A discrepancy principle for Poisson data

2010 Inverse Problems 26 105004

(http://iopscience.iop.org/0266-5611/26/10/105004)

IOP PUBLISHING INVERSE PROBLEMS

Inverse Problems **26** (2010) 105004 (20pp) doi:10.1088/0266-5611/26/10/105004

A discrepancy principle for Poisson data

M Bertero[1], **P Boccacci**[1], **G Talenti**[2], **R Zanella**[3] **and L Zanni**[3]

[1] Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Via Dodecaneso 35, I 16146 Genova, Italy

[2] Dipartimento di Matematica 'Ulisse Dini', Università di Firenze, Viale Morgagni 67A, I 50134 Firenze, Italy

[3] Dipartimento di Matematica Pura e Applicata, Università di Modena e Reggio Emilia, Via G Campi 213/b, I 41100 Modena, Italy

E-mail: bertero@disi.unige.it

Received 27 January 2010, in final form 13 July 2010
Published 26 August 2010
Online at stacks.iop.org/IP/26/105004

Abstract

In applications of imaging science, such as emission tomography, fluorescence microscopy and optical/infrared astronomy, image intensity is measured via the counting of incident particles (photons, γ-rays, etc). Fluctuations in the emission-counting process can be described by modeling the data as realizations of Poisson random variables (Poisson data). A maximum-likelihood approach for image reconstruction from Poisson data was proposed in the mid-1980s. Since the consequent maximization problem is, in general, ill-conditioned, various kinds of regularizations were introduced in the framework of the so-called Bayesian paradigm. A modification of the well-known Tikhonov regularization strategy results in the data-fidelity function being a generalized Kullback–Leibler divergence. Then a relevant issue is to find rules for selecting a proper value of the regularization parameter. In this paper we propose a criterion, nicknamed discrepancy principle for Poisson data, that applies to both denoising and deblurring problems and fits quite naturally the statistical properties of the data. The main purpose of the paper is to establish conditions, on the data and the imaging matrix, ensuring that the proposed criterion does actually provide a unique value of the regularization parameter for various classes of regularization functions. A few numerical experiments are performed to demonstrate its effectiveness. More extensive numerical analysis and comparison with other proposed criteria will be the object of future work.

1. Introduction

Poisson data typically occur in those imaging processes where images are obtained by counting the particles, in general photons, that hit the image domain. In the case of radioactive decay, fluorescence emission or similar phenomena, the number of particles counted by a given detector element of area A, within a given time interval T, is a random variable with a *Poisson distribution* [13, 32], i.e. the probability of receiving n particles is given by

$$p(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \qquad n = 0, 1, 2, \ldots, \tag{1}$$

where λ equals the expected value of the counts and is proportional to A and T. This statistical model is appropriate for describing data in e.g. fluorescence microscopy, emission tomography, optical/infrared astronomy, etc. Even if the wavelength of the photons is different in different applications, the statistics of the data is identical.

Just for completeness we recall that in the case of images acquired with a CCD camera, as in astronomy or some modalities of microscopy, a more refined model taking into account both Poisson noise and read-out noise (modeled as a Gaussian process) has been introduced [33]. In such a case the likelihood has a more complex form and an EM algorithm has been proposed for its maximization [33]. However this refinement seems to be useful only in very particular circumstances.

Research on image reconstruction methods, based on the statistical property (1) of the noise, started in the mid-1980s with the seminal papers of Shepp and Vardi [31] and Geman and Geman [14]. This research applies both to denoising and deblurring problems. A maximum-likelihood (ML) approach is proposed in [31], while a related regularization strategy, based on the Bayesian paradigm, is proposed in [14]. A large body of literature exists on this topic, mainly driven by research on medical imaging. A recent application to image denoising is investigated in [9, 20, 36], while a review of the applications to image deblurring is given in [5].

The previous approach leads in many instances to the minimization of a function which involves a free parameter mimicking a Tikhonov-like regularization parameter [12]. The main purpose of this paper is to investigate a principle for the choice of this parameter. Such a principle, proposed in [36] for denoising, is extended to deblurring and is based on the Kullback–Leibler (KL) divergence between computed and detected data. We believe that it is an improvement of a similar discrepancy principle based on a quadratic approximation of KL [3]: it enables a theoretical analysis and preliminary numerical results indicate that it may provide more reliable estimates.

In section 2 we sketch a scenario in which the denoising/deblurring of Poisson data is portrayed as a convex optimization problem. The objective function is the sum of two terms: the former is a data-fidelity function, eligible for dealing with Poisson data; the latter is a penalty function, serving regularization purposes. This last term is multiplied by a free parameter, which we denote by β and call *the regularization parameter*. At the end of the section we formulate a lemma (proved in the appendix) characterizing the dependence on β of the two terms evaluated at the minimizer of the objective function. In section 3 we motivate and formulate our principle for determining β, which we call the *discrepancy principle*. It implies solving a transcendental equation called the *discrepancy equation*. We briefly discuss the relationship between the lemma of section 2 and the solution of this equation. In section 4 we prove uniqueness of the solution of the discrepancy equation in the one-dimensional case for a number of penalty functions: Tikhonov penalty in terms of the ℓ_2 norm of the object or of its gradient and edge-preserving penalty. In section 5 we investigate the existence of the solution under the conditions of section 4. In section 6 we consider the multi-dimensional case. For simplicity, we introduce some restrictive conditions on the data and on the imaging matrix ensuring that the objective function is strictly convex, so that results on the discrepancy equation can be easily derived. However, these conditions can be actually removed [7]. Finally in section 8 we test the effectiveness of our criterion by means of a few numerical simulations. As far as deblurring is concerned, we show that the criterion can be used for stopping relevant iterations in the case where the *expectation maximization* (EM) [31] algorithm or one of its accelerated version, such as the *scaled gradient projection* (SGP) method [6], is used.

2. Formulation of the problem

We denote the data as $y = \{y_i\}_{i \in S}$, where i is an index (in the case of spectra measurements, for instance) or a multi-index (in the case of 2D or 3D images), and S is an appropriate range of i. We denote the cardinality of S by M. Each y_i is a nonnegative integer and realizes a Poisson random variable Y_i with (unknown) the expected value $\tilde{y}_i$. Y_i are statistically independent, so their joint probability distribution is the product of their Poisson distributions. We set $\tilde{y} = \{\tilde{y}_i\}_{i \in S}$ and $Y = \{Y_i\}_{i \in S}$.

The data y are the image of some unknown object $x = \{x_j\}_{j \in R}$; in image deblurring, image and object may have different dimensions, and in such a case we denote the cardinality of R by N. In the Bayesian paradigm x is also assumed to realize a multi-valued random variable $X = \{X_j\}_{j \in R}$ with (unknown) the expected value $\tilde{x} = \{\tilde{x}_j\}_{j \in R}$. In the case of denoising the relationship between image and object is simply $Y = X$; hence, $\tilde{y} = \tilde{x}$. In the case of deblurring the relationship is more complex. If the distortions introduced by the imaging system can be described by a linear relationship, then $Y = HX$, where H is an imaging matrix satisfying the following conditions:

$$H_{i,j} \geqslant 0, \qquad \sum_{i \in S} H_{i,j} > 0, \qquad \forall j \in R, \qquad \sum_{j \in R} H_{i,j} > 0, \qquad \forall i \in S. \tag{2}$$

In other words, for each fixed value of the index or multi-index i or j, there exists at least one non-zero entry.

Remark 1. The previous assumptions imply that if the null space of the matrix H is nontrivial, then all its elements must have at least one negative component. It follows that $||Hx||$ *is coercive on the nonnegative orthant*. Indeed, it cannot be zero on the intersection of the nonnegative orthant with the unit sphere and therefore has a positive minimum α. Since it is homogeneous of order 1 it follows $||Hx|| \geqslant \alpha ||x||$, $x \geqslant 0$.

In the following it is convenient to assume that the normalization conditions

$$\sum_{i \in S} H_{i,j} = 1, \qquad \forall j \in R \tag{3}$$

are satisfied. As a consequence we have $H^T e_M = e_N$, where e_M, e_N are the vectors (or arrays or stacks of arrays) in $\mathbb{R}^M$ and $\mathbb{R}^N$, respectively, with all entries equal to one.

Remark 2. As stated in the seminal paper of Shepp and Vardi [31], these conditions do not involve a loss of generality of the results obtained by means of them, but make their interpretation simpler. For instance, if we set $y = Hx$, then they imply equality of the ℓ_1 norms of x and y and the interpretation of this relationship is that 'the total number of photons is the same in the original object and in the image', or also, that 'the object and the image have the same total flux'.

If the imaging matrix does not satisfy these conditions, then they can be introduced by a renormalization of the 'columns' of the matrix. However this procedure may not be convenient in practice when designing a reconstruction algorithm. For instance, if the matrix has a Toeplitz structure, then this is destroyed by the renormalization. On the other hand, if the imaging matrix is approximated by the cyclic convolution of the object with a periodic and nonnegative *point-spread function* (PSF) $K = \{K_i\}_{i\in S}$, i.e. $Hx = K * x$, then the normalization conditions are satisfied if the PSF is normalized as follows:

$$\sum_{i\in S} K_i = 1, \tag{4}$$

a condition required in the everyday use of deconvolution in astronomy and microscopy. In such a case we also have $He_N = e_N$.

In addition to the object emission, very often there also exists a background emission, described by a Poisson process with a constant expected value, i.e. independent of the index i, and known. We denote this value by b and we assume that b is strictly positive, even if it can be quite small. The assumption of a constant background may be unrealistic in many applications, but it greatly simplifies the analysis without invalidating the significance of our main purposes. In conclusion, if we denote by B the multi-valued random variable whose expected value is be_M, then the complete relationship between image and object is given by $Y = HX + B$.

According to the previous assumptions, for a given y the likelihood function of the problem is defined by

$$\mathcal{L}_y^Y(x) = \prod_{i\in S} \frac{\mathrm{e}^{-(Hx+be_M)_i}(Hx + be_M)_i^{y_i}}{y_i!}, \tag{5}$$

and a ML estimate of $\tilde{x}$ is any maximizer x^* of $\mathcal{L}_y^Y(x)$.

By taking the negative logarithm (neglog) of the likelihood function and readjusting terms independent of x, one easily finds that maximizing $\mathcal{L}_y^Y(x)$ is equivalent to minimizing the following generalized KL divergence:

$$f_0(x) = \sum_{i\in S}\left\{ y_i \ln\frac{y_i}{(Hx + be_M)_i} + (Hx + be_M)_i - y_i \right\}, \tag{6}$$

which will be thought of as the *data-fidelity function*. Of course, the case of denoising is obtained by setting $H = I$ and $b = 0$. This function is nonnegative, convex and coercive on the nonnegative orthant. The last property is obvious in the case of denoising and follows from remark 1 in the case of deblurring.

A priori information on the solution is introduced in the Bayesian approach by assuming that the probability distribution of the random variable X is also known. A very frequent assumption is that X is a Markov random field or, equivalently, a Gibbs random field [14] whose distribution, usually called the *prior*, obeys

$$p_X(x) = \frac{1}{Z}\,\mathrm{e}^{-\beta f_1(x)}. \tag{7}$$

Here Z is a normalization constant, and $f_1(x)$ is given and plays the role of a *penalty*.

Once the prior has been chosen, the determination of the *maximum a posteriori* (MAP) estimate of $\tilde{x}$ is equivalent to maximizing the so-called *posterior probability distribution* of X, given by

$$\mathcal{P}_y^X(x) = \frac{\mathcal{L}_y^Y(x)\,p_X(x)}{p_Y(y)}. \tag{8}$$

Taking the neglog, readjusting terms and considering that only nonnegative solutions are relevant, entails that a tractable MAP estimate is a solution to the following constrained minimization problem:

$$\begin{aligned} &\text{minimize} \quad f_\beta(x) = f_0(x) + \beta f_1(x) \\ &\text{subject to} \quad x \in \Omega, \end{aligned} \tag{9}$$

where Ω is a closed and convex subset of the nonnegative orthant to be appropriately specified from time to time. We denote throughout by x_β^* any solution to this problem.

We assume that the penalty function $f_1(x)$ is nonnegative and convex, conditions that are in general satisfied in practical applications. In addition we assume that it is also differentiable. Then, the following lemma is proved in the appendix.

Lemma 1. *If the function $f_\beta(x)$ is strictly convex, coercive and differentiable on Ω, then $f_0(x_\beta^*)$ and $f_1(x_\beta^*)$ are respectively an increasing and a decreasing function of β.*

3. The discrepancy principle

In the case of data perturbed by additive Gaussian noise, i.e. in the case of Tikhonov regularization, the data-fidelity function $f_0(x)$ is the square of the metric distance between detected and computed data, and the choice of the regularization parameter is claimed by the well-known *Morozov discrepancy principle* [12]. For the deblurring problem, the value of β is given by the unique solution of the equation

$$\|Hx_\beta + be_M - y\|^2 = \sigma^2 M, \tag{10}$$

where x_β is the Tikhonov regularized solution and σ^2 is the variance of the noise. This recipe is frequently used in practice even if it can over-smooth the solution [15]. For this reason criteria based on risk minimization, derived from the assumption of Gaussian noise [26] or of a mixture of Gaussian and Poisson noise [27], have been proposed.

In the case of Poisson data, a counterpart of Morozov principle, based on the data-fidelity function (6), is proposed in [36] for the denoising problem and rests upon the following result, also proved in [36].

Lemma 2. *Let Y_λ be a Poisson random variable with expected value λ and consider the function of Y_λ defined by*

$$F(Y_\lambda) = 2\left\{Y_\lambda \ln\left(\frac{Y_\lambda}{\lambda}\right) + \lambda - Y_\lambda\right\}. \tag{11}$$

Then the following estimate of the expected value of $F(Y_\lambda)$ holds true for large λ:

$$E\left\{F(Y_\lambda)\right\} = 1 + O\left(\frac{1}{\lambda}\right). \tag{12}$$

Replacing x by $\tilde{x}$ in equation (6), and letting the components of $\tilde{x}$ be sufficiently large, makes each term the realization of a random variable whose expected value is approximately 1/2. Consequently these terms fluctuate around 1/2 and the positive fluctuations may compensate the negative ones. In other words we expect that the value of the sum, for $x = \tilde{x}$, is just 1/2 multiplied by the cardinality of S.

Motivated by the above arguments, we suggest the following recipe for both denoising and deblurring: search for a value of β such that the quantity defined by

$$D_y(\beta) = \frac{2}{M} f_0(x_\beta^*) \tag{13}$$

obeys

$$D_y(\beta) = 1. \tag{14}$$

If x_β^* is unique then $D_y(\beta)$ is a well-defined function. Therefore uniqueness is one of the key issues investigated in the following. We call $D_y(\beta)$ the *discrepancy function* for Poisson data and call equation (14) the *discrepancy equation*.

Remark 3. The choice of the value 1 in (14) may not be 'optimal'. In general, one should replace (14) by $D_y(\beta) = 1+\epsilon$ where ϵ is a small positive or negative number. The discussion of the choice of ϵ is beyond the scope of this paper and, presumably, a range of possible values can be derived from statistical results more accurate than that contained in lemma 2. We only point out that our results on uniqueness and existence of the solution of the discrepancy equation can be easily extended to the case where 1 is replaced by a positive constant greater or smaller than 1. In particular, lemma 1 implies that $D_y(\beta)$ is an increasing function of β when it is well defined. Therefore, the equation $D_y(\beta) = 1+\epsilon$ will provide a value of β which is smaller than that provided by equation (14) if $\epsilon < 0$ and greater if $\epsilon > 0$.

A criterion similar to (14) is proposed in a recent paper [3], in connection with a quadratic approximation to the generalized KL divergence. In our notation and in the case of image reconstruction it is based on the following discrepancy function:

$$\bar{D}_y(\beta) = \frac{1}{M}\left\|\frac{Hx^*_\beta + be_M - y}{\sqrt{Hx^*_\beta + be_M}}\right\|^2, \tag{15}$$

where the quotient of vectors (or arrays or stacks of arrays) is in the Hadamard sense, i.e. component-wise. This notation will also be used in the following. We also point out that the condition $\bar{D}_y(\beta) \simeq 1$ is used in [18] for defining a set of objects that are 'weakly feasible' with respect to the detected image.

4. Uniqueness of the solution: one-dimensional problems

In this section we first consider 1D problems because in this case we can prove that x^*_β is unique, both for denoising and deblurring with three relevant regularizers, without additional assumptions on the data and the imaging matrix. As a consequence the discrepancy function is well defined. The discussion of 2D an 3D cases is deferred to section 6. First we give the properties of the data-fidelity functions $f_0(x)$ and subsequently those of the penalty functions $f_1(x)$ we are considering.

Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be the index sets defined by

$$\mathcal{I}_1 = \{i \in S | y_i > 0\}, \qquad \mathcal{I}_2 = \{i \in S | y_i = 0\} = S \setminus \mathcal{I}_1. \tag{16}$$

- *Denoising problem*

In this case $S = R = \{1, 2, \ldots, N\}$, i.e. the index i takes values from 1 to N. Then the problem can be formulated as in (9), with $f_0(x)$ given by

$$f_0(x) = \sum_{i=1}^{N}\left\{y_i \ln\frac{y_i}{x_i} + x_i - y_i\right\}, \tag{17}$$

($y_i \ln y_i = 0$ if $y_i = 0$) and

$$\Omega = \{x \in \mathbb{R}^N | x_i \geqslant \eta, i \in \mathcal{I}_1; x_i \geqslant 0, i \in \mathcal{I}_2\}, \tag{18}$$

where η is any constant such that $0 < \eta < \min\{y_i, i \in \mathcal{I}_1\}$. Any η between 0 and 1 works well (for instance, 10^{-5} as in [36]) because $\min\{y_i, i \in \mathcal{I}_1\} \geqslant 1$.

The function $f_0(x)$ is strictly convex if and only if $y > 0$. It has a unique nonnegative minimizer given by $x^* = y$, with $f_0(x^*) = 0$; thus, the ML estimate is trivial. The gradient and the Hessian are given by

$$\nabla f_0(x) = e_N - \frac{y}{x}, \qquad \nabla^2 f_0(x) = \mathrm{diag}\left(\frac{y}{x^2}\right). \tag{19}$$

If some components of y are zero, then the null space of the Hessian is

$$\mathcal{N}[\nabla^2 f_0(x)] = \{u \in \mathbb{R}^N | u_i = 0, i \in \mathcal{I}_1\}, \tag{20}$$

and therefore is not trivial.

- *Deblurring problem*

In this case $S = \{1, 2, \ldots, M\}$, $R = \{1, 2, \ldots, N\}$ and $M = N$ as far as image deconvolution is dealt with. Then, the data-fidelity function is given by

$$f_0(x) = \sum_{i=1}^{M} \left\{ y_i \ln \frac{y_i}{(Hx + be_M)_i} + (Hx + be_M)_i - y_i \right\}. \tag{21}$$

Its domain equals $\mathcal{D}_0 = \{x \in \mathbb{R}^N | (Hx+be_M)_i > 0, i \in \mathcal{I}_1\}$ and is broader than the nonnegative orthant; in this case, the feasible set is just

$$\Omega = \{x \in \mathbb{R}^N | x \geqslant 0\}. \tag{22}$$

The function $f_0(x)$ is nonnegative, coercive and convex on the nonnegative orthant where it has at least one minimizer x^*. Its gradient and Hessian are given by

$$\nabla f_0(x) = e_N - H^T \frac{y}{Hx + be_M}, \qquad \nabla^2 f_0(x) = H^T \frac{y}{(Hx + be_M)^2} H. \tag{23}$$

If $y > 0$ and $\mathcal{N}(H) = \{0\}$, $f_0(x)$ is strictly convex and the constrained minimizer x^* is unique. In the event that some components of y vanish, the null space of the Hessian, which obeys

$$\mathcal{N}[\nabla^2 f_0(x)] = \{u \in \mathbb{R}^N | (Hu)_i = 0, i \in \mathcal{I}_1\}, \tag{24}$$

contains the null space of H, and might be nontrivial even if $\mathcal{N}(H) = \{0\}$.

The constrained minimizers of $f_0(x)$ belong, in general, to the boundary of the nonnegative orthant and do not provide sensible solutions of the deblurring problem. Sometimes they are called *night-sky* [4] or also *checker-board reconstructions* [23]. The next lemma provides a sufficient condition ensuring that the minimizer is not trivial.

Lemma 3. *The null element 0 is not a minimizer of $f_0(x)$ if*

$$\frac{1}{N} \sum_{j=1}^{N} (H^T y)_j > b. \tag{25}$$

Proof. We consider the restriction of $f_0(x)$ to the set of vectors parallel to e_N, and let

$$\psi(c) = f_0(ce_N). \tag{26}$$

Using the normalization condition (3), manipulation shows

$$\psi'(c) = N - \sum_{i \in \mathcal{I}_1} \frac{\mathcal{H}_i y_i}{\mathcal{H}_i c + b}, \qquad \mathcal{H}_i = \sum_{j=1}^{N} H_{i,j}. \tag{27}$$

Function ψ has a positive minimizer if $\psi'(0) < 0$, and this is just inequality (25). □

Remark 4. Condition (25) has a simple interpretation. Indeed, the vector $H^T y$ is the back-projection of the data and the condition requires that its average value is greater than the background. If a non-zero signal is superimposed to the background then this condition must be satisfied. In the case of image deconvolution, from condition (4) we obtain $\mathcal{H}_i = 1$, so equation (25) implies

$$\bar{y} = \frac{1}{N} \sum_{i \in \mathcal{I}_1} y_i > b, \tag{28}$$

where $\bar{y}$ is the average of the detected values.

Next we consider the following penalty functions.

- *Tikhonov regularization of order 0.*

  In this case we have

$$f_1(x) = \frac{1}{2}\sum_{i=1}^{N} x_i^2 \tag{29}$$

and this function is nonnegative, coercive and strictly convex.

- *Tikhonov regularization of order 1.*

  Tikhonov regularization in terms of the ℓ_2 norm of the discrete gradient leads to

$$f_1(x) = \frac{1}{2}\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2; \tag{30}$$

this function is nonnegative and convex and its minimizers are the constant vectors. The components of its gradient are given by

$$\begin{aligned} &[\nabla f_1(x)]_1 = -x_2 + x_1, \qquad [\nabla f_1(x)]_N = x_N - x_{N-1} \\ &[\nabla f_1(x)]_i = 2x_i - x_{i-1} - x_{i+1}, \qquad i = 2, \ldots, N-1, \end{aligned} \tag{31}$$

and satisfy the following identity:

$$\sum_{i=1}^{N} [\nabla f_1(x)]_i = 0. \tag{32}$$

The Hessian is the symmetric matrix whose non-zero elements from the upper triangle are given by

$$\begin{aligned} &[\nabla^2 f_1(x)]_{1,1} = [\nabla^2 f_1(x)]_{N,N} = 1, \\ &[\nabla^2 f_1(x)]_{i,i} = 2, \qquad i = 2, \ldots, N-1, \\ &[\nabla^2 f_1(x)]_{i-1,i} = -1, \qquad i = 2, \ldots, N. \end{aligned} \tag{33}$$

It is positive semidefinite and its null space is the set of the constant vectors

$$\mathcal{N}[\nabla^2 f_1(x)] = \{u \in \mathbb{R}^N | u = c e_N\}. \tag{34}$$

- *Edge-preserving regularization.*

  A class of edge-preserving potentials for 2D image denoising is considered in [36]. In the 1D case they lead to the following function:

$$f_1(x) = \sum_{i=1}^{N-1} \sqrt{(x_{i+1} - x_i)^2 + \delta^2}, \tag{35}$$

where δ is a 'scaling' or 'thresholding' parameter, tuning the 'jumps' in the values of the discrete signal. This potential function, proposed in [10] and called *hypersurface regularization*, is convex and its minimizers are the constant vectors. For very small values of δ it provides a numerical approximation of the *total variation* (TV) [35]. Otherwise it is an approximation of the length of the graph of a function $x(t)$ with samples $\{x_i\}_{i=1}^N$ and sampling distance δ.

If we let

$$\Delta_i = (x_{i+1} - x_i)^2 + \delta^2, \qquad h_i = \frac{\delta^2}{\Delta_i^{3/2}}, \qquad i = 1, \ldots, N-1, \tag{36}$$

then the gradient of $f_1(x)$ is given by

$$
\begin{aligned}
&[\nabla f_1(x)]_1 = -\frac{x_2 - x_1}{\Delta_1^{1/2}}, \qquad [\nabla f_1(x)]_N = \frac{x_N - x_{N-1}}{\Delta_{N-1}^{1/2}} \\
&[\nabla f_1(x)]_i = \frac{x_i - x_{i-1}}{\Delta_{i-1}^{1/2}} - \frac{x_{i+1} - x_i}{\Delta_i^{1/2}}, \qquad i = 2, \ldots, N-1,
\end{aligned}
\tag{37}
$$

and satisfies identity (32). The Hessian is symmetric, and its nonzero entries from the upper triangle are

$$
\begin{aligned}
&[\nabla^2 f_1(x)]_{1,1} = h_1, \qquad [\nabla^2 f_1(x)]_{N,N} = h_{N-1}, \\
&[\nabla^2 f_1(x)]_{i,i} = h_{i-1} + h_i, \qquad i = 2, \ldots, N-1, \\
&[\nabla^2 f_1(x)]_{i-1,i} = -h_{i-1}, \qquad i = 2, \ldots, N.
\end{aligned}
\tag{38}
$$

Elementary computations show

$$
(\nabla^2 f_1(x) u, u) = \sum_{i=1}^{N-1} h_i (u_i - u_{i+1})^2, \tag{39}
$$

for any $u \in R^n$. Therefore the Hessian of $f_1(x)$ is positive semidefinite and its null space contains only the constant vectors (see equation (34)).

Remark 5. Penalties (30) and (35) can be obtained by extending the sum up to N and setting $x_{N+1} = x_N$, i.e. by assuming a Neumann boundary condition. We remark that equations (32) and (34) still hold true if we assume 'periodic' boundary conditions, i.e. $x_{N+1} = x_1$, as we did in [36], even if the expressions of the gradient and Hessian are slightly different.

Theorem 1. *The function $f_\beta(x)$ is strictly convex in the case of denoising and deblurring with anyone of the penalty functions discussed above. Therefore x^*_β is unique and the discrepancy function $D_y(\beta)$ is well defined.*

Proof. The strict convexity of $f_\beta(x)$ is obvious in the case of Tikhonov regularization of order 0. In the case of the two other penalty functions we first remark that, since $f_0(x)$ is nonnegative, convex and coercive and $f_1(x)$ is nonnegative and convex, then also $f_\beta(x)$ is nonnegative, convex and coercive. Therefore, the result follows if the null space of the Hessian of $f_\beta(x)$ is $\{0\}$. But, since all the relevant Hessians are positive semidefinite, this null space is just the intersection of the null spaces of $f_0(x)$ and $f_1(x)$ and, from equations (20), (24) and (34), it follows that it is just $\{0\}$. □

Theorem 1 and lemma 1 imply that the corresponding discrepancy functions are increasing. Therefore the following result is obvious.

Theorem 2. *For the denoising and deblurring problems considered above, the solution of the discrepancy equation (14), if it exists, is unique.*

5. Existence of the solution: one-dimensional problems

In order to establish conditions for the existence of the solution we consider separately the different penalties.

5.1. Tikhonov regularization of order 0

The analysis of the corresponding denoising problem is elementary because the explicit form of the unique minimizer can be computed by solving a quadratic algebraic equation. However we briefly discuss it because the solution of this problem is the M-step in the EM-algorithm proposed in [11] for image deblurring with Poisson data and Tikhonov regularization. It is given by

$$[x^*_\beta]_i = \max\left\{-\frac{1}{2\beta} + \frac{1}{2\beta}\sqrt{1+4\beta y_i}, \eta\right\}, \tag{40}$$

if $i \in \mathcal{I}_1$, while $[x^*_\beta]_i = 0$ if $i \in \mathcal{I}_2$.

The limit of x^*_β as $\beta \to +\infty$ is the vector x_η defined by $(x_\eta)_i = \eta$ if $i \in \mathcal{I}_1$, and $(x_\eta)_i = 0$ if $i \in \mathcal{I}_2$. Consequently, a solution of the discrepancy equation exists if and only if

$$\frac{1}{N}\sum_{i\in\mathcal{I}_1}\left(y_i \ln\frac{y_i}{\eta} + \eta - y_i\right) > \frac{1}{2}. \tag{41}$$

This condition is satisfied if we take η sufficiently small and therefore there always exists a unique solution of the discrepancy equation (14).

In the case of deblurring with this regularization (see, for instance, [1, 11, 22]), we remark that $x^*_\beta \to x^*$ as $\beta \to 0$, where x^* is a minimizer of $f_0(x)$ while $x^*_\beta \to 0$ as $\beta \to \infty$. Therefore the discrepancy principle provides a solution if and only if the following conditions are satisfied

$$f_0(x^*) < \frac{M}{2}, \qquad f_0(0) > \frac{M}{2}. \tag{42}$$

The latter condition is satisfied if

$$\frac{1}{M}\sum_{i\in\mathcal{I}_1} y_i \ln y_i > \frac{1}{2} + (\bar{y} - b) + \bar{y}\ln b. \tag{43}$$

5.2. Tikhonov regularization of order 1

Lemma 4. *Let $\bar{y}$ be the average value of the components of y, equation (28). In the case of image denoising if $\bar{y} > \eta$, then we have the following limits:*

$$\lim_{\beta\to 0} x^*_\beta = y, \qquad \lim_{\beta\to+\infty} x^*_\beta = \bar{y}e_N. \tag{44}$$

Moreover, in the case of image deblurring if condition (25) is satisfied, then the following limits hold true:

$$\lim_{\beta\to 0} x^*_\beta = x^*, \qquad \lim_{\beta\to+\infty} x^*_\beta = \bar{c}e_N, \tag{45}$$

where x^* is a minimizer of (21) and $\bar{c}$ is the unique positive solution of the equation

$$\sum_{i\in\mathcal{I}_1}\frac{\mathcal{H}_i y_i}{\mathcal{H}_i\bar{c} + b} = N. \tag{46}$$

In the case of image deconvolution ($M = N$, $\mathcal{H}_i = 1$) the solution is $\bar{c} = \bar{y} - b$.

Proof. Consider first image denoising. Since x^*_β depends continuously on β, the limit of x^*_β for $\beta \to 0$ is the minimizer of the generalized KL divergence (17), i.e. y. On the other hand, the limit for $\beta \to +\infty$ is a minimizer of $f_1(x)$, hence a constant vector ce_N. The constant is greater than η. Indeed, from the Karush–Kuhn–Tucker (KKT) condition for x^*_β and equations (32) and (19) we have

$$0 \leqslant \sum_{i=1}^{N} [\nabla f_\beta(x^*_\beta)]_i = \sum_{i=1}^{N} [\nabla f_0(x^*_\beta)]_i = N - \sum_{i\in\mathcal{I}_1} \frac{y_i}{[x^*_\beta]_i}, \tag{47}$$

and therefore, in the limit $\beta \to +\infty$

$$c \geqslant \bar{y} > \eta. \tag{48}$$

As a consequence, x^*_β is interior to Ω for β sufficiently large and the equality holds true in equation (47). The limit ce_N must satisfy this same condition, so that $c = \bar{y}$.

In the case of deblurring the limit for $\beta \to 0$ is obvious. We only remark that if $f_0(x)$ has more than one minimizer, then the regularization selects one of them. As far as the latter limit is concerned, KKT conditions and equations (32) and (23) imply that x^*_β satisfies the inequality

$$0 \leqslant \sum_{j=1}^{N} [\nabla f_\beta(x^*_\beta)]_j = \sum_{j=1}^{N} [\nabla f_0(x^*_\beta)]_j = N - \sum_{j=1}^{N} \left(H^T \frac{y}{Hx^*_\beta + b} \right)_j. \tag{49}$$

Since the limit for $\beta \to \infty$ is a constant vector $\bar{c}e_N$, from the previous equation we have

$$0 \leqslant N - \sum_{i\in\mathcal{I}_1} \frac{\mathcal{H}_i y_i}{\mathcal{H}_i \bar{c} + b}. \tag{50}$$

Thanks to condition (25) this inequality is not satisfied by $\bar{c} = 0$, while it is satisfied by sufficiently large values of $\bar{c}$. It follows that $\bar{c} > 0$. But if the constant is positive, x^*_β is interior to the nonnegative orthant for β sufficiently large, and the equality holds true in equation (49). It follows that $\bar{c}$ is the unique solution of equation (46), and the lemma is proved. □

Theorem 3. *Under the assumptions of lemma 4, in the case of denoising the discrepancy equation (14) has a solution if and only if the data satisfy the following condition:*

$$\frac{1}{N} \sum_{i\in\mathcal{I}_1} y_i \ln y_i > \frac{1}{2} + \bar{y} \ln \bar{y}, \tag{51}$$

while in the case of deblurring the solution exists if and only if

$$f_0(x^*) < \frac{M}{2}, \qquad f_0(\bar{c}e_N) > \frac{M}{2}, \tag{52}$$

where x^* is the relevant nonnegative minimizer of $f_0(x)$ and $\bar{c}$ is the unique solution of equation (46). In the case of image deconvolution ($\mathcal{H}_i = 1$, $M = N$) the latter condition in (52) amounts to (51).

Proof. In the case of denoising, thanks to lemma 4 we have $D_y(0) = 0$ and

$$\lim_{\beta\to+\infty} D_y(\beta) = \frac{2}{N} f_0(\bar{y}e_N) = \frac{2}{N} \sum_{i\in\mathcal{I}_1} y_i \ln y_i - 2\bar{y} \ln \bar{y}. \tag{53}$$

It follows that the function $D_y(\beta)$ can take the value 1 if and only if this limit is greater than 1 and this is just condition (51).

In the case of deblurring, since $D_y(0) = 2f_0(x^*)/M$, the former condition (52) is obvious as well as the latter condition because the limit of $D_y(\beta)$ is $2f_0(\bar{c}e_N)/M$ as $\beta \to +\infty$. In the case of image deconvolution, equations (21) and (4) yield ($M = N$)

$$f_0(\bar{c}e_N) = f_0[(\bar{y} - b)e_N] = \sum_{i\in\mathcal{I}_1} y_i \ln y_i - N\bar{y} \ln \bar{y} > \frac{N}{2}, \tag{54}$$

and this condition coincides with that required in the case of denoising. □

5.3. Edge-preserving regularization

The analysis performed in the case of Tikhonov regularization of order 1 is based on lemma 4. The main ingredient in the proof of this lemma is identity (32). Since it holds true also in the case of edge-preserving regularization, we can conclude that lemma 4 and theorem 3 apply also to this case.

6. Multi-dimensional problems

We consider the case where the data depend on a multi-index i, that is, a pair of indices $\{i_1, i_2\}$ in the case of 2D images, and a triple of indices $\{i_1, i_2, i_3\}$ in the case of 3D images. This indexing of the data is convenient when FFT is used for matrix computation.

The data-fidelity function $f_0(x)$ for both denoising and deblurring is given by equation (6). The gradients and Hessians of these two functions can be written as in equations (19) and (23), and the null spaces of the Hessians are given by equations (20) and (24), provided that now i is a multi-index and R is the range of i.

The analysis is greatly simplified if we make the following additional assumptions:

- the data are strictly positive, $y > 0$, and
- $\mathcal{N}(H) = \{0\}$.

Indeed, if both assumptions are satisfied, then the null space of the Hessian of $f_0(x)$ equals $\{0\}$, both for denoising and deblurring; the data-fidelity function is strictly convex in both cases and lemma 1 applies, providing the following general result.

Theorem 4. *For any nonnegative, convex and differentiable penalty function $f_1(x)$, the solution of equation (14), if it exists, is unique. This result also holds true when 1 is replaced by another positive constant in the rhs of the equation.*

The former assumption above is certainly satisfied when a sufficiently large background contributes to the detected images (as in some astronomical images), while the latter assumption is likely satisfied in most practical applications of image deconvolution, and is fulfilled in the numerical simulations of the subsequent section. However, in applications to microscopy or medical imaging, the number of counts can be small and therefore components of y (sometimes, several components) can be zero. Again in the case of medical imaging the matrix H is sparse and not square, so that it may have a nontrivial null space. For these reasons it is important to remove the previous conditions, as we did in the 1D case, for some relevant regularizations of $f_0(x)$. The extension of the results of section 4 to the multi-dimensional case is not easy and is done in [7].

As concerns the existence of the solution of the discrepancy equation, it is easy to extend the results of section 5 to the case of the following penalty functions.

- Tikhonov regularization of order 0

$$f_1(x) = \frac{1}{2}\sum_{i\in S} x_i^2, \tag{55}$$

- Tikhonov regularization of order 1

$$f_1(x) = \frac{1}{2}\sum_{i\in S}(\Delta_i - \delta^2), \tag{56}$$

- edge-preserving regularization

$$f_1(x) = \sum_{i\in S} \sqrt{\Delta_i}, \tag{57}$$

where Δ_i is defined by

$$\Delta_i = \left(x_{i_1+1,i_2} - x_{i_1,i_2}\right)^2 + \left(x_{i_1,i_2+1} - x_{i_1,i_2}\right)^2 + \delta^2 \tag{58}$$

in the 2D case, and

$$\Delta_i = \left(x_{i_1+1,i_2,i_3} - x_{i_1,i_2,i_3}\right)^2 + \left(x_{i_1,i_2+1,i_3} - x_{i_1,i_2,i_3}\right)^2 + \left(x_{i_1,i_2,i_3+1} - x_{i_1,i_2,i_3}\right)^2 + \delta^2 \tag{59}$$

in the 3D case (in both cases, the Δ_i can be defined for all values of $i \in S$ via some extension of x—see remark 5).

For penalty (55), the denoising problem is trivial as in the 1D case. Moreover, $x^*_\beta \to x_\eta$ (see section 5) for $\beta \to +\infty$ and therefore the discrepancy principle provides a solution if and only if $f_0(x_\eta) > N/2$. As in the 1D problem this condition is always satisfied if η is sufficiently small. In the case of the deblurring problem, $x^*_\beta \to x^*$ for $\beta \to 0$, where x^* is the unique minimizer of $f_0(x)$, and $x^*_\beta \to 0$ for $\beta \to \infty$. Therefore the discrepancy equation provides a solution if and only if the two conditions $f_0(x^*) < M/2$ and $f_0(0) > M/2$ are satisfied. equation (6) implies that the latter inequality is true if the data satisfy the condition

$$\frac{1}{M}\sum_{i\in S} y_i \ln y_i > \frac{1}{2} + (\bar{y} - b) + \bar{y}\ln b, \tag{60}$$

which is an extension of condition (43) derived in section 5.

For penalties (56) and (57), the key point is the identity

$$\sum_{i\in S} [\nabla f_1(x)]_i = 0, \tag{61}$$

which is proved in [7] for both the 2D and the 3D setting and allows one to extend lemma 4 to the multi-dimensional case. Therefore, for the denoising problem a solution of the discrepancy equation exists if and only if the data satisfy the condition

$$\frac{1}{N}\sum_{i\in S} y_i \ln y_i > \frac{1}{2} + \bar{y}\ln \bar{y}, \tag{62}$$

which is satisfied in all the numerical simulations described in the next section. In the case of deblurring, the limit of x^*_β for $\beta \to 0$ is x^*, the unique minimizer of $f_0(x)$. Again one can easily extend lemma 4 and theorem 3 to the multi-dimensional case.

7. Numerical experiments

We test the effectiveness of the previous criterion with a few numerical experiments in the case of 2D images. A more accurate validation is deferred to a future paper.

7.1. Denoising

The denoising of 2D images corrupted by Poisson noise is investigated in [9, 20, 36]. In the first two papers TV regularization is used, while in the third one the penalty functions (56) and (57) are used as well as another penalty derived from a Markov random field model (which is also considered in [7]). The minimization of the corresponding functions $f_\beta(x)$ is obtained by adapting to these problems the general SGP algorithm proposed in [6].

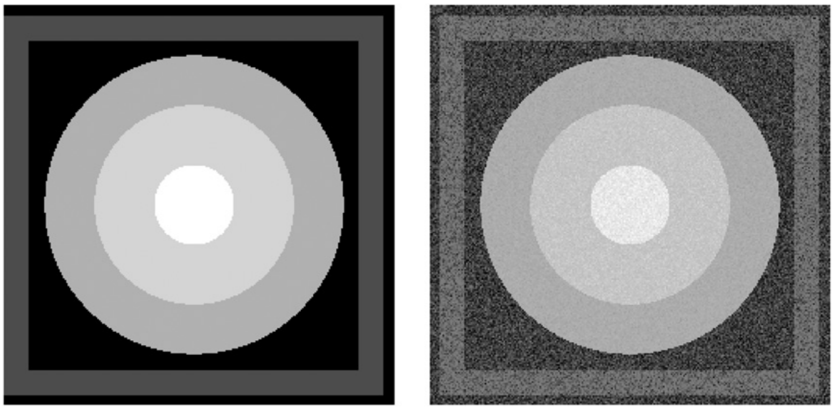

Figure 1. Left panel: the LCR_2 phantom. Right panel: the corresponding noisy version. Image display intensity has been adjusted to make the frame visible.

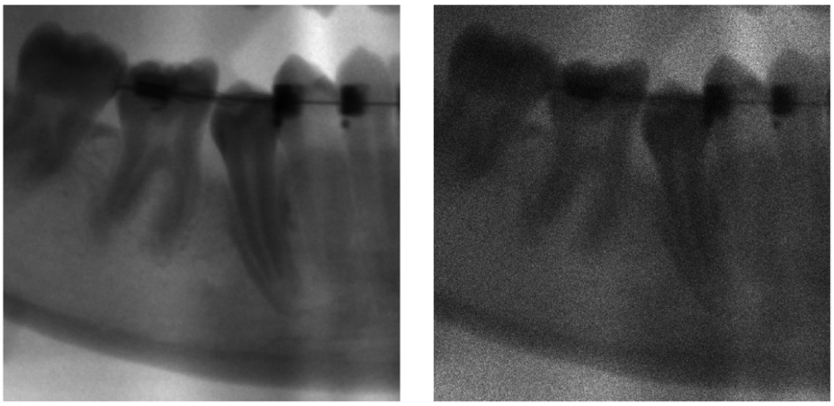

Figure 2. Left panel: the dental radiograph. Right panel: the lower dose version.

The discrepancy principle investigated in this paper is already proposed in [36] for the specific problem of denoising and the numerical results derived in that paper indicates that it provides sensible values of β. Another result derived in that paper is that the choice of δ seems to be crucial for improving the quality of the reconstruction. A very small value of δ corresponds approximately to TV regularization, but such a small value does not provide the best results. However, the choice of an 'optimal' δ in the denoising of real data is a completely open problem.

For this reason we attempt to use the discrepancy principle for obtaining some hint in the problem of estimating δ. The idea is to compute the value of β provided by equation (14) for different values of δ and the corresponding relative error in ℓ_2 norm, defined by

$$\rho(\beta) = \frac{\|x^*_\beta - \tilde{x}\|}{\|\tilde{x}\|}, \tag{63}$$

where $\tilde{x}$ is the object used in the numerical experiment for generating the noisy image y, and search for some correlation between these values.

Table 1. Values of β provided by the discrepancy principle for different values of δ and the corresponding relative ℓ_2 errors (expressed in %).

	LCR_1		LCR_2		LCR_3		Radiograph		Nebula	
δ	β	%	β	%	β	%	β	%	β	%
10^{1}	–	–	–	–	–	–	–	–	0.071	14.1
10^{0}	0.137	1.43	0.526	4.49	1.442	9.87	1.020	4.04	0.034	12.4
10^{-1}	0.123	1.22	0.408	3.04	0.790	5.42	0.563	3.63	0.039	13.1
10^{-2}	0.124	1.22	0.392	2.67	0.714	4.33	0.583	3.77	0.052	11.5
10^{-3}	0.146	1.37	0.435	2.81	0.797	4.39	0.814	3.55	0.073	11.7
10^{-4}	0.174	1.54	0.538	3.15	0.938	5.18	0.992	3.23	0.348	17.1
10^{-5}	0.173	1.54	0.717	3.49	1.784	5.95	2.933	5.12	1.367	18.7
10^{-6}	0.176	1.55	0.704	3.48	1.983	6.12	3.774	4.27	–	–
10^{-7}	0.174	1.54	0.704	3.46	2.077	6.33	4.047	4.30	–	–
10^{-8}	0.170	1.52	0.695	3.47	2.131	6.29	3.798	4.23	–	–

We first consider three different versions of a phantom already used in [36] and proposed in [20], here denoted respectively as LCR_1, LCR_2 and LCR_3, and corresponding to three different values of the central disk: 2000, 200 and 50. The second phantom is obtained by dividing the first by 10 and the third by 40. The phantom LCR_2 and its noisy version are shown in figure 1. Since the number of counts decreases from LCR_1 to LCR_3, the noise increases.

For each β the solution x^*_β is computed as in [36] and a secant-like method is used for solving equation (14) within a tolerance of 10^{-3}. In general, ten steps are required at most; the corresponding CPU time, on an AMD Athlon X2 Dual-Core at 3.1 GHz, is about ten seconds for a 256×256 image [30].

The numerical results are reported in the first columns of table 1. Since β, as given by (14), and the corresponding relative error (63) slowly depend on δ, at each step we reduce this parameter by a factor 10.

This table makes evident an interesting correlation, namely a minimum in the value of β corresponds to a minimum in the relative ℓ_2 error. This result, if confirmed by other simulations, could suggest a practical recipe: compute β by means of the discrepancy principle and select the value of δ corresponding to the minimum value of β.

The important point is that the values of the parameters obtained by this rule do not differ significantly from the 'best' values obtained in [36] by minimizing the relative ℓ_2 error. For completeness we report here these values. For LCR_1: $\delta = 3 \times 10^{-2}$, $\beta = 0.05$, error $= 0.8\%$; for LCR_2: $\delta = 10^{-2}$, $\beta = 0.2$, error $= 2.35\%$; for LCR_3: $\delta = 3 \times 10^{-3}$, $\beta = 0.67$, error $= 4.2\%$.

The LCR phantom is very simple (a piece-wise constant function) with well-defined edges. More complex phantoms were considered in [36]. The first is a good quality dental radiograph 512×512 shown in figure 2 with its noisy version (simulating an image obtained with a lower dose). The second one is the image of the nebula NGC5979 displayed in the central panel of figure 3.

In the last columns of table 1 we give the results obtained for these two phantoms by the same procedure used in the case of the LCR phantoms. The behavior is not so regular as in the previous case. The parameter β as a function of δ has only one minimum while the relative error has two minima. However, as a rule, we find that, for decreasing values of δ, the minimum of β corresponds to a relative minimum of the error, with a value not significantly

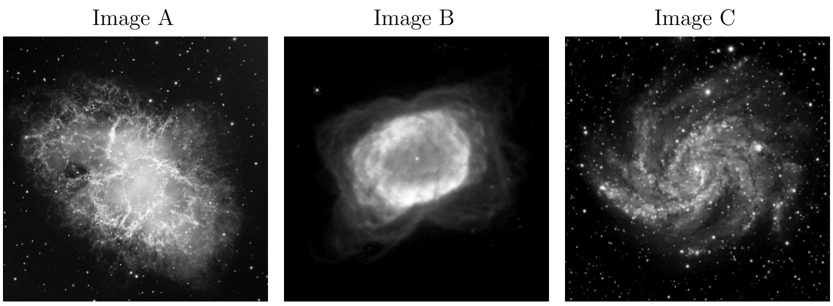

Figure 3. Original images. A is the image of the Crab nebula, B the image of the nebula NGC5979 and C the image of a spiral galaxy.

greater than that of the absolute minimum. We also remark that in the case of the radiograph, the optimal values providing the minimum reconstruction error of 2.79% are $\delta = 0.1$ and $\beta = 0.25$, while in the case of the nebula, the minimum error of 10.4% is obtained with $\delta = 0.1$ and $\beta = 0.02$. We must remark that the nebula is a smooth object and therefore the use of edge-preserving regularization is not the best strategy. As shown in [36] the 'best' reconstruction is affected by strong 'cartoon' artifacts.

The previous results suggest the following thumb rule for estimating δ in the case of real data: start with a large value of δ and compute β by the discrepancy principle; then decrease δ by factors of 10 and compute β for each of them; choose the value of δ corresponding to the first minimum of β.

In our experiments we also compute the discrepancy function (15) and we find that it is lower than (13); so it provides a larger value of β.

7.2. Deblurring

Several algorithms have been proposed for the computation of x^*_β with different penalty functions $f_1(x)$. Let us mention some of them, without claiming completeness. For instance, in the case of standard Tikhonov regularization, a computational method is proposed in [1] and an EM method is investigated in [11]; as already remarked, the simple denoising method (40) is just the M-step of the latter iterative algorithm. TV regularization is investigated in [2, 8] and a quasi-Newton method for the computation of x^*_β is used in [2] while a primal-dual Bregman method is proposed in [8]. In [3] the gradient-projection-reduced Newton method, outlined in [35], is used in the case of a penalty of the form $f_1(x) = (Cx, x)$, with three different choices of the matrix C. From our side, work is in progress for applying the SGP algorithm [6] to this problem in the case of edge-preserving regularization, with a choice of the scaling suggested by the split-gradient method (SGM) [17] and already used in [36]. Preliminary results indicate that this approach provides a very simple and very efficient algorithm and that the criterion proposed in this paper works well.

However, it is known that a suitable stopping of some iterative methods amounts to a kind of 'regularization'. In the case of Poisson data, the prototype of these methods is the EM algorithm proposed by Shepp and Vardi for emission tomography [31] and previously proposed by Richardson [28] and Lucy [21] for image deconvolution. The algorithm is as follows:

$$x^{(k+1)} = x^{(k)} H^T \frac{y}{Hx^{(k)} + be_M}, \tag{64}$$

and is, in general, initialized with a constant image.

The problem of suitable stopping rules has been investigated by several authors. We mention a few. In [18, 19, 34] a definition of 'feasible objects', as those objects that could have generated the detected data, is given and feasibility tests are introduced for stopping the iterations. The generalized cross-validation is used in [24] to the same purpose, while a comparison of the iteration with the solution provided by the Wiener filter is proposed in [16]. Finally another statistical method is investigated in [25].

Since in the case of Gaussian noise, the Morozov discrepancy principle is also used as a stopping rule for iterative methods, such as Landweber or the conjugate gradient, that converge to a minimizer of the least-square function [12], it is attractive to test whether the proposed discrepancy principle may be used for stopping EM or similar methods.

In this vein we propose the following stopping criterion:

- at each iteration compute the discrepancy $D_y^{(k)} = 2 f_0(x^{(k)})/M$ and stop the iterations as soon as $D_y^{(k)} < 1$.

Indeed, if condition (62) is satisfied, by initializing the iteration with a constant image equal to $\bar{y}$, then $D_y^{(0)} > 1$; moreover, from general results on the EM method, it is known that $D_y^{(k)}$ is a decreasing function of k. Therefore, if the condition $f_0(x^*) < M/2$ is also satisfied, then during the iteration $D_y^{(k)}$ must cross 1. This criterion is very simple and, even if it does not provide the final answer to the problem of stopping EM iteration, it may deserve some consideration.

In addition to EM, we can apply the criterion also to the algorithm SGP proposed in [6], that is much more efficient than EM and, in the monotone version, provides iterations with decreasing values of $D_y^{(k)}$.

To test the criterion in both cases we consider a problem of image deconvolution ($M = N$) using the same images as in [6]. They are obtained from three astronomical objects shown in figure 1. These objects, 256×256, each one normalized to three different total fluxes, are convolved with an ideal diffraction limited PSF, the so-called Airy pattern, and the results are perturbed with Poisson noise, thus obtaining for each object three images with different noise levels. Then each image is deconvolved using the two algorithms EM and SGP and in both cases the iterations are stopped via the criterion indicated above (discrepancy stopping) and via the criterion (optimal stopping) that consists in searching the iteration which renders

$$\rho^{(k)} = \frac{\|x^{(k)} - \tilde{x}\|}{\|\tilde{x}\|}, \tag{65}$$

a minimum. The results are reported in table 2. With the exception of the EM algorithm applied to image B with the largest noise level, in all the other cases the number of iterations required by the 'discrepancy criterion' is much smaller than that required by the 'optimality criterion' and provides a reconstruction error that is not significantly greater than the optimal one. The reason of this success is due to the fact that the minimum of the reconstruction error is very flat (extremely flat in the case of EM). We remark that in these examples the same results are obtained by means of the discrepancy function (15).

Since SGP has been already implemented on graphics processors (GPU) [29], implementing the discrepancy stopping rule might provide a very fast tool for image deblurring with automatic stopping of iterations. Potential applications are open to image deblurring in astronomy and microscopy. In the last case an extension to 3D images is required.

Table 2. Application of the discrepancy principle to the stopping of the EM and SGP iterative algorithms and comparison with the stopping corresponding to the minimum reconstruction error.

	SGP				EM			
	Optimal stop		Discrepancy stop		Optimal stop		Discrepancy stop	
Flux	Iteration	Error	Itearation	Error	Iteration	Error	Iteration	Err
				Image A				
4.43×10^9	380	0.1851	59	0.1865	10 000*	0.1852	992	0.1864
7.02×10^8	103	0.1866	26	0.1889	4047	0.1865	249	0.1889
4.43×10^7	21	0.1944	8	0.1965	414	0.1942	112	0.1964
				Image B				
4.43×10^9	251	0.0513	69	0.0539	10 000*	0.0512	1151	0.0537
7.02×10^8	157	0.0542	53	0.0559	4185	0.0541	414	0.0570
4.43×10^7	26	0.0689	21	0.0689	500	0.0687	812	0.0695
				Image C				
4.43×10^9	736	0.2922	143	0.2949	10 000*	0.2929	4341	0.2941
7.02×10^8	374	0.2953	66	0.2984	10 000*	0.2945	1305	0.2988
4.43×10^7	41	0.3127	15	0.3161	1459	0.3110	293	0.3159

(* indicates that the minimum error has not been reached after 10000 iterations.)

Appendix.

Proof of lemma 1. Let x_β^* be the unique minimizer of $f_\beta(x)$ and let $0 \leqslant \beta_1 < \beta_2$; we need to prove the inequalities: (i) $f_0(x_{\beta_1}^*) < f_0(x_{\beta_2}^*)$ and (ii) $f_1(x_{\beta_1}^*) > f_1(x_{\beta_2}^*)$.

(i) The convexity of $f_0(x)$ gives

$$f_0(x_{\beta_2}^*) \geqslant f_0(x_{\beta_1}^*) + \nabla f_0(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*). \tag{A.1}$$

Since $x_{\beta_1}^*$ solves problem (9), it satisfies the optimality condition

$$\nabla f_{\beta_1}(x_{\beta_1}^*)^T(x - x_{\beta_1}^*) \geqslant 0, \qquad \forall x \in \Omega; \tag{A.2}$$

then letting $x = x_{\beta_2}^*$ we get

$$\left(\nabla f_0(x_{\beta_1}^*) + \beta_1 \nabla f_1(x_{\beta_1}^*)\right)^T(x_{\beta_2}^* - x_{\beta_1}^*) \geqslant 0, \tag{A.3}$$

that is

$$\nabla f_0(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*) \geqslant -\beta_1 \nabla f_1(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*). \tag{A.4}$$

The optimality of $x_{\beta_2}^*$ and the strict convexity of $f_\beta(x)$ yield

$$f_{\beta_2}(x_{\beta_1}^*) \geqslant f_{\beta_2}(x_{\beta_2}^*) > f_{\beta_2}(x_{\beta_1}^*) + \nabla f_{\beta_2}(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*) \tag{A.5}$$

so that

$$\nabla f_{\beta_2}(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*) < 0, \tag{A.6}$$

which implies

$$-\nabla f_0(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*) > \beta_2 \nabla f_1(x_{\beta_1}^*)^T(x_{\beta_2}^* - x_{\beta_1}^*). \tag{A.7}$$

Adding inequalities (A.4) and (A.7) and using $(\beta_2 - \beta_1) > 0$ gives

$$\nabla f_1\big(x^*_{\beta_1}\big)^T\big(x^*_{\beta_2} - x^*_{\beta_1}\big) < 0; \tag{A.8}$$

thus, from (A.4) we obtain

$$\nabla f_0\big(x^*_{\beta_1}\big)^T\big(x^*_{\beta_2} - x^*_{\beta_1}\big) > 0 \tag{A.9}$$

and from inequality (A.1) we conclude $f_0\big(x^*_{\beta_1}\big) < f_0\big(x^*_{\beta_2}\big)$.

(ii) The convexity of $f_1(x)$ implies

$$f_1\big(x^*_{\beta_1}\big) \geqslant f_1\big(x^*_{\beta_2}\big) + \nabla f_1\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big). \tag{A.10}$$

From the optimality condition for $x^*_{\beta_2}$ we have

$$\big(\nabla f_0\big(x^*_{\beta_2}\big) + \beta_2 \nabla f_1\big(x^*_{\beta_2}\big)\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big) \geqslant 0, \tag{A.11}$$

and consequently

$$\nabla f_0\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big) \geqslant -\beta_2 \nabla f_1\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big). \tag{A.12}$$

Moreover, the optimality of $x^*_{\beta_1}$ and the strict convexity of $f_\beta(x)$ imply

$$f_{\beta_1}\big(x^*_{\beta_2}\big) \geqslant f_{\beta_1}\big(x^*_{\beta_1}\big) > f_{\beta_1}\big(x^*_{\beta_2}\big) + \nabla f_{\beta_1}\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big), \tag{A.13}$$

so that

$$\nabla f_{\beta_1}\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big) < 0, \tag{A.14}$$

which implies

$$-\nabla f_0\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big) > \beta_1 \nabla f_1\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big). \tag{A.15}$$

By adding inequalities (A.12) and (A.15) and using $(\beta_1 - \beta_2) < 0$ we obtain

$$\nabla f_1\big(x^*_{\beta_2}\big)^T\big(x^*_{\beta_1} - x^*_{\beta_2}\big) > 0, \tag{A.16}$$

and from inequality (A.10) we conclude $f_1\big(x^*_{\beta_1}\big) > f_1\big(x^*_{\beta_2}\big)$. □

References

[1] Bardsley J M and Laobeul N 2008 Tikhonov regularized Poisson likelihood estimation: theoretical justification and a computational method *Inverse Problems Sci. Eng.* **16** 199–215

[2] Bardsley J M and Luttman A 2009 Total variation-penalized Poisson likelihood estimation for ill-posed problems *Adv. Comput. Math.* **31** 35–59

[3] Bardsley J M and Goldes J 2009 Regularization parameter selection methods for ill-posed Poisson maximum-likelihood estimation *Inverse Problems* **25** 095005

[4] Barrett H H and Meyers K J 2003 *Foundations of Image Science* (New York: Wiley) pp 1047–48

[5] Bertero M, Boccacci P, Desiderá G and Vicidomini G 2009 Image deblurring with Poisson data: from cells to galaxies *Inverse Problems* **25** 123006

[6] Bonettini S, Zanella R and Zanni L 2009 A scaled gradient projection method for constrained image deblurring *Inverse Problems* **25** 015002

[7] Bonettini S and Ruggiero V 2010 A discrepancy principle for Poisson data: uniqueness of the solution for 2D and 3D data http://eprints.unife.it/195

[8] Brune C, Sawatzky A and Burger M 2010 Primal and dual Bregman methods with application to optical nanoscopy *Int. J. Comput. Vis.* (to appear) (doi:10.1007/s11263-010-0339-5)

[9] Chan R H and Chen K 2007 Multilevel algorithms for a Poisson noise removal model with Total-variation regularization *Int. J. Comput. Math.* **84** 1183–98

[10] Charbonnier P, Blanc-Féraud L, Aubert G and Barlaud A 1997 Deterministic edge-preserving regularization in computed imaging *IEEE Trans. Image Processing* **6** 298–311

[11] Conchello J A and McNally J G 1996 Fast regularization technique for expectation maximization algorithm for optical sectioning microscopy *Proc. SPIE* **2655** 199–208
[12] Engl H W, Hanke M and Neubauer N 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[13] Feller W 1968 *An Introduction to Probability Theory and its Applications* (New York: Wiley)
[14] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41
[15] Hall P and Titterington D M 1987 Common structure of techniques for choosing smoothing parameters in regression problem *J. R. Stat. Soc.* B **49** 184–98
[16] Lantéri H, Soummer R and Aime C 1999 Comparison between ISRA and RLA algorithms. Use of a Wiener filter based criterion *Astron. Astrophys. Suppl. Ser.* **140** 235–46
[17] Lantéri H, Roche M and Aime C 2002 Penalized maximum-likelihood image restoration with positivity constraints: multiplicative algorithms *Inverse Problems* **18** 1397–419
[18] Llacer J and Veklerov E 1989 Feasible images and practical stopping rules for iterative algorithms in emission tomography *IEEE Trans. Med. Imaging* **8** 186–93
[19] Llacer J 1990 On the validity of hypothesis testing for feasibility of image reconstruction *IEEE Trans. Med. Imaging* **9** 226–30
[20] Le T, Chartran R and Asaki T J 2007 A variational approach to reconstructing images corrupted by Poisson noise *J. Math. Imaging Vis.* **27** 257–63
[21] Lucy L B 1974 An iterative technique for the rectification of observed distributions *Astron. J.* **79** 745–54
[22] Markham J and Conchello J A 2001 Fast maximum-likelihood image-restoration algorithms for three-dimensional fluorescence microscopy *J. Opt. Soc. Am.* **A-18** 1062–71
[23] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (Philadelphia: SIAM)
[24] Perry K M and Reeves S J 1994 Generalized cross-validation as a stopping rule for the Richardson–Lucy algorithm *The Restoration of HST Images and Spectra* 2nd edn ed R J Hanisch and R L White (Baltimore: Space Telescope Science Institute) 97–103
[25] Prasad S 2002 Statistical-information-based performance criteria for Richardson–Lucy image deblurring *J. Opt. Soc. Am.* **A-19** 1286–96
[26] Ramani S, Blu T and Unser M 2008 Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms *IEEE Trans. Image Process.* **17** 1540–54
[27] Ramani S, Vonesch C and Unser M 2008 Deconvolution of 3D fluorescence micrographs with automatic risk minimization *Proc. ISBI* 732–5
[28] Richardson W H 1972 Bayesian based iterative method of image restoration *J. Opt. Soc. Am.* **62** 55–9
[29] Ruggiero V, Serafini T, Zanella R and Zanni L 2009 Iterative regularization algorithms for constrained image deblurring on graphics processors *J. Glob. Optim.* **48** 148–57
[30] Serafini T, Zanella R and Zanni L 2010 Gradient projection methods for image deblurring and denoising on graphics processors *Advances in Parallel Computing* vol 19 ed B Chapman, F Desprez, G R Joubert, A Lichnewsky, F Peters and T Priol (Amsterdam: IOS Press) pp 59–66
[31] Shepp L A and Vardi Y 1982 Maximum-likelihood reconstruction for emission tomography *Trans. Med. Imaging* **MI-1** 113–22
[32] Snyder D L and Miller M I 1991 *Random Point Processes in Time and Space* (New York: Springer)
[33] Snyder D L, Hammoud A M and White R L 1993 Image recovery from data acquired with a charged-coupled-device camera *J. Opt. Soc. Am.* A **10** 1014–23
[34] Veklerov E and Llacer J 1987 Stopping rule for the MLE algorithm based on statistical hypothesis testing *Trans. Med. Imaging* **6** 313–9
[35] Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia: SIAM)
[36] Zanella R, Boccacci P, Zanni L and Bertero M 2009 Efficient gradient projection methods for edge-preserving removal of Poisson noise *Inverse Problems* **25** 045010