THE BAYESIAN MAXIMUM ENTROPY APPROACH TO IMAGE IMPROVEMENT
PART II
A PRACTICAL ALGORITHM FOR IMPROVING X-RAY DIGITAL MAMMOGRAMS

RADU MUTIHAC*, ALBERTO A. COLAVITA
ICTP-INFN-UNU Microprocessor Laboratory, Trieste 34100, Italy
(Received March 27, 1997)

Abstract. An efficient iterative algorithm for maximum entropy image improvement from large data sets based on the conjugate gradient method and Lagrange multipliers in nonlinear optimisation of a specific potential function has been developed. The point spread function (PSF) has been defined following numerical simulations of homogeneous mammalian tissue with microcalcification inclusions of various opacities. Digitised mammograms have been subject of comparative analyses in terms of signal-to-noise ratio, contrast, and number of distinct patterns.

Key words: conjugate gradient method, point spread function, Lagrange multipliers.

1. INTRODUCTION

Mammography is currently the best radiological technique for early detection of breast cancer. An efficient and robust algorithm has been developed to process digitised X-ray mammographic images in order to detect subtle abnormalities, such as different kind of lesions (microcalcifications, opacities or stellate patterns of straight lines), which may indicate the presence of malignancies. The study is intended to improve the diagnosis of mammographic lesions by using computer image analysis methods to derive quantitative description of their visual characteristics like edge definition, extent and texture. Though experienced radiologists may often establish the malignant or benign nature of a lesion on the basis of these characteristics, the image perception by human observers is subject to complex interpretation by the visual system. The consequence is that the perceived image does not always correspond to the actual data contained in the image, which might lead to inaccurate or incorrect conclusions [1].

Pattern recognition methods can be further developed for computer assisted diagnosis and finally, expert systems may come out to perform all tasks required by digital mammography and computed tomography (CT).

* Permanent address: Faculty of Physics, University of Bucharest, Romania.
II. IDEAL IMAGING PROCESSES

1. PHYSICAL BACKGROUND

The basic unit of information in an imaging system is the detected photon. The physical image corresponds to the recorded number of photons impinging on some photosensitive device at specified locations. Photon emission is a Poissonian process so that the number of detected photons will be distributed according to Poisson statistics. Therefore, if in a unit time and per unit area one counts an average number $n$ of photons, the standard deviation will be $\sqrt{n}$. This fluctuation in a single measurement is called quantum noise. It is intrinsic to the mechanism of photon counting and therefore one may speak of a quantum limited image.

An ideal imaging device can be thought as a spatial array of individual photon receptors and counters, each having identical properties. It is assumed that each receptor can receive light quanta (photons) independent of its neighbouring receptors. The image state of a detector is completely determined by the number of quanta it receives and records, and each detector can be in one of a finite number of distinguishable image states. Since the possible number of states of a detector is finite, after a detector attains the final state, all other additional incident quanta will remain unrecorded, that is the detector gets saturated.

In case of a digital image system, each pixel can be viewed as a receptor. The spatial resolution of the system depends on the spatial size of the pixel, whereas each pixel can have only a finite number of states until reaching its saturation level. The observed grey level of a pixel is the consequence of the received quanta by the corresponding receptor. Up to a certain level, the larger the number of the recorded quanta, the larger the grey value.

2. MAIN CHARACTERISTICS OF AN X-RAY IMAGE

The knowledge of the dose and the energy spectrum of the X-rays delivered to the subject during exposure allows in principle the computation of photon number per unit surface of the image system. Let us consider a sample to be imaged by means of a detector whose surface $A$ is subdivided in identical square pixels of side $d$ [2]. Let us denote by $n_b$ the number of photons incident on a pixel looking at the “background”, and $n_t$ the number of photons incident on a pixel containing some texture to be imaged assuming that a percentage $p$ of the original incident photons has been absorbed, that is $n_d = (1 - p)n_b$. Further, let us define the signal as $n_b - n_d = pn_b$ and the noise as $\sqrt{n_b + n_d} = \sqrt{n_b(2 - p)}$. Then it only appears natural to express the contrast $c$ as
The Bayesian maximum entropy (II)

$$c = \frac{n_b - n_d}{n_b + n_d} = \frac{2p}{2-p}$$  \hspace{1cm} (2.1)

while the signal-to-noise ratio $k$ can be given by

$$k = \frac{pn_b}{\sqrt{n_b(2-p)}}.$$  \hspace{1cm} (2.2)

Hence the number of background photons required to reveal a certain target detail with a given contrast $c$ is $n_b = 2k^2/pc$. If $p < 0.2$ then $c \approx p$ and it results $n_b \approx 2(k/c)^2$. The result holds for an imaging system that is completely noise-free, i.e., the target is visible against a smooth background.

Image noise degrades the overall quality of an image and makes it difficult to reveal certain details of interest. The major source of noise in X-ray and nuclear imaging is the random distribution of photons over the surface of the image. The amount of noise, or variation in photon concentration from area to area (e.g., from one detector cell to another) is inversely related to the number of photons used to form the image. Therefore, by increasing the total number of photons, the noise of an image can be reduced. It is important that the signal-to-noise ratio be large enough so that in a set of pixels the probability of any pixel giving accidentally a false signal may be negligible. In the case of $10^5 \div 10^6$ pixels the condition is accomplished by setting $k = 5$ [3]. If $N$ photons are incident on the surface $A$ of the detector, then $N/n_b = A/d^2$ and it follows that

$$c = \frac{\sqrt{2k}}{d\sqrt{N/A}}.$$  \hspace{1cm} (2.3)

Inserting typical numerical values in eq. (2.3) such as $N/A = 4 \times 10^6$ photons/mm$^2$, $k = 5$, and the pixel side of our detector $d = 0.2$ mm$^2$, it comes out the contrast value $c = 1.7 \times 10^{-2}$. This value is extremely low since the practical rule is currently setting the minimum contrast perceivable by the human eye to be of about 2.5% over a smooth background. From the statistical point of view, the contrast value obtained according to eq. (2.3) infers that the number of photons per unit surface employed in a conventional radiological examination is quite excessive. Consequently, a detector with near-unity efficiency and capable of single photon counting would require less photons for the same quality of radiological images and, therefore, a lower radiation dose for the subject as well.
The signal-to-noise ratio in the case of soft tissues, as in mammography for instance, is a function of energy and Monte-Carlo simulations [4] indicated a maximum to be found in the range of 20 to 25 keV. However, our numerical simulations have considered homogeneous mammalian tissue of 5 cm thickness with microcalcification inclusions and various opacities, while the X-ray energy has been set throughout at the level of 18 keV [5].

3. DIGITAL DETECTORS

Suppose that a detector has 100% efficiency and only the quantum noise is taken into account. Then eq. (2.3) holds for the smallest detail to be imaged down to the dimension of one pixel. For objects which are different from the pixel size, a more complex equation must be used

\[ C_m = \frac{1}{\sqrt{N/A}} \frac{k(a)}{\sqrt{a}} \]  

(2.4)

where \( k(a) \) is a rather complex function of the object area, and \( C_m \) is the minimum contrast which can be detected using a given number of photons per unit area and which could ideally be achieved with a 100% efficient noiseless detector [6].

Both of the above expressions, eqs. (2.3) and (2.4), show that contrast varies inversely with the square root of the superficial dose (which is proportional to the number of photons per unit surface impinging on the detector). Qualitatively, if a small object is to be detected, then it must have a large contrast, whereas large objects may be imaged with a relatively smaller contrast. However, practical detectors with their electronic parts included have an efficiency less than 100%, thereby reducing \( N/A \) and adding noise to the photon counters. If \( R \) is the ratio between the total noise (including the detector noise) and the intrinsic quantum noise, then the minimum detectable contrast becomes

\[ C = \frac{R}{\sqrt{\varepsilon}} C_m \]  

(2.5)

where \( \varepsilon \) is the detector efficiency. The last expression (2.5) shows that both a poor efficiency (small values of \( \varepsilon \)) and a noisy detector (small values of \( R \)) yield a minimum detectable contrast higher than the ideal case \( C_m \), degrading therefore the performance of the detection system.

Experimental investigations have revealed the superiority of digital detectors over conventional X-ray tubes and films as detectors in two directions at least [2]. The phantoms routinely used as test objects in mammography are already visible at a fluence by 3 orders of magnitude lower by the digital recorders. An increase in fluence corresponds to an increase in image contrast which means an increase in object visibility. Secondly, once a digital detector yielded an im-
age as a set of numerical data, the grey scale can be adapted as to enhance the contrast, and even more, tones of greys can be substituted by false colours.

III. DERIVATION OF THE POTENTIAL FUNCTION $Z$

With a space invariant linear system with additive measurement errors (noise), so that

$$g = Rf + b$$

(3.1)

a $\chi^2$-constraint for handling the errors $b_m$, $m = 1, 2, \ldots, M$, is natural [7]. Hence

$$\chi^2 (f) = \sum_{i=1}^{M} \sum_{j=1}^{M} \left[ g_i - \sum_{n=1}^{N} R_{in} f(t_n) \right] S^{-1} \left[ g_j - \sum_{n=1}^{N} R_{jn} f(t_n) \right]$$

(3.2)

$$\chi^2 (f) = (g - R f)^T S^{-1} (g - R f)$$

(3.3)

where $S^{-1}$ is the inverse of the covariance matrix. If noise is independent among pixels, then the covariance and the correlation vanish, so that

$$\text{cov}[b_k, b_m] = S_{km} = \sigma_k \sigma_m \delta_{km}, \quad k, m = 1, 2, \ldots, M$$

(3.4)

and the covariance matrix becomes diagonal, $S_{mm} = \sigma_m^2, \quad m = 1, 2, \ldots, M$, leading to

$$\chi^2 (f) = \sum_{m=1}^{M} \left( \frac{g_m - \sum_{n=1}^{N} R_{mn} f(t_n)}{\sigma_m} \right)^2 = \sum_{m=1}^{M} \frac{b_m^2}{\sigma_m^2}.$$  

(3.5)

The task is to reconstruct an image in the form of a probability distribution $f = \{p_n\}$, where $p_n = f_n / F$, $n = 1, 2, \ldots, N$, given the measured “blurred” data $g_m$, $m = 1, 2, \ldots, M$, affected by the errors $b_m$, $m = 1, 2, \ldots, M$, and which complies to a finite set of constraints

$$\sum_{n=1}^{N} p_n = 1$$

(3.6)

$$g_m = \sum_{n=1}^{N} F \cdot R_{mn} \cdot p_n + b_m$$

(3.7)

$$\sum_{m=1}^{M} \frac{b_m^2}{\sigma_m^2} = \Omega$$

(3.8)
where we denoted by $\Omega$ the expected value of statistical misfit $\chi^2$. While the restrictions (3.6) and (3.7) are quite evident, the last equation (3.8) explicitly assumes that the errors $b_m$, $m = 1, 2, \ldots, M$ are normally distributed with zero mean, $\overline{b} = 0$, and variances $\sigma_m^2$, $m = 1, 2, \ldots, M$. Generally, the chi-square statistic associated with $b_m$, $m = 1, 2, \ldots, M$ independent Gaussian random variables

$$\chi^2 = \frac{\sum_{m=1}^{M} (b_m - \overline{b})^2}{\sigma_m^2}$$

(3.9)

has the expectation $M \pm \sqrt{M}$. Since the mean is fitted by setting $\overline{b} = \frac{1}{M} \sum_{m=1}^{M} b_m = 0$ which is inferred by the conservation of the number of quanta spread over the pixels, one degree of freedom has been already used up. Consequently, we reduce the expectation of $\chi^2$ by one and set $\Omega = M - 1$. It is common for this distinction to be ignored, but in some cases, such as interpolation where the number of free parameters is similar to the number of data points, it is essential to make the difference [8]. However, the procedure is effective if reliable estimates of the standard errors $\sigma_m^2$, $m = 1, 2, \ldots, M$ are a priori selected on some theoretical and/or experimental base.

Among all admissible probability distributions of the true image, $f = \{p_n\}$, that satisfy the system of eqs. (3.6), (3.7), and (3.8), the one which complies with the ME principle is the most unbiased selection to estimate the true image. Accordingly, maximising the image entropy

$$S(f) = -\sum_{n=1}^{N} f_n \ln (f_n/F) = -F \sum_{n=1}^{N} p_n \ln p_n$$

(3.10)

leads to the smoothest and uniform distribution, $\overline{f}$, among the admissible set of images.

The Lagrangean associated with the objective function (the image entropy) and the specific constraints (3.6), (3.7), and (3.8), may be exhaustively written in the form

$$L(p_1, p_2, \ldots, p_N, b_1, \ldots, b_M, \lambda_0, \ldots, \lambda_M, \rho) =$$

$$= -\sum_{n=1}^{N} Fp_n \ln p_n - \lambda_0 \left( \sum_{n=1}^{N} p_n - 1 \right) -$$

$$- \sum_{m=1}^{M} \lambda_m \left( \sum_{n=1}^{N} FR_m p_n + b_m - g_m \right) - \rho \left( \sum_{m=1}^{M} \frac{b_m^2}{\sigma_m^2} - \Omega \right)$$

(3.11)
The Bayesian maximum entropy (II) 25

where $\lambda_0, \lambda_1, \ldots, \lambda_M$ and $\rho$ are the Lagrange multipliers and the errors, $b_m$, $m = 1, 2, \ldots, M$, were included in the argument since they were evidently unknown. Applying the Lagrange method by setting $\partial L / \partial p_n = 0$, $n = 1, 2, \ldots, N$ and $\partial L / \partial b_m = 0$, $m = 1, 2, \ldots, M$, we obtain the following set of $N + M$ equations, respectively,

$$\ln p_n = -1 - \lambda_0 / F - \sum_{m=1}^{M} \lambda_m R_{mn}, \quad n = 1, 2, \ldots, N \tag{3.12}$$

$$b_m = -\lambda_m \frac{\sigma_m^2}{2\rho}, \quad m = 1, 2, \ldots, M. \tag{3.13}$$

The Lagrangean arguments, $p_1, \ldots, p_N$, $b_1, \ldots, b_M$, $\lambda_0$, $\lambda_1, \ldots, \lambda_M$, $\rho$, that is $N + 2M + 2$ unknown variables, are now perfectly matched by the number of equations (3.6), (3.7), (3.8), (3.12) and (3.13). By solving eq. (3.12) for $p_n$ and using the total flux constraint (3.6) to eliminate $\lambda_0$, each $p_n$ can be expressed as a function of $\lambda_1, \lambda_2, \ldots, \lambda_M$

$$p_n = p_n(\lambda_1, \lambda_2, \ldots, \lambda_M) = \frac{\exp\left(-\sum_{m=1}^{M} \lambda_m R_{mn}\right)}{\sum_{n=1}^{N} \exp\left(-\sum_{m=1}^{M} \lambda_m R_{mn}\right)}, \quad n = 1, 2, \ldots, N. \tag{3.14}$$

By substitution of the eqs. (3.14) and (3.13) into the remaining equations (3.7) and (3.8), respectively, we get the following non-linear system of $M + 1$ equations with $M + 1$ unknown, $\lambda_1, \lambda_2, \ldots, \lambda_M$, $\rho$, namely

$$\sum_{n=1}^{N} F R_{mn} p_n(\lambda_1, \lambda_2, \ldots, \lambda_M) - \lambda_m \frac{\sigma_m^2}{2\rho} - g_m = 0, \quad m = 1, 2, \ldots, M \tag{3.15}$$

$$\frac{1}{4\rho^2} \sum_{m=1}^{M} \sigma_m^2 \lambda_m^2 - \Omega = 0$$

The left hand side of the nonlinear system (3.15) is the gradient of the potential function

$$Z(\lambda_1, \lambda_2, \ldots, \lambda_M, \rho) = -F \cdot \ln \left[\sum_{n=1}^{N} \exp\left(-\sum_{m=1}^{M} \lambda_m R_{mn}\right)\right] - \frac{1}{4\rho^2} \sum_{m=1}^{M} \lambda_m^2 \sigma_m^2 - \sum_{m=1}^{M} \lambda_m g_m + \rho \Omega. \tag{3.16}$$
Hence the whole problem may be put in terms of finding the extremum of the potential function $Z$ in the $M + 1$ dimension space of the Lagrange multipliers, so that the system (3.15) is equivalent with solving

$$
\nabla Z(\lambda_1, \lambda_2, \ldots, \lambda_M, \rho) = 0.
$$

(3.17)

The higher accuracy in solving numerically the system (3.15) is demanded, the more stringent the memory requirements should be. Wilczek and Drapatz [9] suggested Newton’s iteration method as offering high accuracy results. Though at each step of the iteration the Jacobian of the system (3.15) has to be evaluated, the computing requirements may be significantly reduced by exploiting its symmetric structure. Ortega and Rheinbolt [10] adopted a so-called continuation technique for the very few cases where Newton’s method fails to converge. Basically, the chi-square function is considered as a parameter and the iteration starts with such of its value that the last equation in our system (3.15) be (almost) satisfied; then it is gradually adjusted to its originally intended value $\Omega$.

IV. MULTIDIMENSIONAL OPTIMIZATION

The techniques presented above are practically successful for relatively small data sets only. In our approach addressed to large data sets, we minimised the potential function $Z$ by numerical calculus of highly nonlinear equations. Unfortunately, there is no perfect optimisation algorithm and comparative approaches are necessary to determine the most effective method for a specific problem. We shall hereafter refer to the optimisation of a given function by using the term of minimisation.

There are two major families of algorithms for multidimensional minimisation in the $N$-dimensional space, which assume the calculation of first derivatives. Both of them require a one-dimensional minimisation sub-algorithm which may or may not use the derivative information. The first family is known under the name of conjugate gradient algorithms, the most typical being Fletcher-Reeves and Polak-Ribiere algorithms. Conjugate gradient algorithms only require order a few times $N$ storage. The second family goes under such names as quasi-Newton or variable metric methods as typified by the David-Fletcher-Powell (DFP) algorithm or the closely related Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. These methods require order $N^2$ storage. Since both families imply successive line minimisations, they are generally referred to as direction set methods. There is no definite evidence in favour of one out of the two.

The central problem for a direction set (search) method is to come up with a set of $N$ linearly independent, mutually conjugate directions. Then, one pass of $N$ line minimisations will put it exactly at the minimum of a quadratic form. For functions that are not exactly quadratic forms, it won’t be exactly at the mini-
mum; but repeated cycles of $N$ line minimisations will in due course converge quadratically to the minimum.

V. CONJUGATE GRADIENT METHODS

Let $Z : \mathbb{R}^{M+1} \rightarrow \mathbb{R}$, $P \in \mathbb{R}^{M+1}$ and both $Z(P)$ and $\nabla Z(P)$ can be calculated. Methods which provide iteratively new directions of minimisation not down the new gradient but rather along the lines constructed to be conjugated to the old gradient and, insofar as possible, to all previous directions traversed are generally called conjugate gradient methods. The space of $M + 1$ dimension is corresponding to the Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_M, \rho$.

The conjugate gradient algorithm is applied to the minimisation of arbitrary nonlinear functions $Z$ which are able to be approximated by quadratic forms like

$$Z(x) = Z(P) + \sum_{i=1}^{M+1} \frac{\partial Z}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j=1}^{M+1} \frac{\partial^2 Z}{\partial x_i \partial x_j} x_i x_j + \ldots \approx Z(P) - b \cdot x + \frac{1}{2} x \cdot A \cdot x$$

(5.1)

where

$$b = -\nabla Z|_P \quad \text{and} \quad A_{ij} = \frac{\partial^2 Z}{\partial x_i \partial x_j}|_P, \quad i, j = 1, 2, \ldots, M + 1$$

(5.2)

and the matrix $A = \left\| A_{ij} \right\|_{j=1,2,\ldots,M+1}$ is called the Hessian matrix of the function $Z$ at $P$.

We start with an arbitrary initial vector $g_0$ and letting $h_0 = g_0$, then the conjugate gradient method constructs two sequences of vectors by the recurrence

$$g_{i+1} = g_i - \lambda_i A \cdot h_i \quad \text{and} \quad h_{i+1} = g_{i+1} + \gamma_i h_i, \quad i = 1, 2, \ldots$$

(5.3)

The vectors satisfy the orthogonality and conjugacy conditions

$$g_i \cdot g_j = 0, \quad h_i \cdot A \cdot h_j = 0, \quad g_i \cdot h_j = 0, \quad i \neq j.$$

(5.4)

The scalars $\lambda_i$ and $\gamma_i$ are given by

$$\lambda_i = \frac{g_i \cdot g_i}{h_i \cdot A \cdot h_i} \quad \text{and} \quad \gamma_i = \frac{g_{i+1} \cdot g_{i+1}}{g_i \cdot g_i}$$

(5.5) \hspace{1cm} (5.6)

Suppose we have $g_i = -\nabla Z(P_i)$ for some point $P_i$ where $Z$ can be expressed as in the form (5.1). We proceed from $P_i$ along the direction $h_i$ to the
local minimum of $Z$ located at some point $P_{i+1}$ and then set $g_{i+1} = -\nabla Z(P_{i+1})$. It turns out that $g_{i+1}$ is the same vector as it would have been constructed by equation (5.3), though we have constructed it without the knowledge of $A$. We have then the basis of an algorithm similar to Fletcher-Reeves which requires neither knowledge of the Hessian matrix $A$, nor even the storage capacity necessary to store such a matrix $A$. The sequence of directions $h_i$ is constructed using only line minimisations and an auxiliary vector to store the latest in the sequence of $g$’s.

A further refined ingredient consists in a minor modification (Polak-Ribiére), namely

$$\gamma_i = \frac{(g_{i+1} - g_i) \cdot g_{i+1}}{g_i \cdot g_i}$$

which is a similar orthogonality condition as eq. (5.6). Practical functions are generally not exactly quadratic forms and one may still want to proceed for another set of iterations after arriving at the supposed minimum of an alleged quadratic form. There is evidence [11] that the last variant (5.7) accomplishes more elegantly the transitions to further iterations in the sense that it tends to reset $h$ to be down the local gradient, which is equivalent to restarting the conjugate gradient algorithm.

VI. RESULTS AND DISCUSSION

1. IMAGE NOISE

The expression of the PSF has been derived from simulations performed in our laboratory with parameters as close as possible to the mammalian tissue and featuring the characteristics of our incoming solid-state original detector. A 25-pixel map describes the errors in our image processing system (Table 1). A proper and realistic PSF is fundamentally related with the overall image quality which is based on quite subjective evaluations.

The error range describes how far a single measurement value might deviate, or miss, the true count value of a sample. Error ranges can be expressed in units, or increments of standard deviations. For values distributed in a Gaussian manner, the relationship between the probability of a value falling within a specific narrow range remains constant when expressed in terms of standard deviations. The confidence level expresses the probability, or chance, that a single measurement will fall within a specific error range. The relationship between confidence level and error range expressed in standard deviations does not change for measurement values distributed in a Gaussian manner.

The standard deviation of the count densities (photon concentrations) is the best quantitative estimator of the noise in an image. Theoretically, the value of the
The Bayesian maximum entropy (II)

The significance is that the precision of a radiation measurement increases as the number of counts in an individual measurement is increased. Though we have assumed Gaussian errors with zero mean value, it is appropriate to express the photon fluctuations in each pixel in terms of the Poissonian standard deviation as long as the quantum noise is dominant. The more efficient the detector is, the better this statement holds. In photon distributions, the square root of the average number of photons per pixel provides a close estimate of the standard deviation value in each pixel [12]. Therefore, we have used throughout the following expressions for the noise standard deviations.
\( \sigma_m = \alpha \cdot (g_m)^{1/2}, \quad m = 1, 2, \ldots, M \)  

(6.1)

where \( g_{\text{max}} \) denotes the maximum number of photons in any pixel \( g_m, m = 1, 2, \ldots, M \). According to our experiments, the convergence of the algorithm depends on the signal-to-noise ratios in each pixel \( g_m/\sigma_m, m = 1, 2, \ldots, M \), which set the values of the constant \( \alpha \) within the range \( 10^{-6} \pm 10^{-4} \). Lower values of \( \alpha \) are related with better contrast and higher brightness.

2. IMAGE ENTROPY

The expression derived for the entropy of an image represents the conformational entropy of the image and it is quite different in form and meaning from the informational entropy introduced by Shannon in statistics or the thermodynamic entropy of a beam of photons. However, there are still debates on the correct definition for the entropy of an image and the particular mathematical functions to describe it.

3. ALGORITHM IMPLEMENTATION

We have derived an efficient algorithm for minimising the potential function \( Z \) by generalisation of the conjugate gradient method for nonlinear equations with positive constraints and using Lagrange multipliers method. As long as practical images contain a large number of unknown pixels, say \( N = 10^4 \pm 10^6 \) and even more, it is essential that the number of storage locations needed by the algorithm be only of order \( N \). Accordingly, variable metric methods or other techniques which require storage locations of order \( N^2 \) are to be discarded from the very beginning. Our algorithm is a trade-off between the task of processing large data sets and the computational time and power requirements. It contains flavours from the algorithms developed by Fletcher-Reeves, Polak-Ribière, and Cornwell-Krueger [13] in the form of some adapted subroutines presented in Numerical Recipes in C [14], and A Numerical Library in C for Scientists and Engineers [15].

Basically, the algorithm is aiming to minimise the potential function \( Z \) in an \( M \)-dimensional space using gradient information as well. It starts from an approximate solution as a function of the Lagrange multipliers for the predicted image, which is chosen as perfectly plain. While the initial values of the Lagrange multipliers provide the coordinates of the starting point \( P \), the search line is determined along the gradient direction of the potential function \( Z \) as evaluated at the input point \( P \). Then a subroutine treats the potential \( Z \) as a function of position along this line and minimises \( Z \) with a conventional one-dimensional routine. New values for the solution are computed and consequently \( P \) is re-
placed by a new point, the search direction is redefined and a further minimum of \( Z \) is searched along the new line conjugated to the previous one. Lower initial values for the Lagrange multipliers yield higher-valued pixels of the predicted image. The iterations are performed until some convergence criterion is satisfied, \textit{e.g.}, the chi-square function reaches its statistical expected interval.

The program has been designed using Microsoft Visual C++ version 5.0, it fully complies with ANSI C [16] and runs under Windows 95 on a 450 MHz Pentium II. The choice of C++ is mainly based on its successful combination of control structures featured by high-level languages, exhibiting subtlety and elegance, with the power and flexibility typically associated with low-level (assembly) languages.

As a general policy, pointers have been used throughout the program in view to speed up running since incrementing a pointer is faster than indexing an array. Most data have been of floating point type as a compromise between precision on the one hand and memory requirements and speed on the other hand.

Input images have to be in \textit{pbm} (Portable BitMap graphics file format) and the output processed image comes out in \textit{pgm} (Portable GreyMap graphics file format). Maximum running time required by full processing of digitised images of \( 1000 \times 1000 \) pixel size has been of the order of 6 to 10 minutes without derivative information and around 1 hour if derivatives were taken into account, depending on the number of control subroutines included and texture richness.

\section*{VII. CONCLUSIONS}

The main computational desiderata in any image improvement technique are obvious: fast algorithms, small memory requirements and compliance with cheap computers. Whether the computational effort is dominated by the cost of evaluating the function and its derivatives, eventually, a less stringent demand is replacing the above, namely, the evaluation of the function to be performed as few times as possible.

The choice of the starting image approximation has been reported as [17], an opinion which is supported by our data processing. One may start from a grey picture with all pixel values equal (\textit{no a priori} information and maximum non-committal choice), or directly from the observed blurred image. Nevertheless, incorporating \textit{a priori} knowledge is highly recommended along with the increasing number of pixels in the image.

Test objects, so-called “phantoms”, are routinely used in mammography to test the overall performance of imaging systems. The numerical data processing of output images in the Microprocessor Laboratory of the Abdus Salam ICTP produced results in very good accordance with the simulated targets within the breast tissue (Fig. 1). The size of the smallest detail which could be detected, and
therefore the spatial resolution of both the detector and the final image, has been assumed to coincide with the pixel size. Accordingly, the resolution of our images is at least about 0.2 mm × 0.2 mm. Spatial resolution, however, need not be limited by the pixel size.

![Image](image1.png)

Fig. 1. – Simulated phantom of a microcalcium grain embedded in an homogeneous mammalian tissue: a) simulated image of $31 \times 31$ pixel size; b) improved image after 11 iterations of the reconstruction algorithm.

![Image](image2.png)

Fig. 2. – Digitised images of an X-ray mammogram, a) scanned mammographic image of $1000 \times 1000$ pixel size; b) improved image after the reconstruction processing.

Until the release of our detector, preliminary tests have been carried out by scanning routine photographic mammograms and downloading digitised images from the Internet. The finally processed images proved to be of a fairly good
quality in terms of contrast, signal-to-noise ratio and number of patterns (Fig. 2). However, since the PSF has been defined in accordance with the characteristics of our incoming solid-state detector, ever better results are expected.

Acknowledgements. The present research has been carried out in the Microprocessor Laboratory of the International Centre for Theoretical Physics, Trieste, Italy, within the framework of the X-ray ICTP-INFN-UNU Co-operation Project 95A2MC02C. Special thanks to our pertinent referees who thoroughly corrected the manuscripts and kindly suggested further improvements.

REFERENCES