Iterative image reconstruction: a point of view

Mario Bertero, Henri Lantéri and Luca Zanni

Abstract. Several iterative methods are available for solving the ill-posed problem of image reconstruction. They are motivated by different approaches and may derive from methods used for the solution of linear equations or the minimization of suitable functionals. In this paper we adopt the approach flowing from maximum likelihood to Bayesian formulation of image reconstruction and providing a generalization of the classical regularization theory. This approach leads to the minimization of functionals derived from properties of the noise and, possibly, from additional information on the solution. We investigate a class of scaled gradient methods, based on a suitable decomposition of the gradient, and we show that this class contains some of the methods used for the solution of maximum likelihood problems in image reconstruction. We also obtain very simple regularized versions of these methods. Constraints of non-negativity and flux conservation are taken into account by considering scaled gradient projection (SGP) methods, derived from the previous approach, and for them a convergence proof can be given. Numerical experience on a particular problem shows that SGP can provide a considerable increase in efficiency with respect to the standard algorithm used for that problem. Work is in progress in order to understand whether a similar gain can be achieved in other cases.

Introduction

Image reconstruction is an important problem in several domains of applied science such as Medical Imaging, Microscopy and Astronomy. In many instances it can be formulated as a linear inverse and ill-posed problem. In general, it is a large scale problem, since nowadays it is quite common to process images of the order of several mega-pixels (or mega-voxels in the 3D case). Therefore efficient numerical methods are required and the use of iterative methods, except a few special cases, is an almost compelling choice. A plethora of these methods is available, in general derived from different approaches, ranging from iterative methods for the solution of linear equations to methods for solving specific

M. Bertero and L. Zanni are partly supported by MUR Grant 2006018748.

variational problems. For these reasons we think that it is important to clearly define the framework where iterative methods are introduced. In this paper we choose a specific one that we motivate with a brief survey of some important achievements in the theory of inverse and ill-posed problems.

It is well known that the main difficulty in the treatment of ill-posed problems is the lack of continuous dependence of the solution on the data. After the basic work of Jacques Hadamard [22,23], only well-posed problems are investigated for a long period. However, around the fifties, it is shown that, in the case of ill-posed problems for partial differential equations, it is possible to restore continuous dependence by restricting attention to approximate solutions satisfying a prescribed bound [27,38]; it is also demonstrated that these problems may be of interest in several applications. Similar results are obtained for analytic continuation [11] and a general formulation of these approaches is given in terms of least-squares problems with prescribed bounds [34].

Another research line is opened by a paper of Tikhonov [44] where the continuity of the inverse mapping is proved by assuming that the direct mapping is restricted to a compact set. In that paper only exact data are considered but, in the case of inexact data, least-squares solutions restricted to a compact set are investigated and their continuous dependence on the data is proved in [26], where the case of weak compactness is also considered.

In both approaches the relevance of additional information on the solution is evident. In the first one, it consists of prescribed bounds on the solution while in the second one it is expressed as the requirement of searching for a solution in a compact set. The role of additional information is less evident in *regularization theory*, that, proposed in 1963 [45,46], has dominated (and is still dominating) the scene of inverse problems. In this approach the focus is on the approximation of an ill-posed problem by a family of well-posed ones.

However, the relevance of additional information reappears, in a statistical form, in a number of papers devoted to the application of Wiener-Kolmogorov filtering to the solution of first-kind Fredholm integral equations [17,43]. A more general and abstract formulation of this approach is given in [18]. Moreover, a statistical regularization derived from Bayes formula is investigated in [47]. Even if these approaches provide a new insight in the problem, the main limitation is due to the fact that the analysis is restricted to the case of linear estimation or, equivalently, to the case of Gaussian random variables. In particular, additive Gaussian noise is assumed, so that the problem is reduced again to the minimization of a penalized least-squares functional.

Therefore new perspectives are open by the paper of Shepp and Vardi [41] on positron emission tomography (PET) where a different noise model is considered, the so-called Poisson noise due to photon counting. A further step is done in a paper by Snyder *et al.* [42], where a more complex noise model is introduced. All these contributions enlarge the field of applicability of statistical methods and, in this framework, regularization theory appears as a particular case (although extremely rich of mathematical results) of a more general theory based on Bayes formula. For this reason we adopt this approach in this paper.

In Section 1 we analyze the properties of the data of an imaging system, showing that it is quite natural to look for a statistical formulation of an image reconstruction problem. In Section 2 we introduce the two approaches known as maximum likelihood estimate and maximum a posteriori estimate and we consider, as particular examples, the three noise models indicated above and denoted, respectively, as Gauss, Poisson and Gauss+Poisson noise. In all cases we have to minimize specific functionals with the additional constraint of non-negativity of the solution.

In Section 3 we introduce a scaled gradient method that has been recently proposed [30,31] for solving these constrained minimization problems. In the case of maximum likelihood estimation we re-obtain, for the three noise models above, three well-known iterative methods. Moreover the approach is able to provide very simple regularized versions of these algorithms.

Finally, in Section 4, we introduce an additional constraint on the 1-norm of the solution and we use the scaling of the gradient discussed in the previous Section for introducing a scaled gradient projection (SGP) method. Its convergence can be proved [10] but, in this paper, we only give algorithmic details and we discuss a specific application. Indeed, the method has been implemented for the deconvolution of images corrupted by Poisson noise. We discuss the numerical results we have obtained and we show that SGP provides a considerable computational gain with respect to the standard expectation maximization (EM) method [41]. This suggests to apply the method also to regularized versions of this algorithm and possibly also of other ones. Work is in progress in this direction.

ACKNOWLEDGEMENTS. This work was partially supported by MUR (Italian Ministry for University and Research), Grant 2006018748.

1 Mathematical modeling of data acquisition

Since image reconstruction is an example of ill-posed problem an accurate mathematical modeling is a necessary prerequisite for obtaining a correct formulation of the problem.

An imaging system consists, in general, of two parts.

- The first is an apparatus (formed by physical components such as sources, collimators, mirrors, lenses etc.) able to transform the radiation (microwaves, photons, X-rays, γ-rays, ultrasound etc.) emitted or transmitted by the sample to be imaged (in the following called **the object**) into a detectable radiation containing useful information about the spatial properties of the object.
- The second is a detector providing measured values of the incoming radiation; this is the part of the system introducing sampling and noise.

Therefore general features of the data are the following.

- **Data are discrete** and the discretization is not decided by the mathematician but by the physicist or engineer who designed the imaging system; in general, sampling theorems are taken into account in the design and, very often, data are oversampled. In any case they consist of a set of numbers characterized by a multi-index (a pair of integers in 2D imaging and a triple in 3D imaging). However, as usual, we assume that these numbers are ordered to form a vector with m components (m is the number of data) and denoted by $y = \{y_i\}_{i=1}^m \in \mathbb{R}^m$.
- **Data are realizations of random variables**, as a consequence of the noise introduced by the detection system. Therefore, a component y_i of the data vector is the realization of a random variable Y_i. We will denote by $Y = \{Y_i\}_{i=1}^m$ the vector valued random variable corresponding to the data.

The modeling of the apparatus implies the use of the equations that describe the propagation of the radiation used in the imaging process. The goal is to get a mapping that transforms the spatial distribution of the object into the radiation incoming on the detector. In this paper we only consider the case where, thanks to the use of suitable physical approximations, this mapping is **linear**.

Therefore, if we denote by $x(s)$ a function of the space variables describing the properties of the object, we obtain a semi-discrete mapping [5] transforming this function into sampled values of the radiation before detection (also called exact data values), that we denote by $\bar{y}_i$. In many instances one must also take into account the existence of some

background radiation, denoted by b_i, so that we have

$$\bar{y}_i = (Hx)_i + b_i = (Hx + b)_i, \tag{1.1}$$

where H is the semi-discrete mapping and $b = \{b_i\}_{i=1}^m$. In general, one can assume that b is a constant vector.

An approach based on the use of functional spaces for describing the properties of the object can be very elegant but it can be unnecessarily heavy if one is looking for probabilistic approaches [4]. For this reason we prefer a completely discrete model, remarking that, now, the sampling of the object is at disposal of the mathematician, if this degree of freedom appears to be useful. Therefore, we assume that the object consists of a set of numbers, also characterized by a multi-index, and again these can be ordered to form a vector with n components (n is the number of unknowns of the problem) and denoted by $x = \{x_j\}_{j=1}^n \in \mathbb{R}^n$. In general, one has $n \neq m$.

The mapping describing the transformation from x to the exact data values is now a matrix that will be denoted again by H. We assume that it satisfies the following conditions

$$H_{i,j} \geq 0; \sum_{i=1}^{m} H_{i,j} > 0\ ,\ \forall j;\ \sum_{j=1}^{n} H_{i,j} > 0\ ,\ \forall i. \tag{1.2}$$

In other words we assume that each row or column contains at least one non-zero element. We will use the following notation for the sum of the entries of one column of this matrix

$$h_j = \sum_{i=1}^{m} H_{i,j}; \quad j = 1, \ldots, n, \tag{1.3}$$

and we will denote by h the corresponding vector. Moreover, products and quotients of vectors will be intended in the Hadamard sense, *i.e.* component by component

$$(hx)_j = h_j x_j, \left(\frac{x}{h}\right)_j = \frac{x_j}{h_j}. \tag{1.4}$$

About the detection system, we already remarked that it introduces sampling and noise, and that noise is a random process, so that the detected values are realizations of random variables. Therefore a modeling of this system requires a modeling of the noise, *i.e.* a model of its probability density (a function or a distribution). This density depends on the object x and therefore we denote it as $p_Y(y; x)$. The following assumptions are, in general, accepted as reasonable ones.

- The random variables Y_i and Y_l associated to different elements of the detector are statistically independent, so that we can write

$$p_Y(y;x) = \prod_{i=1}^{m} p_{Y_i}(y_i;x). \tag{1.5}$$

- The expected value of Y_i is just given by the exact value of the incoming radiation so that we have

$$E\{Y\} = \int y\ p_Y(y;x)\ dy = Hx + b. \tag{1.6}$$

Example 1.1. The first example is provided by the so-called **additive white Gaussian noise**. In this example Y is given by

$$Y = Hx + b + E \tag{1.7}$$

where E is a vector-valued random variable with statistically independent components, all having the same Gaussian distribution, with expected value 0 and variance σ^2, so that

$$P_E(e) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^m \exp\left(-\frac{1}{2\sigma^2}||e||^2\right), \tag{1.8}$$

where $||\cdot||$ denotes the usual 2-norm. Therefore the statistical model for the detected data is given by

$$p_Y(y;x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^m \exp\left(-\frac{1}{2\sigma^2}||y-(Hx+b)||^2\right). \tag{1.9}$$

Example 1.2. The second example is the so-called **Poisson noise**, describing, in general, the noise affecting counting processes (sometimes it is also called "photon noise"). In such a case each Y_i is a Poisson random variable with expected value given by Equation (1.1)

$$Y_i \sim \text{Poisson}\{(Hx+b)_i\}, \tag{1.10}$$

so that its probability density is a distribution with support the set of the non-negative integers (each y_i is a non-negative integer). We have

$$p_Y(y;x) = \prod_{i=1}^{m} \frac{\mathrm{e}^{-(Hx+b)_i}(Hx+b)_i^{y_i}}{y_i\ !}. \tag{1.11}$$

Example 1.3. The third example is the so-called **Gauss+Poisson noise** and is a more refined model of the noise affecting data detected by a charged-coupled-device (CCD) camera, as described in [42]. It is given by

$$Y = Z + E, \tag{1.12}$$

where Z is a Poisson process as in Example 1.2 while E is an additive white Gaussian noise as in Example 1.1. In such a case the probability density is given by

$$p_Y(y; x) = \prod_{i=1}^{m} \left(\sum_{l=0}^{+\infty} \frac{\mathrm{e}^{-(Hx+b)_i} (Hx+b)_i^l}{l!} \frac{\mathrm{e}^{-\frac{1}{2\sigma^2}(y_i - l)^2}}{\sqrt{2\pi}\sigma} \right). \tag{1.13}$$

In conclusion, we have a complete model of the process of data formation and acquisition when we know the imaging matrix H, the background b and the probability density $p_Y(y; x)$.

2 Statistical formulations of the problem of image reconstruction

Let us assume that we have a complete model in the sense specified above and that we have a detected image y (for simplicity, we do not introduce some specific notation for the detected image), *i.e.* a realization of the random variable Y. The problem of image reconstruction is to find an estimate $\bar{x}$ of the unknown object corresponding to the image y. The trivial approach should be to look for a solution of the linear equation $Hx + b = y$, but, as we know, in general this approach is not successful since the matrix H is ill-conditioned. The fact that the most frequently used algorithm in tomography, namely the *filtered back-projection*, is just coming from the solution of the linear equation is an exception to this rule. Information about statistical properties of the data suggests to look for statistical approaches to the problem.

2.1 Maximum likelihood formulation

Since we assume to know the probability density $p_Y(y; x)$ of the data and since, in this density, the unknown object appears as a set of unknown parameters, at first glance the problem of image reconstruction appears as a classic problem of parameter estimation. Then the standard approach is the so-called **maximum likelihood** (ML) estimation. In our specific application, for a given detected image y, it consists in introducing the *likelihood function* defined by

$$L_y^Y(x) = p_Y(y; x); \tag{2.1}$$

clearly this is only a function of x since y is given and is just the detected image. Then the ML-estimate of the unknown object is any object x^* that maximizes the likelihood function

$$x^* = \operatorname*{argmax}_{x \in \mathbb{R}^n} \; L_y^Y(x). \tag{2.2}$$

It is obvious that this definition is meaningful if the likelihood function has maximum points.

In our applications the likelihood function is the product of a very large number of factors, so that it is convenient to take the logarithm of this function; moreover, if we consider the negative logarithm (the so-called neglog) the maximization problem is transformed into a minimization one. Therefore we introduce the functional

$$J_0(x; y) = -A \ln L_y^Y(x) + B, \tag{2.3}$$

where A, B are suitable constants that can be introduced in order to simplify the expression of the functional. Since the neglog function is strictly convex, the problem of Equation (2.2) is equivalent to the following one

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^n} \; J_0(x; y). \tag{2.4}$$

We reconsider now the three examples of the previous section.

Example 2.1. In the case of additive white Gaussian noise, by a suitable choice of the constants A, B, we obtain

$$J_0(x; y) = ||Hx + b - y||^2, \tag{2.5}$$

and therefore the ML approach coincides with the well-known least-squares (LS) approach. It is also well-known that the functional of Equation (2.5) is convex, and strictly convex if and only if the equation $Hx = 0$ has only the solution $x = 0$. Moreover it has always global minima, *i.e.* the LS-problem has always a solution; but this problem is ill-conditioned, in the case of image reconstruction, since it is equivalent to the solution of the Euler equation

$$H^T H \, x = H^T (y - b), \tag{2.6}$$

and the condition number of the matrix H can be very large. Indeed, the continuous version of this problem is ill-posed (the matrix H comes from the discretization of an integral operator, very often a compact one) and this ill-posed problem is the starting point of the so-called *Tikhonov regularization theory* (see, for instance, [16,46]). Therefore, this theory is

based on the tacit assumption that the noise affecting the data is additive and Gaussian.

We remark that, in the case of object reconstruction, since objects are non-negative, we should consider the minimization of the functional of Equation (2.5) on the non-negative orthant. With such a constraint the problem is not treatable in the standard framework of regularization theory.

Example 2.2. In the case of Poisson noise, if we introduce the so-called Kullback-Leibler (KL) divergence of a vector z from a vector y, defined by

$$D_{\mathrm{KL}}(y, z) = \sum_{i=1}^{m} \left\{ y_i \ln \frac{y_i}{z_i} + z_i - y_i \right\}, \tag{2.7}$$

then, with a suitable choice of the constants A, B, the functional $J_0(x; y)$ is given by

$$\begin{aligned} J_0(x; y) &= D_{\mathrm{KL}}(y; Hx + b) \\ &= \sum_{i=1}^{m} \left\{ y_i \ln \frac{y_i}{(Hx + b)_i} + (Hx + b)_i - y_i \right\}. \end{aligned} \tag{2.8}$$

It is quite natural to take the non-negative orthant as the domain of this functional. Moreover, it is well-known that it is convex (strictly convex if the equation $Hx = 0$ has only the solution $x = 0$) [41], non-negative and locally bounded. Therefore it has global minima.

The properties of the continuous version of this functional and its minimization are investigated in [35–37]. In particular, in [37] an example is given where the functional does not have a minimum in the classical sense, hence proving the ill-posedness of this minimization problem. As a consequence, we should expect that noise strongly affects the minima of the discrete problem. Indeed this is the case and the specific effect of the noise in this problem is known as *checkerboard effect*, since many components of the minima are zero.

Example 2.3. In the case of Gauss+Poisson noise, the functional $J_0(x; y)$ is given by

$$\begin{aligned} J_0(x; y) &= -\sum_{i=1}^{m} \ln M_i(x; y) \\ M_i(x; y) &= \sum_{l=0}^{+\infty} \frac{\mathrm{e}^{-(Hx+b)_i} (Hx + b)_i^l}{l!} \mathrm{e}^{-\frac{1}{2\sigma^2}(l - y_i)^2}. \end{aligned} \tag{2.9}$$

For future use we also introduce the function

$$N_i(x;\, y) = \sum_{l=0}^{+\infty} \frac{\mathrm{e}^{-(Hx+b)_i}(Hx+b)_i^l}{l!} \mathrm{e}^{-\frac{1}{2\sigma^2}(l+1-y_i)^2}. \tag{2.10}$$

It has been recently proved [3] that also this functional is convex (strictly convex if the equation $Hx = 0$ has the unique solution $x = 0$), non-negative and locally bounded. Therefore it also has global minima on the non-negative orthant. As far as we know, no result is available about the ill-posedness of this minimization problem. However, numerical experience demonstrates that also in this case the minimum points are affected by the checkerboard effect [3].

Remark 2.4. The previous examples demonstrate that, in the case of image reconstruction, ML problems are ill-posed or ill-conditioned. That means that one is not interested in computing the minimum points x^* of the functionals corresponding to the different noise models because they do not provide sensible estimates $\bar{x}$ of the unknown object.

However, as it is known, the ML approach deserves an accurate analysis. The previous remark only implies that one must be very careful in applying to these problems methods derived from optimization theory. In particular, in our opinion, very efficient methods, such as second order methods, pointing directly to a minimum can be dangerous. On the other hand, numerical experience (and, in some cases, also theoretical results) demonstrates that first order methods can provide acceptable (regularized) solutions by early stopping. In the framework of regularization theory the study of iterative methods with such a property (we only mention Landweber, steepest descent and conjugate gradient methods) is a widely investigated topic.

Finally we point out that, since objects are non-negative, the non-negativity constraint must always be introduced in the formulation of the previous minimization problems.

2.2 Bayesian formulation

The previous remark is not surprising in the framework of inverse problem theory. Indeed it is generally accepted that, if the formulation of the problem does not use some additional information on the object, then the resulting problem is ill-posed. This is what happens in the maximum likelihood approach because we only use information about the noise with, possibly, the addition of the constraint of non-negativity.

The additional information may consist, for instance, in prescribed bounds on the solution and/or its derivatives up to a certain order (in general not greater than two). These prescribed bounds can be introduced in

the problem as additional constraints in the variational formulation provided by ML. However, in this paper, we adopt a completely probabilistic approach, called **Bayesian approach**, where the additional information is given in the form of statistical properties of the object.

In other words, one assumes that the unknown object x is also a realization of a (vector valued) random variable X. Then, a different interpretation of the probability density $p_Y(y; x)$ is introduced: this is considered as the **conditional probability density** of Y when the random variable X assumes the value x

$$p_Y(y; x) = p_Y(y|X = x). \tag{2.11}$$

For simplicity we will write $p_Y(y|x)$.

Then additional information on the unknown object x is introduced by providing the probability density of X, the so-called **prior**, that will be denoted by $p_X(x)$. The most frequently used priors are of the Gibbs type, i.e they have the following form

$$p_X(x) = \frac{1}{Z} \mathrm{e}^{-\mu \Omega(x)}, \tag{2.12}$$

where Z is a normalization constant, μ is a positive parameter (a hyperparameter in the statistical language, a regularization parameter in the language of regularization theory), while $\Omega(x)$ is a functional, possibly convex.

The previous assumptions imply that the joint probability density of the random variables X, Y is given by

$$p_{XY}(x, y) = p_Y(y|x) p_X(x). \tag{2.13}$$

If we introduce the marginal probability density of Y

$$p_Y(y) = \int p_{XY}(x, y)\, dx, \tag{2.14}$$

from *Bayes formula* we obtain the conditional probability density of X for a given value y of Y

$$p_X(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_Y(y|x) p_X(x)}{p_Y(y)}. \tag{2.15}$$

If in this equation we insert the detected value y of the image, we obtain the *a posteriori* probability density of X

$$P_y^X(x) = p_X(x|y) = L_y^Y(x) \frac{p_X(x)}{p_Y(y)}. \tag{2.16}$$

Then, a **maximum a posteriori** (MAP) estimate of the unknown object is defined as any object x^* that maximizes the a posteriori probability density

$$x^* = \operatorname*{argmax}_{x \in \mathbb{R}^n} P_y^X(x). \tag{2.17}$$

As in the case of the likelihood it is convenient to consider the neglog function of $P_y^X(x)$. If we assume a Gibbs prior as that given in Equation (2.12) and we take into account the definition of Equation (2.3), we can introduce the following functional

$$\begin{aligned} J(x;\, y) = &-A \ln P_y^X(x) + B - A \ln Z \\ &- A \ln p_Y(y) = J_0(x;\, y) + \mu J_R(x), \end{aligned} \tag{2.18}$$

where $J_R(x) = A\Omega(x)$. This notation is introduced because the functional coming from the Gibbs prior is conceived as a *regularization functional*.

Therefore the MAP estimates are also given by

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^n} J(x;\, y) \tag{2.19}$$

and again one must look for the minimum points satisfying the non-negativity constraint.

We conclude by remarking that it is not obvious that a minimum point x^* of $J(x;\, y)$ is a sensible estimate $\bar{x}$ of the unknown object. In fact, in this formulation we have a free parameter μ (that, for analogy with regularization theory, we will call *regularization parameter*). In the classical regularization theory, a wide literature exists on the problem of the optimal choice of this parameter [16] but, as far as we know, this problem has not yet been thoroughly investigated in the more general framework provided by Bayesian regularization.

3 Scaled gradient methods based on a gradient decomposition

The formulations discussed in the previous section lead to the following general problem

$$\begin{aligned} &\text{minimize} \quad J(x;\, y) = J_0(x;\, y) + \mu J_R(x) \\ &\text{subject to} \quad x \geq 0, \end{aligned} \tag{3.1}$$

where $J_0(x;y)$ is coming from the neglog-likelihood function while $J_R(x)$ is coming from the neglog-prior function. As follows from the examples discussed above, we can assume that both are convex so that we have

a convex minimization problem. In the case $\mu = 0$ this is just the ML problem while in the case $\mu \neq 0$ it is a MAP problem.

As already remarked, the ML problem can not be treated as a standard optimization problem because it is ill-posed and we do not want to reach the minimum; on the other hand this is just what we want in the case of the MAP problem. Therefore it seems that it is necessary to use different methods in the two cases and, in fact, this is what people are usually doing. In this section we discuss an idea that can provide a unified approach to both problems. For simplicity, our presentation is heuristic.

Since $J(x; y)$ is convex all its minima are global. Then the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient conditions for a point x^* to be a minimum of $J(x; y)$

$$\begin{aligned} &x^* \nabla J(x^*; y) = 0, \\ &x^* \geq 0, \ \nabla J(x^*; y) \geq 0. \end{aligned} \tag{3.2}$$

Let us consider now the following decomposition of the gradient [30, 31]

$$-\nabla J(x; y) = U(x; y) - V(x; y); \ U(x; y) \geq 0, \ V(x; y) > 0. \tag{3.3}$$

It is obvious that such a decomposition always exists but is not unique. Different choices of the vectors U, V can be used and this non-uniqueness may be an advantage in some cases. However, the applicability of the approach is based on the fact that in all models that have been introduced for image reconstruction a natural decomposition of the gradient of this kind can be found, with explicit expressions of U, V. This point will become clear from the inspection of the examples that will be discussed in the following.

By assuming that we have selected a decomposition of the gradient in the previous form, then we can write the first KKT condition as a fixed point equation

$$x^* = T_y(x^*), \tag{3.4}$$

with

$$T_y(x) = x \frac{U(x; y)}{V(x; y)}. \tag{3.5}$$

The operator $T_y(\cdot)$ is well defined, since $V(x; y) > 0$. Moreover it is continuous if the functional $J(x; y)$ is continuously differentiable, as it is assumed, because, in such a case, it is possible to choose continuous functions U, V.

By applying the method of successive approximations we get the following iterative algorithm

- give $x^{(0)} > 0$
- given $x^{(k)}$ compute

$$x^{(k+1)} = x^{(k)} \frac{U(x^{(k)}; y)}{V(x^{(k)}; y)}. \tag{3.6}$$

About the convergence of the algorithm nothing can be said at this stage of the analysis since the operator $T_y(\cdot)$ is not in general a contraction. However we can remark some interesting features suggesting that it can deserve further considerations.

The first is that all the iterates are automatically non-negative. The second is that the algorithm is a scaled-gradient method, with step-size 1, since it can be written in the following form

$$x^{(k+1)} = x^{(k)} - S_k \nabla J(x^{(k)}; y) \tag{3.7}$$

where

$$S_k = \text{diag} \left\{ \frac{x_j^{(k)}}{V_j(x^{(k)}; y)} \right\}. \tag{3.8}$$

Remark 3.1. It is important to remark that, in [30, 31] the algorithm is presented as a descent method with a step-size selection. Indeed, it is written in the following form

$$x^{(k+1)} = x^{(k)} + \lambda_k \frac{x^{(k)}}{V(x^{(k)}; y)} \left\{ U(x^{(k)}; y) - V(x^{(k)}; y) \right\} \tag{3.9}$$

and the step-size $\lambda_k > 0$ is chosen in the following way. First an upper bound $\lambda_k^{(0)}$ is determined in order to ensure that $x^{(k+1)} \geq 0$. This is obtained by looking at the values of j such that $x_j^{(k)} > 0$ and $[\nabla J(x^{(k)}; y)]_j > 0$. If we denote by I_+ the set of these index values, then it is easy to see that

$$\lambda_k^{(0)} = \min_{j \in I_+} \left\{ \frac{V_j(x^{(k)}; y)}{V_j(x^{(k)}; y) - U_j(x^{(k)}; y)} \right\} \geq 1. \tag{3.10}$$

Next, the step-size λ_k is optimized by a line search in the interval $(0, \lambda_k^{(0)}]$ using, for instance, Armijo rule. In such a way convergence of the method is ensured.

A third property of the algorithm of Equation (3.6) is contained in the following proposition.

Proposition 3.2. *If the sequence of the iterates* $\{x^{(k)}\}$ *is convergent to* x^* *and if* $U(x; y) > 0$ *for any* $x > 0$*, then* x^* *solves* (3.1).

Proof. It is sufficient to prove that x^* satisfies the KKT conditions. The first one is satisfied because, thanks to the continuity of $T_y(\cdot)$, x^* is a fixed point of $T_y(\cdot)$. Moreover, the assumption $U(x; y) > 0$ for any $x > 0$ implies that all the iterates $x^{(k)}$ are strictly positive if $x^{(0)} > 0$, as one can easily prove by induction. It follows that $x^* \geq 0$. Therefore we have only to check that also the third KKT condition is satisfied.

It is certainly satisfied for all the values of the index such that $x_j^* > 0$, because it follows from the first condition. It is also satisfied if $x_j^* = 0$ and $U_j(x^*; y) = 0$ because, in such a case, the corresponding component of the gradient is strictly positive. Then, let us assume that it is not satisfied for a value of the index such that $x_j^* = 0, U_j(x^*; y) > 0$, *i.e.* let us assume that

$$x_j^* = 0, \frac{U_j(x^*; y)}{V_j(x^*; y)} > 1. \tag{3.11}$$

It follows that there exists k_0 such that, for any $k \geq k_0$, we have

$$\frac{U_j(x^{(k)}; y)}{V_j(x^{(k)}; y)} > 1. \tag{3.12}$$

Since all the iterates are strictly positive, we get $x_j^{(k+1)} > x_j^{(k)}$, in contradiction with the assumption that the limit of $x_j^{(k)}$ is zero. □

3.1 Maximum likelihood estimates

In Table 1 we give possible choices of the functions $U_0(x; y)$, $V_0(x; y)$ associated to the functionals $J_0(x; y)$ for the three noise models discussed in Section 2. It is obvious that it is possible to obtain other acceptable choices by adding, for instance, a suitable constant to both functions.

The interesting point is that, if we particularize the general algorithm of Equation (3.6) to the three noise models, we obtain three well-known algorithms proposed for image reconstruction.

Indeed, in the case of **Gaussian noise** we obtain

$$x^{(k+1)} = x^{(k)} \frac{H^T y}{H^T H x^{(k)} + b}, \tag{3.13}$$

and this is the image iterative space reconstruction algorithm (ISRA), introduced in [14], whose asymptotic convergence is proved in [15]. More precisely the original algorithm is with $b = 0$, but the proof of convergence can be easily extended to the case $b \neq 0$.

Table 1. The functions U_0, V_0 for the three noise models. The functions M, N of the third line are defined respectively in Equations (2.9) and (2.10), while h is defined in Equation (1.3).

	$U_0(x; y)$	$V_0(x; y)$
Gauss	$2H^T y$	$2(H^T Hx + b)$
Poisson	$H^T \dfrac{y}{Hx + b}$	h
Gauss+Poisson	$H^T \dfrac{N(x; y)}{M(x; y)}$	h

In the case of **Poisson noise** we obtain

$$x^{(k+1)} = \frac{x^{(k)}}{h} H^T \frac{y}{Hx^{(k)} + b}, \tag{3.14}$$

and this is the expectation maximization (EM) algorithm proposed in [41] and known as Richardson-Lucy (RL) algorithm in image deconvolution [33,39]. More precisely Equation (3.14) is the modified version of the algorithm introduced in [42] for taking into account background emission. In the case $b = 0$ several convergence proofs of the algorithm to a ML solution are available [24, 25, 29, 36, 48]; all these proofs utilize a nice property of the iterates, that holds true also for the minimum points x^* of the functional. If the matrix H is normalized in such a way that $h = 1$, then

$$\sum_{j=1}^{n} x_j^* = \sum_{j=1}^{n} x_j^{(k)} = \sum_{i=1}^{m} y_i. \tag{3.15}$$

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 $b \neq 0$ and the convergence of the algorithm does not appear to be proved in such a case.

Finally, in the case of **Gauss+Poisson noise** we obtain

$$x^{(k+1)} = \frac{x^{(k)}}{h} H^T \frac{N(x^{(k)}; y)}{M(x^{(k)}; y)}, \tag{3.16}$$

and this is the algorithm proposed in [42]. We recall that the functions M, N are defined respectively in Equations (2.9) and (2.10). As far as we know, also for this algorithm no convergence proof is available.

3.2 Maximum a posteriori estimates

In the case of a regularized functional, the general algorithm of Equation (3.6) takes the following form

$$x^{(k+1)} = x^{(k)} \frac{U_0(x^{(k)}; y) + \mu U_R(x^{(k)})}{V_0(x^{(k)}; y) + \mu V_R(x^{(k)})}, \tag{3.17}$$

where $U_0(x; y)$, $V_0(x; y)$ come from the likelihood while $U_R(x)$, $V_R(x)$ come from the prior. We will give examples of these functions in the next section, for different kinds of regularization functionals. Here we only remark that this algorithm has a very simple structure (linear fractional dependence of the iterates on the regularization parameter μ) suggesting an implementation based on an auxiliary function that can be called for different kinds of noise and regularization. In this way one can have a general algorithm working for all known kinds of likelihoods and priors. Unfortunately a convergence analysis is lacking.

4 Scaled gradient projection methods

If the reconstructed image has to be used for quantitative analysis (this may be required in domains of application such as Nuclear Medicine, Microscopy and Astronomy) it is important to guarantee flux conservation. In the case of zero background, this condition is given by Equation (3.15) and is automatically satisfied by the EM algorithm. However it is not satisfied in all the other cases so that it is important to introduce this condition as an additional constraint. We also remark that, in the case $b \neq 0$, Equation (3.15) must be modified for taking into account the effect of the background.

For simplicity we assume that the matrix H in normalized in such a way that $h = 1$, so that

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} H_{i,j} x_j \right) = \sum_{j=1}^{n} x_j. \tag{4.1}$$

Then, in the case $b \neq 0$, condition (3.15) is replaced by the following one

$$\sum_{j=1}^{n} x_j = \sum_{i=1}^{m} \{y_i - b_i\} \doteq c, \tag{4.2}$$

that will be called the *flux condition* or flux constraint.

If we introduce this additional constraint, the problem (3.1) is modified as follows

$$\begin{aligned} &\text{minimize} \quad J(x;y) = J_0(x;y) + \mu J_R(x) \\ &\text{subject to} \quad x \geq 0, \sum_{j=1}^{n} x_j = c. \end{aligned} \tag{4.3}$$

Again, in the case $\mu = 0$ we have ML problems, while in the case $\mu \neq 0$ we have MAP problems.

We denote by $\mathcal{C}$ the closed and convex set which is the intersection of the non-negative orthant with the affine subspace defined by the flux condition. We remark that $\mathcal{C}$ is compact so that any sequence contained in $\mathcal{C}$ will contain convergent subsequences.

In order to solve problem (4.3), a Scaled Gradient Projection (SGP) method is proposed in [10], which can be considered a generalization of the scaled gradient method (3.9).

The following notations are useful to describe the method. For a given vector $x \in \mathbb{R}^n$, we denote by $\|x\|_D$ the norm induced by the $n \times n$ symmetric positive definite matrix D, that is, $\|x\|_D = \sqrt{x^T Dx}$. Furthermore, for some given positive scalars c_1 and c_2, let $\mathcal{D}$ be the set of the $n \times n$ symmetric positive definite matrices D such that

$$c_1\|x\|^2 \leq x^T Dx \leq c_2\|x\|^2, \quad \forall\, x \in \mathbb{R}^n. \tag{4.4}$$

Finally, we denote by $P_{\mathcal{C},D}(x)$ the projection of $x \in \mathbb{R}^n$ over $\mathcal{C}$ in the norm $\|\cdot\|_D$, that is

$$P_{\mathcal{C},D}(x) = \operatorname*{argmin}_{z\in\mathcal{C}} \|z - x\|_D = \operatorname*{argmin}_{z\in\mathcal{C}} \left(\frac{1}{2} z^T Dz - z^T Dx\right). \tag{4.5}$$

The main steps of the SGP method are stated in Algorithm SGP. Several reasons make this approach appealing for solving problem (4.3). First of all, it is very simple: it belongs to the class of standard scaled gradient methods [6] with variable step-length α_k and non-monotone line-search strategy [8]. Secondly, due to the special constraints of the problem and to appropriate choices of D_k, the projection operation in Step 2 can be computationally non-expensive. Finally, the iterative scheme can achieve good convergence rate by exploiting the effective step-length selection rules recently proposed in literature.

Algorithm SGP: *Scaled Gradient Projection Method*

1. *Initialization.* Let $\alpha_{\min}, \alpha_{\max} \in \mathbb{R}$ be such that $0 < \alpha_{\min} < \alpha_{\max}$, $\beta, \gamma \in (0, 1)$ and let M be a positive integer.
   Set $x^{(0)} \in \mathcal{C}, \quad D_0 \in \mathcal{D}, \quad \alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$.

For $k = 0, 1, 2, \ldots$

2. *Projection.* Compute the descent direction
   $d^{(k)} = P_{\mathcal{C}, D_k^{-1}}(x^{(k)} - \alpha_k D_k \nabla J(x^{(k)}; y)) - x^{(k)}$.
3. *Line-search.* Set $\lambda_k = 1$ and $\bar{J} = \max_{0 \le j \le \min\{k, M-1\}} J(x^{(k-j)}; y)$.
   While $J(x^{(k)} + \lambda_k d^{(k)}; y) > \bar{J} + \gamma \lambda_k \nabla J(x^{(k)}; y)^T d^{(k)}$
   $\lambda_k = \beta \lambda_k$
   end.
   Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$.
4. *Update.* Define $D_{k+1} \in \mathcal{D}$ and $\alpha_{k+1} \in [\alpha_{\min}, \alpha_{\max}]$.

end

For the sake of completeness, we report some important details on the SGP implementation evaluated in this work. The choice of the scaling matrix D_k must avoid to introduce significant computational costs and, in particular, it must keep the projection $P_{\mathcal{C}, D_k^{-1}}(\cdot)$ in Step 2 computationally non-expensive. This can be done for example by using a diagonal scaling, so that the projection is obtained by solving a separable quadratic program for which efficient linear-time solvers can be used [12, 28]. According to the considerations of the previous section, we choose the following modification of the scaling matrix defined in (3.8)

$$D_k = \operatorname{diag}\left\{\max\left\{c_1, \frac{x_j^{(k)}}{V_j(x^{(k)}; y)}\right\}\right\}, \tag{4.6}$$

where $c_1 > 0$ is a prefixed threshold. Therefore (4.4) is satisfied, c_1 being just this threshold and $c_2 = c/\nu$, with c the flux constant and $\nu > 0$ such that

$$\nu = \min_j \left\{\min_{x \in \mathcal{C}}\{V_j(x; y)\}\right\}. \tag{4.7}$$

The line-search step of the SGP consists in a non-monotone strategy that uses successive reductions of λ_k to make $J(x^{(k+1)}; y)$ lower than the maximum of the objective function on the last M iterations [8, 21]. Of course, if $M = 1$ then the strategy reduces to the standard Armijo rule.

The updating rule for the step-length α_k is crucial for improving the convergence rate of the scheme; we use special step-length selections derived by the two Barzilai-Borwein (BB) rules [2], as usually done in many effective gradient methods [12, 13, 19, 40, 49]. In case of scaled gradient methods, by proceeding as for the derivation of the BB rules, we can regard the matrix $B(\alpha_k) = (\alpha_k D_k)^{-1}$ as an approximation of the Hessian $\nabla^2 J(x^{(k)}; y)$ and force a quasi-Newton property on $B(\alpha_k)$

$$\alpha_k^{\mathrm{BB1}} = \operatorname*{argmin}_{\alpha\in\mathbb{R}} \|B(\alpha)s^{(k-1)} - z^{(k-1)}\| \tag{4.8}$$

or

$$\alpha_k^{\mathrm{BB2}} = \operatorname*{argmin}_{\alpha\in\mathbb{R}} \|s^{(k-1)} - B(\alpha)^{-1}z^{(k-1)}\|, \tag{4.9}$$

where $s^{(k-1)} = \left(x^{(k)} - x^{(k-1)}\right)$ and $z^{(k-1)} = \left(\nabla J(x^{(k)}) - \nabla J(x^{(k-1)})\right)$; in this way, the following step-lengths are obtained

$$\alpha_k^{\mathrm{BB1}} = \frac{{s^{(k-1)}}^T D_k^{-1} D_k^{-1} s^{(k-1)}}{{s^{(k-1)}}^T D_k^{-1} z^{(k-1)}} \tag{4.10}$$

and

$$\alpha_k^{\mathrm{BB2}} = \frac{{s^{(k-1)}}^T D_k z^{(k-1)}}{{z^{(k-1)}}^T D_k D_k z^{(k-1)}}. \tag{4.11}$$

The step-length selection rule implemented within SGP is the $\mathrm{ABB}_{\mathrm{min1}}$ strategy proposed in [19], that consists in an adaptive alternation between the values

$$\bar{\alpha}_k = \max\{\alpha_{\min}, \min\{\alpha_{\max}, \alpha_k^{\mathrm{BB1}}\}\} \tag{4.12}$$

and

$$\hat{\alpha}_k = \max\{\alpha_{\min}, \min\{\alpha_{\max}, \alpha_k^{\mathrm{BB2}}\}\}. \tag{4.13}$$

A convergence analysis of the SGP method is carried out in [10] for the general case of the minimization of differentiable functions on closed convex sets. This analysis is based on several well-known technical results on gradient projection type methods [6, 8, 9] and it is not included in the present discussion. Here, we simply recall that when algorithm SGP is applied to problem (4.3), the following proposition may be derived from [10].

Proposition 4.1. *Let $\{x^{(k)}\}$ be the sequence generated by applying algorithm* SGP *to problem* (4.3)*. Every accumulation point x^* of $\{x^{(k)}\}$ is a constrained stationary point, that is*

$$\nabla J(x^*; y)^T (x - x^*) \geq 0 \quad \forall\, x \in \mathcal{C}. \tag{4.14}$$

If $J(x; y)$ is a convex function, then every accumulation point of $\{x^{(k)}\}$ is a solution of problem (4.3)*.*

The SGP algorithm has been implemented in the case of the deconvolution of 2D images corrupted by Poisson noise, without regularizing the neglog of the likelihood. Therefore, we have $\mu = 0$ in (4.3) and $J_0(x; y)$ is given in Equation (2.8). Moreover H is a block circulant matrix with circulant blocks so that both Hx and $H^T y$ can be computed by means of FFT. We recall that Hx is just the cyclic convolution of x with an array that is usually called the *point spread function* (PSF). We considered images obtained by convolving different objects with different PSFs, with the addition of a background. Moreover we also considered different noise levels (in the case of Poisson noise, this is obtained by changing the total number of counts). Our sample consisted of about 18 images. Detailed numerical results on this SGP implementation are reported in [10].

We evaluated the SGP behaviour on these test problems in comparison with the basic algorithm EM (or RL), as defined in (3.14), and the EM accelerated version proposed by Biggs and Andrews [7], that is implemented in the MATLAB code *deconvlucy* of the *Image Processing* toolbox (EM_MATLAB). We point out that, as far as we know, no convergence proof of the Biggs-Andrews algorithm is available. In all cases we obtained very similar results as concerns the computational efficiency. Therefore in Table 2 we only report results about a single image, sized 256×256, with three different noise levels.

Table 2. Behaviour of SGP, EM_MATLAB and EM.

	SGP			EM_MATLAB			EM		
c	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
4.43×10^9	339	0.185	63.7	388	0.185	64.9	3500*	0.186	458.9
7.02×10^8	108	0.187	20.3	141	0.187	23.0	3500*	0.187	447.4
4.43×10^7	20	0.195	4.0	46	0.195	7.7	414	0.194	54.5

In detail, in Table 2 we report the flux constant c controlling the different noise levels (we recall that, in the case of Poisson noise, the noise level is increasing when the total flux/number of counts is decreasing), the numbers of iterations required by the three methods (it_opt), the corresponding computational times in seconds (sec) and the relative reconstruction error (err_opt), defined as $\|x^{(k)} - x\|/\|x\|$, x being the object to be reconstructed. For each method, the results refer to the iteration where the minimum of the reconstruction error has been obtained, except for the cases marked with an asterisk in which the minimum is not reached within the maximum number of iterations allowed by the code. In Fig-

ure 1 we also give the behaviour of the relative reconstruction error as a function of the number of iterations for the three methods and for one noise level.

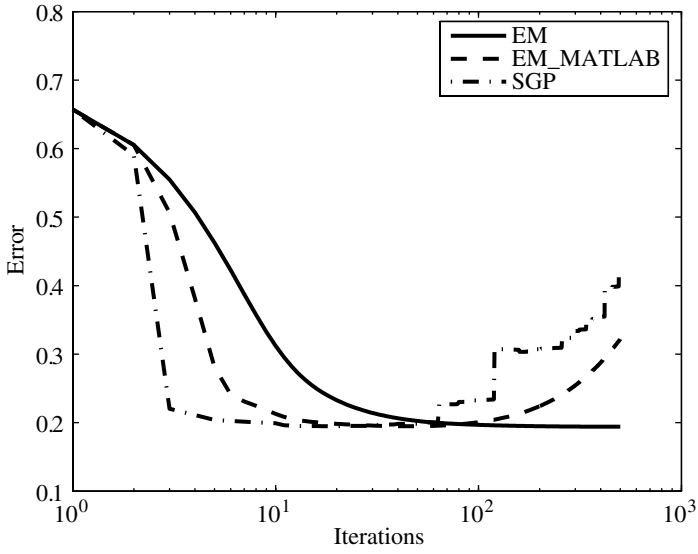

Figure 1. Relative reconstruction error for the case $c = 4.43 \times 10^7$.

We remark that the computational time per iteration in SGP and EM_MATLAB is approximately 40 % and 30 % greater than in EM, respectively. However, SGP and EM_MATLAB exhibit a better convergence rate in comparison with EM, and the time required to obtain a given reconstruction error in these accelerated methods is significantly lower than in EM. In these experiments SGP generally outperforms EM_MATLAB and in some cases it seems largely preferable.

It is also interesting to remark that, in the case of SGP, the reconstruction error drops to a value close to the minimum in very few iterations and that it remains close to this value for a large number of iterations, so that the choice of the optimal number of iterations does not seem to be critical in the case of real images.

The previous results indicate that SGP can be very efficient; therefore it looks quite natural to implement the method not only in the case of the algorithm ISRA, defined in (3.13), but also in the case of the regularized versions of ISRA and EM, as given in (3.17). To this purpose, we give in Table 3 the functions U_R, V_R for some of the most frequently used regularization functionals. Of course these functions provide a decomposition of the gradient of $J_R(x)$ in the sense explained in Section 3. We remark

Table 3. The functions U_R, V_R for different examples of regularization functionals. c is the flux constant of Equation (4.2), while p is a reference vector satisfying the flux condition. In the second line D is a matrix with non-negative entries and this example includes regularization in terms of the discrete Laplacian.

$J_R(x)$	$U_R(x)$	$V_R(x)$
$\frac{1}{2}\|\|x-p\|\|_2^2$	p	x
$\frac{1}{2}\|\|(I-D)x\|\|_2^2$	$(D+D^T)x$	$(I+D^TD)x$
$D_{\mathrm{KL}}(x;p)$	$\sum_{j\in R}\ln\frac{p_j}{c}$	$\sum_{j\in R}\ln\frac{x_j}{c}$
$\|\|x\|\|_1$	0	1

that, in the case of the entropy functional (third line), the flux constraint is essential for introducing a simple decomposition of the gradient.

Simple expressions of the U_R, V_R functions can be obtained also for other important regularization functionals such as total variation or Huber priors or functionals formed by the weighted sum of the squared first differences computed in each pixel. Non-convex regularization functionals, such as those proposed in [20] and [1], could also be considered in this approach. We also point out that, in the implementation of the SGP algorithm, only the gradient and the function V appear explicitly, as follows from (4.6).

Work is in progress in this direction. In particular, we intend to compare SGP with the very efficient second order methods such as quasi-Newton and interior point methods. We recall that these methods must be used with some care in the case of ML estimations because of the ill-posedness of these problems. However, they can certainly be used in the case of MAP estimations. If SGP can compete with these methods from the point of view of numerical efficiency, then it could provide a very useful and simple approach to iterative image reconstruction.

References

[1] B. Anconelli, M. Bertero, P. Boccacci, M. Carbillet, and H. Lanteri, *Iterative methods for the reconstruction of astronomical images with high dynamic range*, J. Comp. Appl. Math. **198** (2007), 321–331.

[2] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal. **8** (1988), 141–148.

[3] F. BENVENUTO, A. LA CAMERA, C. THEYS, A. FERRARI, H. LANTÉRI and M. BERTERO, *Study of an iterative method for the reconstruction of images corrupted by Poisson and Gauss noise*, submitted to Inverse Problems, 2007.

[4] M. BERTERO and G. A. VIANO, *On probabilistic methods for the solution of improperly posed problems*, Bollettino UMI **15-B** (1978), 453–508.

[5] M. BERTERO, C. DE MOL and E. R. PIKE, *Linear inverse problems with discrete data: I - General formulation and singular system analysis*, Inverse Problems **1** (1985), 300–330.

[6] D. P. BERTSEKAS, "Nonlinear Programming", Athena Scientific, Belmont, MA, 1999.

[7] D. S. C. BIGGS and M. ANDREWS, *Acceleration of iterative image restoration algorithms*, Applied Optics **36** (1997), 1766–1775.

[8] E. G. BIRGIN, J. M. MARTÍNEZ and M. RAYDAN, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM Journal on Optimization **10:4** (2000), 1196–1211.

[9] E. G. BIRGIN, J. M. MARTÍNEZ and M. RAYDAN, *Inexact spectral projected gradient methods on convex sets*, IMA J. Numer. Anal. **23** (2003), 539–559.

[10] S. BONETTINI, R. ZANELLA and L. ZANNI, *A scaled gradient projection method for constrained image deblurring*, Tech. Rep. Department of Mathematics, University of Modena and Reggio Emilia **78** (2007).

[11] J. CANNON and K. MILLER, *Some problems in numerical analytic continuation*, J. Soc. Indust. Appl. Math. Ser. B Numer. Anal. **2** (1965), 87–96.

[12] Y. H. DAI and R. FLETCHER, *New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds*, Mathematical Programming (Series A) **106:3** (2006), 403–421.

[13] Y. H. DAI, W. W. HAGER, K. SCHITTKOWSKI and H. ZHANG, *The Cyclic Barzilai-Borwein Method for Unconstrained Optimization*, IMA J. Numer. Anal. **26** (2006), 604–627.

[14] M. E. DAUBE-WITHERSPOON and G. MUEHLLEHNER, *An iterative image space reconstruction algorithm suitable for volume ECT*, IEEE Trans. Med. Imaging **5** (1986), 61–66.

[15] A. R. DE PIERRO, *On the convergence of the iterative image space reconstruction algorithm for volume ECT*, IEEE Trans. Med. Imaging **6** (1987), 174–175.

[16] H. W. ENGL, M. HANKE and A. NEUBAUER, "Regularization of Inverse Problems", Kluwer Academic Publishers, Dordrecht, 1996.

[17] M. FOSTER, *An application of the Wiener-Kolmogorov smoothing theory to matrix inversion*, J. Soc. Indust. Appl. Math. **9** (1961), 387–392.

[18] J. N. FRANKLIN, *Well-posed stochastic extensions of ill-posed linear problems*, J. Math. Anal. Appl. **31** (1970), 682–716.

[19] G. FRASSOLDATI, L. ZANNI and G. ZANGHIRATI, *New adaptive step-size selections in gradient methods*, J. of Industrial and Management Optimization, to appear, 2007.
Available at <http://www.optimization-online.org>.

[20] S. GEMAN and D. MC CLURE, *Statistical methods for tomographic image reconstruction*, In: "Proc 46th Session of the ICI", Bulletin of the ICI, 1987, 5–21.

[21] L. GRIPPO, F. LAMPARIELLO and S. LUCIDI, *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis **23** (1986), 707–716.

[22] J. HADAMARD, *Sur les problèmes aux dérivées partielles et leur signification physique*, Bull. Univ. Princeton **13** (1902), 49–52.

[23] J. HADAMARD, "Lectures on Cauchy's problem in Linear Partial Differential Equations", Yale Univ. Press, New Haven, 1923.

[24] A. N. IUSEM, *Convergence analysis for a multiplicatively relaxed EM algorithm*, Math. Methods Appl. Sci. **14** (1991), 573–593.

[25] A. N. IUSEM, *A short convergence proof of the EM algorithm for a specific Poisson model*, REBRAPE **6** (1992), 57–67.

[26] V. K. IVANOV, *On linear problems which are not well-posed*, Soviet Math. Dokl. **3** (1961), 981–983.

[27] F. JOHN, *Continuous dependence on data for solutions of partial differential equations with a prescribed bound*, Comm. Pure Appl. Math. **13** (1960), 551–585.

[28] K. C. KIWIEL, *Breakpoint searching algorithms for the continuous quadratic knapsack problem*, Mathematical Programming, 2007, to appear.

[29] K. LANGE and R. CARSON, *EM reconstruction algorithms for emission and transmission tomography*, J. Computer Assisted Tomography **8** (1984), 306–316.

[30] H. LANTÉRI, M. ROCHE, O. CUEVAS and C. AIME, *A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints*, Signal Process. **81** (2001), 945–974.

[31] H. LANTÉRI, M. ROCHE and C. AIME, *Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms*, Inverse Problems **18** (2002), 1397–1419.

[32] J. LLACER and J. NUÑEZ, *Iterative maximum likelihood and bayesian algorithms for image reconstruction in astronomy*, In: "The Restoration of HST Images and Spectra", R. L. White and R. J. Allen (eds.), The Space Telescope Science Institute, Baltimore, 1990, 62–69.

[33] L. B. LUCY, *An iterative technique for the rectification of observed distributions*, Astron. J. **79** (1974), 745–754.

[34] K. MILLER, *Least squares methods for ill-posed problems with a prescribed bound*, SIAM J. Math. Anal. **1** (1970), 52–74.

[35] H. N. MÜLTHEI and B. SCHORR, *On an iterative method for a class of integral equations of the first kind*, Math. Methods Appl. Sci. **9** (1987), 137–168.

[36] H. N. MÜLTHEI and B. SCHORR, *On properties of the iterative maximum likelihood reconstruction method*, Math. Methods Appl. Sci. **11** (1989), 331–342.

[37] H. N. MÜLTHEI, *Iterative continuous maximum likelihood reconstruction methods*, Math. Methods Appl. Sci. **15** (1993), 275–286.

[38] C. PUCCI, *Sui problemi di Cauchy non "ben posti"*, Atti Acc. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. **18** (1955), 473-477.

[39] W. H. RICHARDSON, *Bayesian based iterative method of image reconstruction*, J. Opt. Soc. Am. **62** (1972), 55–59.

[40] T. SERAFINI, G. ZANGHIRATI and L. ZANNI, *Gradient projection methods for quadratic programs and applications in training support vector machines*, Optimization Methods and Software **20** (2005), 353–378.

[41] L. A. SHEPP and Y. VARDI, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans. Med. Imaging **1** (1982), 113–122.

[42] D. L. SNYDER, A. M. HAMMOUD and R. L. WHITE, *Image recovery from data acquired with a charged-coupled-device camera*, J. Opt. Soc. Am. **A 10** (1993), 1014–1023.

[43] O. N. STRAND and E. R. WESTWATER, *Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind*, J. Assoc. Comput. Mach. **15** (1968), 100–114.

[44] A. N. TIKHONOV, *On the stability of inverse problems*, Dokl. Akad. Nauk. SSSR **39** (1943), 195–200.

[45] A. N. TIKHONOV, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl. **4** (1963), 1035–1038.

[46] A. N. TIKHONOV and V. Y. ARSENIN, "Solutions of ill-posed problems", Wiley, New York, 1977.

[47] V. F. TURCHIN, V. P. KOZLOV and M. S. MALKEVICH, *The use of mathematical-statistics methods in the solution of incorrectly posed problems*, Soviet Phys. Uspekhi **13** (1971), 681–703.

[48] Y. VARDI, L. A. SHEPP and L. KAUFMAN, *A statistical model for positron emission tomography*, J. Am. Stat. Ass. **80** (1985), 8–37.

[49] B. ZHOU, L. GAO and Y. H. DAI, *Gradient methods with adaptive step-sizes*, Computational Optimization and Applications **35** (2006), 69–86.