JOURNAL OF COMPUTATIONAL PHYSICS 24, 339–347 (1977)

A Descent Approach to a Class of Inverse Problems

P. S. KRISHNAPRASAD AND RICHARD BARAKAT*

Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts 02138

Received April 19, 1976; revised September 21, 1976

In this paper, we consider the solution of a class of ill-posed inverse problems occurring in physics and engineering that require the solution to be nonnegative. An iterative algorithm is presented in which this constraint is taken into consideration directly. The algorithm is applied to a typical problem.

1. INTRODUCTION

The inverse problems that we study in this paper are given in terms of the operator equation

$$Af = g, \qquad f \in F, \quad g \in G \tag{1}$$

with the operator A mapping from some complete metric space F onto a complete metric space G. We state the inverse problem as: Given $g \in G$ and A, determine $f \in F$ which satisfies Eq. (1).

It is now reasonably well known that inverse problems are generally numerically unstable due to the ill-posed nature of most inversion problems [1]. Partial stabilization is obtained by imposing constraints, either implicit or explicit, thereby forcing the inverse operator A^{-1} to be better behaved.

In many of the inverse problems encountered in the sciences and engineering, the unknown function f must be nonnegative. Typical examples are: determination of the refractive index structure constant from measurements of the log-amplitude covariance [2], determination of probability density function of aerosol size from spectral attenuation measurements [3], object restoration in incoherent illumination [4, 5], determination of the spectral density function from a finite section of the covariance function [6].

The purpose of the present paper is the construction of an algorithm for inverting Eq. (1) subject to the constraint that f be nonnegative even when g is taken to be noisy.

Our approach is via the quasi-solution idea of Ivanov [7]. A quasi-solution $\tilde{f}$ of the inverse problem is defined by the relation

$$\tilde{f} = \arg \{\min_{f \in M} \| Af - g \|^2\} \tag{2}$$

* Also, Bolt Beranek and Newman Inc., Cambridge, Mass. 02138.

ISSN 0021-9991

where M is a constraint set based on prior information about the solution. By this we mean that $\tilde{f}$ is the solution to the minimization problem in parentheses. The problem of finding a quasi-solution is one of nonlinear programming [8]. The iterative algorithm used in this paper is a modification of the steepest-descent technique and takes into account the nonnegativity constraint. The algorithm picks the optimal direction in which to make a search for the best estimate. In the absence of constraints, the search involves solving a scalar minimization. This is the step involving $\tilde{\alpha}$ in Eq. (18). However, because of the nonnegativity constraint involved the search should be restricted to the feasible set. This is achieved in Eq. (21).

The nonlinear programming approach of this paper can be extended, by the use of Lagrange multipliers, to deal with constraints of the equality type. Such constraints arise, for example, in problems involving a probability density function; here the equality constraint is that f must integrate to a fixed constant.

2. Formulation for a Class of Integral Equations

Let us now specify a class of operators A, the spaces F and G, and the constraint set M of a fairly general nature. Most of the inverse problems listed in Section 1 fall in this category. Consider a Fredholm integral equation of the first kind in n variables, $\hat{x} \equiv (x_1, ..., x_n)$

$$\int_D K(\hat{x}, \hat{y}) f(\hat{y})\, d\hat{y} = g(\hat{x}), \tag{3}$$

where D is a finite domain.

In the problems of interest to us, physical considerations impose nonnegativity of the solution, i.e.,

$$f(\hat{x}) \geqslant 0, \qquad \hat{x} \in D \tag{4}$$

as well as boundedness conditions on the solution, data, and the kernel; as given by

$$\int_D [f(\hat{x})]^2\, d\hat{x} < \infty, \qquad \int_D [g(\hat{x})]^2\, d\hat{x} < \infty \tag{5}$$

and

$$\int_D \int_D [K(\hat{x}, \hat{y})]^2\, d\hat{x}\, d\hat{y} < \infty. \tag{6}$$

From these, it follows that the complete metric spaces F and G can be identified as $L_2(D)$ and the constraint set M is defined by Eq. (4).

A quasi-solution to the problem of inverting Eq. (3) subject to the nonnegativity constraint is

$$\tilde{f} = \underset{f \geqslant 0,}{\arg} \left\{ \min \frac{1}{2} \left\| g - \int_D Kf\, dy \right\|^2 \right\}, \qquad f \in L_2(D), \tag{7}$$

and the norm $\|\cdot\|$ is the usual L_2 norm. For numerical computation the minimization problem in Eq. (7) has to be converted into a problem in finite-dimensional (Euclidean) space.

The discretization is accomplished by considering f and g at a finite set of points in D and approximating the integrand in Eq. (3) by a suitable quadrature rule (e.g., trapezoidal rule). Thus, Eq. (3) is replaced by a system of algebraic equations

$$\hat{A}\hat{f} = \hat{g}. \tag{8}$$

The validity of this process depends on these two approximations and it will be assumed that these approximations are adequate. Attention will be devoted to solving this linear algebraic system. Corresponding to Eq. (2), we have for the quasi-solution

$$\begin{aligned} \tilde{f} &= \arg\{\min \tfrac{1}{2}(\hat{g} - \hat{A}\hat{f})^{+}(\hat{g} - \hat{A}\hat{f})\}, \\ \hat{f} &\geqslant 0. \end{aligned} \tag{9}$$

The problem of two-dimensional object restoration from a noisy image, for example, can be cast into the exact format of the problem posed in this section with the identifications: $f(\hat{x})$ = object intensity distribution, $g(\hat{x})$ = image intensity distribution, $K(\hat{x}, \hat{y})$ = point spread function of optical system. The domain D is the field of certainty (where the unknown object is known to be *a priori*).

In the next section we describe an algorithm to solve Eq. (9) via recursive norm minimization.

3. Inversion Algorithm

Let us define the error function ϕ

$$\phi(\hat{f}) \equiv \tfrac{1}{2}(\hat{g} - \hat{A}\hat{f})^{+}(\hat{g} - \hat{A}\hat{f}) \tag{10}$$

and its gradient $\hat{\psi}$ with respect to $\hat{f}$

$$\hat{\psi}(\hat{f}) \equiv \nabla\phi(\hat{f}) \equiv \hat{A}^{+}(\hat{A}\hat{f} - \hat{g}). \tag{11}$$

The minimization problem of Section 2 can be written compactly as

$$\min \phi(\hat{f}), \qquad \hat{f} \geqslant 0. \tag{12}$$

Any $\hat{f}$ which is nonnegative is said to be a feasible solution and the set

$$\omega \equiv \{x\colon x \in R^N,\ x \geqslant 0\} \tag{13}$$

is called the feasible set. The nonlinear programming algorithm starts at a point in the feasible set and proceeds iteratively to an optimal solution, for Eq. (9), within the feasible set.

Consider the recursion relation

$$\hat{f}^{(k+1)} = \hat{f}^{(k)} + \alpha^{(k)}\hat{d}^{(k)} \tag{14}$$

where

$$\begin{aligned} \hat{d}_i^{(k)} &= -\hat{\psi}_i^{(k)}, \qquad \text{if } \hat{f}_i^{(k)} > \hat{0} \text{ or } \hat{\psi}_i^{(k)} < \hat{0}, \\ &= \hat{0}, \qquad \text{if } \hat{f}_i^{(k)} = \hat{0} \text{ and } \hat{\psi}_i^{(k)} \geqslant \hat{0}, \end{aligned} \tag{15}$$

for $i = 1,..., N$. It can be shown that [8]

a. A necessary condition for a relative minimum at $\hat{f}^{(k)}$ is

$$\langle \hat{\psi}^{(k)}, \hat{d}^{(k)} \rangle \geqslant 0 \tag{16}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual scalar product;

b. $\hat{d}^{(k)} = \hat{0}$ only if $\hat{f}^{(k)}$ satisfies condition (a);

c. there exists an $\bar{\alpha} > 0$ such that for $0 < \alpha^{(k)} \leqslant \bar{\alpha}$

$$\phi[\hat{f}^{(k)} + \alpha^{(k)}\hat{d}^{(k)}] \leqslant \phi[\hat{f}^{(k)}]. \tag{17}$$

The equality being satisfied only when condition (a) is satisfied.

A recursion similar to Eq. (14) was considered by Ho and Kashyap [9, 10]. $\hat{d}^{(k)}$ as given above in Eq. (15) is a feasible search direction. From condition (c) above it is clear that we can make a reduction in the error-function ϕ by proceeding in the $\hat{d}^{(k)}$ direction by a suitable step $\alpha^{(k)}$, as long as we are not at a relative minimum. Thus, $\hat{d}^{(k)}$ is a descent direction. The problem now is to determine $\alpha^{(k)}$ which gives the optimal reduction without violating feasibility.

4. Computation of Step-Size Parameter $\alpha^{(k)}$

The optimal step-size $\bar{\alpha}$ without any restriction such as $\hat{f}^{(k+1)} \in \omega$, will be obtained by solving the minimization problem (involving one-dimensional search),

$$\min_{\alpha>0} \phi[\hat{f}^{(k)} + \alpha\hat{d}^{(k)}]. \tag{18}$$

Let $\bar{\alpha}$ be the solution to Eq. (18). This $\bar{\alpha}$, however, does not guarantee feasibility, that is, it is quite possible that

$$\hat{f}^{(k)} + \bar{\alpha}\hat{d}^{(k)} < \hat{0}. \tag{19}$$

It is easy to see that a step-size α should not be greater than

$$\begin{matrix} \min[-(\hat{f}_i^{(k)}/\hat{d}_i^{(k)})], & \\ & i = 1, 2, \ldots, N. \\ \hat{d}_i^{(k)} < \hat{0}. & \end{matrix} \tag{20}$$

From Eqs. (19) and (20) the optimal step-size is given by

$$\begin{matrix} \alpha^{(k)} = \min\{\bar{\alpha}, [-(\hat{f}_i^{(k)}/\hat{d}_i^{(k)})]\}, & \\ & i = 1, 2, \ldots, N. \\ \hat{d}_i^{(k)} < \hat{0}. & \end{matrix} \tag{21}$$

Computation of $\bar{\alpha}$ involves finding the positive root of the equation

$$\lambda(\alpha) = \langle \hat{\psi}^{(k+1)}, \hat{d}^{(k)} \rangle = 0 \tag{22}$$

where

$$\hat{\psi}^{(k+1)} = \nabla\phi[\hat{f}^{(k)} + \alpha\hat{d}^{(k)}] \tag{23}$$

The determination of the roots of Eq. (22) is accomplished via the method of false position.

Astopping rule for the recursion, Eq. (14), is based on condition (a) of Section 3 and the Aitken δ^2-procedure which is given by the following rule. Stop if

$$|\,\phi[\hat{f}^{(k)}] - \phi^*[\hat{f}^{(k)}]| < \epsilon, \tag{24}$$

where ϵ is a suitable tolerance, and

$$\phi^*[\hat{f}^{(k)}] = \frac{[\phi^{(k)}]^2 - \phi^{(k-1)} \cdot \phi^{(k+1)}}{2\phi^{(k)} - \phi^{(k-1)} - \phi^{(k+1)}}, \tag{25}$$

$$\phi^{(k)} \equiv \phi[\hat{f}^{(k)}]. \tag{26}$$

The algorithm of this paper does not incorporate any smoothing, and the results obtained still compare favorably with other methods (see next section). In case there is a ridge-type minimum or a broad valley for ϕ in Eq. (10), smoothing would help avoid oscillations in the iterative process and give better convergence. This involves replacing the minimization problem of Eq. (12) by a modified minimization problem

$$\min_{\hat{f} \geqslant 0} \phi_\beta(\hat{f}) \tag{12'}$$

where

$$\phi_\beta(\hat{f}) = \tfrac{1}{2}(\hat{g} - \hat{A}\hat{f})^+(\hat{g} - \hat{A}\hat{f}) + \beta\hat{f}^+\hat{f} \tag{10'}$$

and $\beta > 0$ is a small smoothing parameter. The new gradient is now

$$\hat{\psi}_\beta(\hat{f}) = \nabla\phi_\beta(\hat{f}) = \hat{A}^+(\hat{A}\hat{f} - \hat{g}) + 2\beta\hat{f}. \tag{11'}$$

The algorithm is the same as before. One obvious way of determining β is to run some trials with known functions of f. There are some clues as to more systematic ways of determining the optimal smoothing parameter β and even the best discretization. These are the subject of future publications.

4. Numerical Example

In the interest of illustrating the main ideas unencumbered by geometrical complications, we consider a one-dimensional problem of object reconstruction for a slit aperture operating in incoherent illumination.

The object is a one-dimensional unit pulse of half-width x_0 imaged by an aberration-free aperture

$$\begin{aligned} f(x) &= 0, \qquad -\infty < x < -x_0 + \delta, \\ &= 1, \qquad -x_0 + \delta < x < x_0 + \delta, \\ &= 0, \qquad x_0 + \delta < x < \infty, \end{aligned} \tag{27}$$

where δ is the shift from center. Note that the object is always taken to lie within the interval of certainty, i.e.,

$$[-x_0 + \delta, x_0 + \delta]\, D \equiv [-\Omega, \Omega] \tag{28}$$

where Ω is given.

The kernel function is given by [11]

$$K(x, y) = K(x - y) = \frac{1}{\pi} \left[\frac{\sin(x - y)}{(x - y)}\right]^2. \tag{29}$$

The image $g(x)$ is given by

$$g(x) = \frac{1}{\pi} \int_{-x_0+\delta}^{x_0+\delta} \left[\frac{\sin(x - y)}{(x - y)}\right]^2 dy. \tag{30}$$

The integral can be expressed in terms of the sine integral, but we found it more convenient to evaluate the integral numerically.

The integral equation was discretized via a trapezoidal rule. We chose to make the number of sampled points in the image equal to the number of reconstruction points in the object for these illustrative calculations.

Noise was introduced into the image in a multiplicative fashion. We set

$$\hat{g}_{\text{noisy}} = (1 + k\mu)\, \hat{g}_{\text{noiseless}} \tag{31}$$

where μ is a random variable uniformly distributed over $(-\frac{1}{2}, \frac{1}{2})$ and k is a positive constant less than unity. Values of k used in the present calculations are $k = 0.10$ and 0.20, loosely described as 10 and 20% noise.

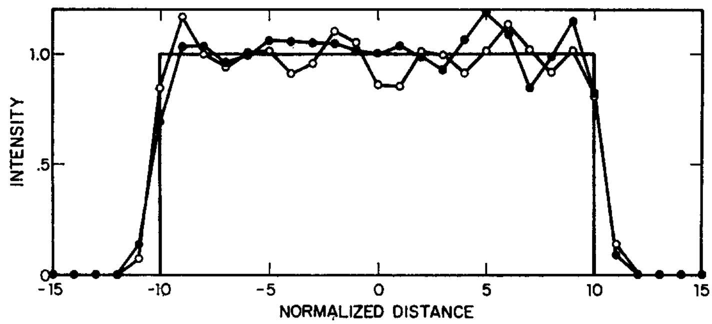

FIG. 1. Object reconstruction for image subject to 10 % noise. — original object, • reconstructed values for first sample realization, ∘ reconstructed values for second sample realization.

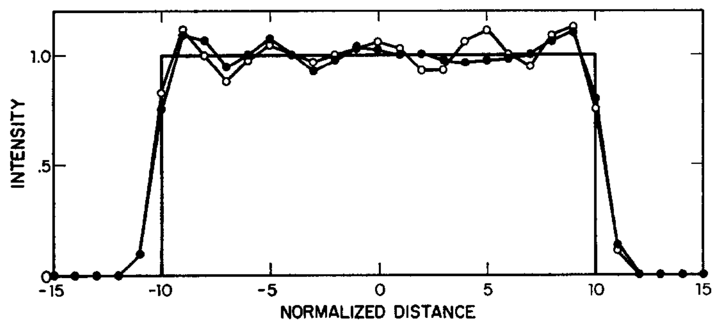

FIG. 2. Object reconstruction for image subject to 20 % noise. — original object, • reconstructed values for first sample realization, ∘ reconstructed values for second sample realization.

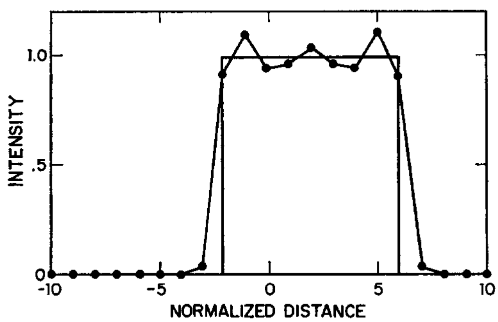

FIG. 3. Object reconstruction for image subject to 10 % noise. — original object, • reconstructed values for a sample realization.

As a first example, an object of half width $x_0 = 10$ centered at $x = 0$ (i.e., $\delta = 0$) was considered and the fundamental interval taken to be $\Omega = 15$. The image was evaluated via Eq. (30) and then corrupted by 10 and 20% noise, respectively. Numerical results are summarized in Figs. 1 and 2 for two sample realizations of the image. The reconstructed object comprises a set of 31 data points which are connected by strighet line segments.

In Fig. 3, we show the reconstruction of an asymmetrically placed object ($x_0 = 4$, $\delta = 2$, $\Omega = 10$) for an image corrupted by 10% noise.

The constraints expressed by Eqs. (4)–(6) imply that the metric spaces F and G be identified as $L_2(D) \equiv L_2(\Omega)$. This means that we cannot expect the reconstructed object to match the original object on a pointwise basis, since we are only requiring that the normed difference be a minimum. Thus, even when the image is essentially noiseless, in the sense that it is known to four–six digits, we cannot expect perfect reconstruction on a pointwise basis. The fact that the object is discontinuous is a further complication and makes the bar target object a severe test of the inversion algorithm. Figure 4 shows the reconstruction of the same object as in Figs. 1 and 2

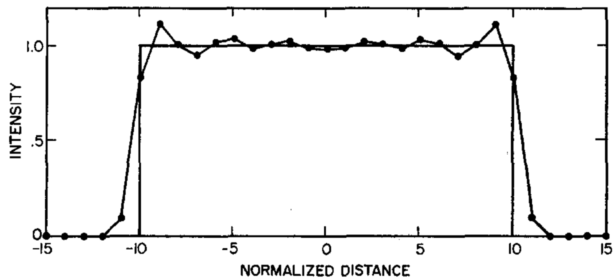

FIG. 4. Object reconstruction for "noiseless" image. — original object, • reconstructed values.

for a "noiseless" image. These results should be compared with those of Barakat and Blackman [12], who employed Tichonov regularization, for the same problem. The present results are superior since they are guaranteed to be nonnegative whereas Tichonov regularization does not possess this feature.

A brief note on the actual iterative process is in order. The iterations proceed by reducing the error function ϕ defined in Eq. (10). The starting point used for all iterations is the vector $f = 0$. A value of $\phi = 0.002$ is used as a threshold at which the recursion is terminated. Other thresholds based on the norm of the descent direction vector and the Aitken δ^2-procedure have also been incorporated.

ACKNOWLEDGMENTS

Krishnaprasad was supported by the Division of Engineering and Applied Physics, Harvard University and ONR under Joint Services Electronics Program, Contract N00014-67A-0298-006. Barakat was supported by Air Force Office of Scientific Research (AFSC) under Contract F44620-72-C-0063.

REFERENCES

1. M. Lavrentiev, "On Some Improperly Posed Problems of Mathematical Physics," Springer-Verlag, Berlin, 1963.
2. J. Heneghan and A. Ishimaru, Remote determination of the profiles of the atmospheric structure constant and wind velocity along a line-of-sight path by a statistical inversion procedure, *IEE Trans. Antennas Propagat.* **AP-22** (1974), 457.
3. A. Fymat, Remote monitoring of environmental particulate pollution: a problem in inversion of first-kind integral equations, *App. Math. Comp.* **1** (1975), 131.
4. B. Hunt, Digital image processing, *Proc. IEEE* **63** (1975), 693.
5. B. Frieden, Image enhancement and restoration, *in* "Topics in Applied Physics," Vol. 6, Springer-Verlag, Berlin, 1975.
6. T. Ulrych and T. Bishop, Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys. Space Phys.* **13** (1975), 183.
7. V. Ivanov, On improperly posed problems, *Mat. Sb.* **61** (1963), 2.
8. D. Luenberger, "Linear and Nonlinear Programming," Wiley, New York, 1972.
9. Y. Ho and R. Kashyap, An algorithm for linear inequalities and its applications, *IEEE Trans. Electronic Computers* **EC-14** (1965), 683.
10. Y. Ho and Kashayap, A class of iterative procedures for linear inequalities, *SIAM J. Control* **4** (1966), 112.
11. J. Goodman, "Fourier Optics," Wiley, New York, 1968.
12. R. Barakat and E. Blackman, Application of the Tichonov regularization algorithm to object restoration, *Opt. Comm.* **9** (1973), 252.