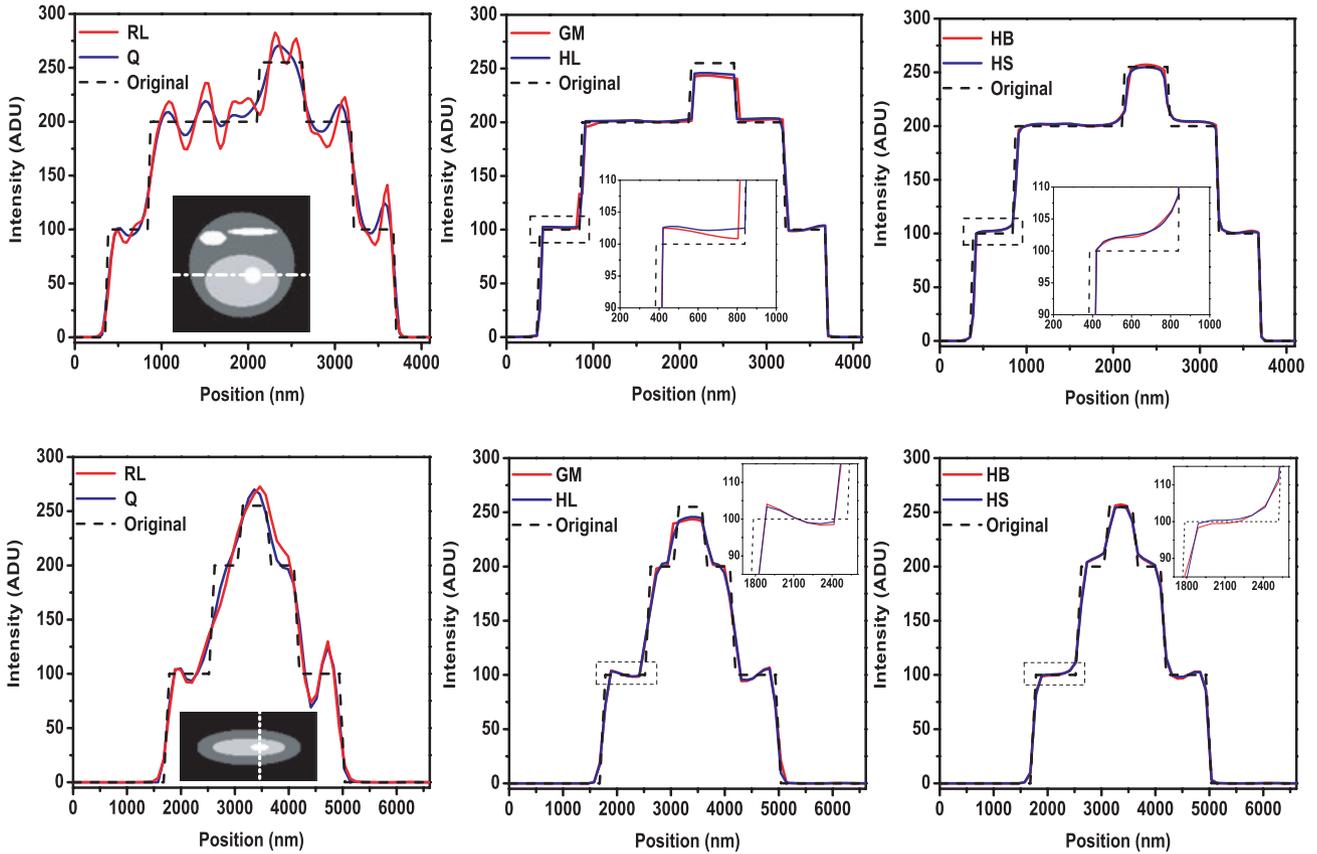


Fig. 5. Intensity profiles along a lateral line (A–C) and an axial line (D–F) in the original object and in the restored object. Insets of (A) and (D) represent, respectively, the regions where the intensity profile are obtained (SNR = 20).

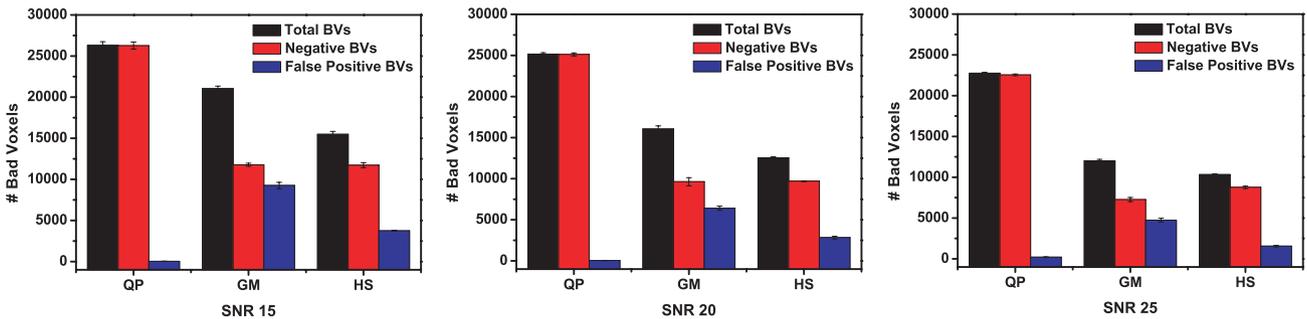


Fig. 6. Plot bar representing the distribution of the numbers of ‘bad voxels’ (BVs), ‘negative bad voxels’ and ‘false-positive bad voxel’ for different restoration methods and for different SNRs. Each value represents the mean value obtained with ten different realizations of noise. The variance of these data is so small that it is not visible.

A threshold of $T = 10^{-4}$ is used for image restoration of both beads and BPAE cells. For mitochondria and F-actin images, a background of 7 units is estimated using the histogram-based methods (van Kempen & van Vliet, 2000a), whereas a null background is estimated for beads image.

In agreement with the numerical results, we do not observe particular differences between GM- and HL-restored images as well as between HS and HB restorations; therefore, we report only the results obtained by GM and HS regularization.

Appreciable restoration of the localized beads can be seen using edge-preserving regularization, whereas oversmoothing is evident in Q-restored images and noisy artefact in RLM-restored images. The drawbacks and the ability of the different restoration methods can be much more appreciated when the previously introduced edge detection algorithm is applied on the respective restored images. The insets of Fig. 7 show how GM and HS thereby improve the performance of the edge detection algorithm. Similar skills

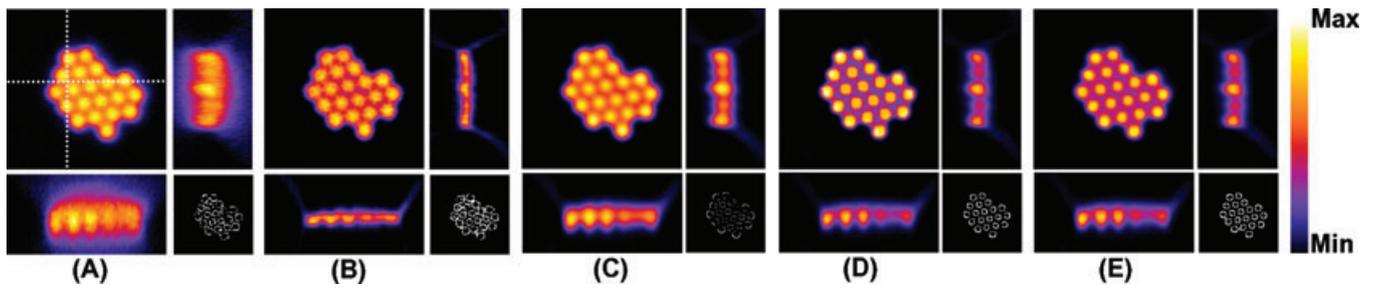


Fig. 7. Restoration of confocal images of fluorescent beads. (A) Raw image (B–E). Restored objects: RLM restoration (B). Regularized restoration using quadratic potential (C), Geman–McClure potential (D) and hyper-surface potential (E). For each raw image and restored object, two axial views along the dotted lines and the result of the edge-detection algorithm applied on the image are reported.

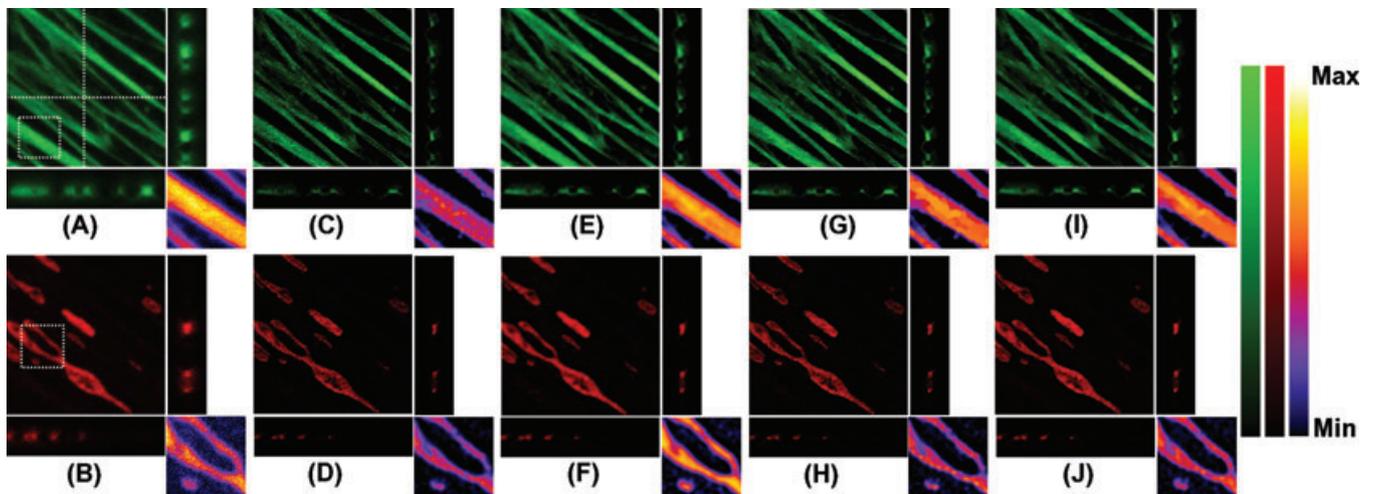


Fig. 8. Restoration of confocal images of BPAE cells. (A) Raw image of F-actin sub-cellular structures. (B) Raw image of mitochondria sub-cellular structures. (C–N) Restored objects associated to F-actin and mitochondria images: RLM restorations (C, D). Regularized restoration using quadratic potential (E, F), Geman–McClure potential (G, H) and hyper-surface potential (I, J). For each raw image and restored object, two axial views along the dotted lines are reported. To better appreciate the differences between the different restorations, the insets show the magnification of the area in the white box.

and drawbacks can be observed from BPAE cell experiment. The insets of Fig. 8 representing a magnification of a particular region of the sample are introduced in the original images and restored images in order to better appreciate the difference between the methods. Checkerboard noise artefacts are clearly observable in RLM restoration; blocky piecewise region presented in the GM restoration are completely avoided in the case of HS.

The improvement obtained by means of deconvolution becomes essential for high-quality 3D volume renderings. Figure 9 shows the 3D volume renderings for the raw and HS-restored stacks of Fig. 8.

Finally, a nice trick to qualitatively compare the performance of the proposed algorithms is to acquire the same sample at different resolution levels using a variable NA objective. Figure 10(A) shows F-actin structures acquired

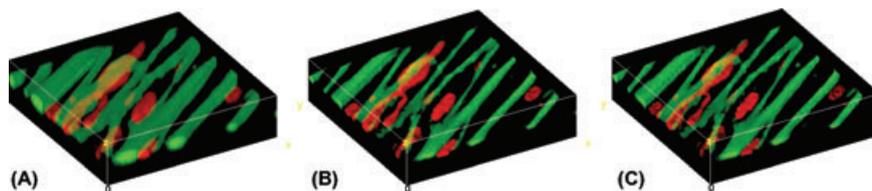


Fig. 9. Three-dimensional volume rendering of raw and deconvolved images of BPAE cells. The volume rendering represents the same structures of Fig. 8. (A) Raw image. (B) Quadratic potential restoration. (C) Hyper-surface potential restoration.

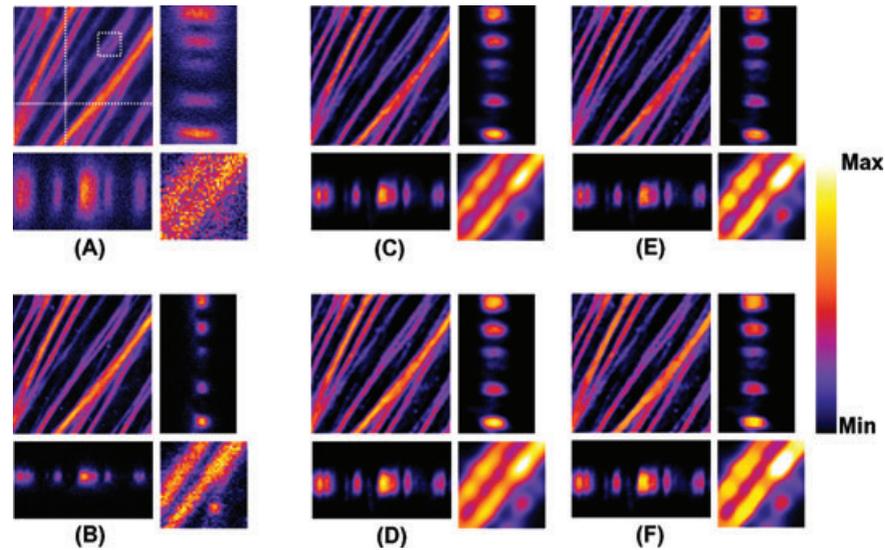


Fig. 10. Comparison between restored low numerical aperture images and raw high numerical aperture images. (A, B) Raw F-actin images obtained using a low numerical aperture objective ($NA = 0.6$) and a high numerical aperture objective ($NA = 1.4$), respectively. Note that both images represent the same structure. Restored objects obtained from low numerical aperture image using RLM (C) and regularized methods with quadratic potential (D), Geman–McClure potential (E) and hyper-surface potential (F). For each raw image and restored object, two axial views along the dotted lines are reported.

using the objective under the minimal NA ($NA = 0.6$). The resolution improvement is evident using the objective under the maximal NA ($NA = 1.4$) (Fig. 10B). The restored images in Figs. 10(C–F) show thereby the increase of information provided by the reconstruction; the previously unresolved F-actin filament structure are fully resolved after image deconvolution (see insets) and artefacts are not generated.

This remark and that at the end of the previous sub-section should be taken into account when deciding what kind of edge-preserving potential can be conveniently used in a practical application.

Concluding remarks

A Bayesian regularization is essential to obtain artefact-free image restoration in 3D confocal microscopy. Quadratic potential function is the most used approach in the Bayesian regularization, but it leads to overstating restoration, with the consequence that edge structures are completely lost. In this paper, we test and compare different edge-preserving potential functions in the MRF framework. Our results show that HS and HB potential functions provide the best reconstructions, both in terms of quantitative KL_D analysis and in terms of qualitative visual inspection. Moreover, we propose the general SGM to devise effective iteration algorithms for maximization of *a posteriori* probability under non-negativity and flux conservation constraints. We believe that such a method thereby helps to test different kinds of regularized potential functions. Slow convergence of the SGM is observed.

However, a modification of this method, known as the SGP, has been recently proposed (Bertero *et al.*, 2008). In this way, one can obtain both a theoretical and a practical improvement of the method. Indeed, it is possible to prove convergence of the modified algorithm, and preliminary numerical results in the case of RLM applied to the deconvolution of 2D images indicate a reduction in the computational time by at least an order of magnitude (Bonettini *et al.*, 2007). The application of this approach to the deconvolution of 3D images in microscopy is in progress.

In this work, we assume that object is piecewise constant. However, we believe that a quadratic piecewise assumption can further improve the results, but this assumption requires higher-order neighbour systems and thus multiple voxel sites, thereby increasing the computational cost of each iteration. For this reason, future work will be addressed not only to test different object models but also to increase the rate of convergence of the SGM along the lines indicated above.

Acknowledgements

We thanks members of the Department of NanoBiophotonics for the support to this work. This work was partially supported by MUR (Italian Ministry for University and Research), grant 2006018748.

References

- Benvenuto, F., Camera, A.L., Theys, C., Ferrari, A., Lanteri, H. & Bertero, M. (2008) The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise. *Inverse Probl.* **24**, 035016.

- Bertero, M. & Boccacci, P. (1998) *Introduction to Inverse Problem in Imaging*. IoP Publishing, Bristol.
- Bertero, M., Lanteri, H. & Zanni, L. (2008) Iterative image reconstruction: a point of view. *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, CRM Series, Vol. 7 (ed. by Y. Censor, M. Jiang & A. Louis), pp. 37–63. Edizioni della Normale, Pisa.
- Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc. B* **36**, 192–236.
- Bonettini, S., Zanella, R. & Zanni, L. (2009) A scaled gradient projection for constrained image deblurring. *Inverse Problems*, **25**, 015002, Department of Mathematics, University of Modena and Reggio Emilia.
- Carrington, W.A. (1990) Image restoration in 3D microscopy with limited data. *Bioimaging and Two-Dimensional Spectroscopy*, Vol. 1205 (ed. by L. C. Smith), pp. 72–83. Society of Photo-Optical Instrumentation Engineers.
- Charbonnier, P., Blanc-Feraud, L., Aubert, G. & Barlaud, M. (1994) Two deterministic half-quadratic regularization algorithms for computed imaging. *Image Processing, 1994. Proceedings. ICIP-94*, Vol. 2, pp. 168–172. Institute of Electrical and Electronics Engineers, Austin, TX, USA.
- Charbonnier, P., Blanc-Feraud, L., Aubert, G. & Barlaud, M. (1997) Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.* **6**, 298–311.
- Conchello, J.A. & McNally, J.G. (1996) Fast regularization technique for expectation maximization algorithm for optical sectioning microscopy. *Three-Dimensional Microscopy: Image Acquisition and Processing III*, Vol. 2655 (ed. C.J. Cogswell, G.S. Kino, G.S. & T. Wilson), pp. 199–208. Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, USA.
- Csiszár, I. (1991) Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.* **19**, 2032–2066.
- Dey, N., Blanc-Feraud, L., Zimmer, C., Roux, P., Kam, Z., Olivo-Marin, J.C. & Zerubia, J. (2006) Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microsc. Res. Tech.* **69**, 260–266.
- Engl, H.W., Hanke, M. & Neubauer, A. (1996) *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht.
- Geman, D. & Reynolds, C. (1992) Constraint restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal.* **14**, 367–383.
- Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal.* **6**, 721–741.
- Geman, S. & McClure, D.E. (1985) Bayesian image analysis: an application to single photon emission tomography. *Proceedings of the Statistical Computation Section*, pp. 12–18. American Statistical Association, Washington, DC.
- Gravier, E. & Yang, Y. (2005) Motion-compensated reconstruction of tomographic image sequences. *IEEE Trans. Nucl. Sci.* **52**, 51–56.
- Green, P.J. (1990) Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imag.* **9**, 84–93.
- Hebert, T. & Leahy, R. (1989) A generalized EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.* **8**, 194–202.
- Holmes, T.J. (1988) Maximum-likelihood image restoration adapted for noncoherent optical imaging. *J. Opt. Soc. Am. A* **5**, 666–673.
- Holmes, T.J. & Liu, Y.-H. (1989) Richardson-Lucy/maximum likelihood image restoration algorithm for fluorescence microscopy: further testing. *Appl. Optics* **28**, 4930–4938.
- Jovin, T.M. & Arndt-Jovin, D.J. (1989) Luminescence digital imaging microscopy. *Annu. Rev. Bioph. Biom.* **18**, 271–308.
- van Kempen, G.M.P. & van Vliet, L.J. (2000a) Background estimation in nonlinear image restoration. *J. Opt. Soc. Am. A* **17**, 425–433.
- van Kempen, G.M.P. & van Vliet, L.J. (2000b). The influence of the regularization parameter and the first estimate on the performance of Tikhonov regularized non-linear image restoration algorithm. *J. Microsc.* **198**, 63–75.
- van Kempen, G.M.P., van Vliet, L.J., Verveer, P.J. & van der Voort, H.T.M. (1997) A quantitative comparison of image restoration methods for confocal microscopy. *J. Microsc.* **185**, 165–181.
- Kullback, S. & Leibler, R.A. (1951) On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86.
- Lange, K. (1990) Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans. Med. Imag.* **9**, 439–446.
- Lanteri, H. & Theys, C. (2005) Restoration of astrophysical images—the case of Poisson data with additive Gaussian noise. *EURASIP J. Appl. Signal Proc.* **15**, 2500–2513.
- Lanteri, H., Roche, M., Cuevas, O. & Aime, C. (2001) A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints. *Signal Process.* **81**, 945–974.
- Lanteri, H., Roche, M. & Aime, C. (2002) Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inverse Probl.* **18**, 1397–1419.
- Lucy, L.B. (1974) An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745–754.
- Mair, B.A. & Zahnen, J. (2006) A generalization of Green's one-step-late algorithm for penalized ML reconstruction of PET images. *Nuclear Science Symposium Conference Record, 2006*, Vol. 5, pp. 2775–2777. Institute of Electrical and Electronics Engineers, San Diego, CA, USA.
- Mondal, P.P., Vicidomini, G. & Diaspro, A. (2007) Markov random field aided Bayesian approach for image reconstruction in confocal microscopy. *J. Appl. Phys.* **102**, 044701.
- Mondal, P.P., Vicidomini, G. & Diaspro, A. (2008) Image reconstruction for multiphoton fluorescence microscopy. *Appl. Phys. Lett.* **92**, 103902.
- Pierro, A.R.D. & Yamagishi, M.E.B. (2001) Fast EM-like methods for maximum a posteriori estimates in emission tomography. *IEEE Trans. Med. Imag.* **20**, 280–288.
- Richards, B. & Wolf, E. (1959) Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system. *Proc. Roy. Soc. Lond. A* **253**, 358–379.
- Richardson, W.H. (1972) Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* **62**, 55–59.
- Rudin, L.I., Osher, S. & Fatemi, E. (1992) Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268.
- Schrader, M., Hell, S.W. & van der Voort, H.T.M. (1998) Three-dimensional super-resolution with a 4Pi-confocal microscope using image restoration. *J. Appl. Phys.* **84**, 4033–4042.
- Schultz, R.R. & Stevenson, R.L. (1995) Stochastic modeling and estimation of multispectral image data. *IEEE Trans. Image Process.* **4**, 1109–1119.
- Shepp, L.A. & Vardi, Y. (1982) Maximum likelihood reconstruction in positron emission tomography. *IEEE Trans. Med. Imag.* **1**, 113–122.
- Snyder, D.L., Hammoud, A.M. & White, R.L. (1993) Image recovery from data acquired with a charge-coupled-device camera. *J. Opt. Soc. Am. A* **10**, 1014–1023.

- Tikhonov, A.N. & Arsenin, V.Y. (1977) *Solutions of Ill-Posed Problems*. Wiley, New York.
- Verbeek, P.W. & van Vliet, L.J. (1994) On the location error of curved edges in low-pass filtered 2D and 3D images. *IEEE Trans. Pattern Anal.* **16**, 726–733.
- Vicidomini, G., Mondal, P.P. & Diaspro, A. (2006) Fuzzy logic and maximum a posteriori-based image restoration for confocal microscopy. *Opt. Lett.* **31**, 3582–3584.
- van der Voort, H.T.M. & Strasters, K.C. (1995) Restoration of confocal images for quantitative image analysis. *J. Microsc.* **178**, 165–181.