A GENERAL FORMULATION OF CONSTRAINED ITERATIVE RESTORATION ALGORITHMS*

A. K. Katsaggelos, J. Biemond**, R. M. Mersereau, R. W. Schafer

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

**Department of Electrical Engineering
Delft University of Technology
Delft, The Netherlands

ABSTRACT

This paper introduces a general formulation of constrained iterative restoration algorithms in which deterministic and/or statistical information about the undistorted signal and statistical information about the noise are directly incorporated into the iterative procedure. This a priori information is incorporated into the restoration algorithm by what we call "soft" or statistical constraints. Their effect on the solution depends on the amount of noise on the data; that is, the constraint operator is "turned off" for noiseless data. The development of the new iterative algorithm is based on results from regularization techniques for stabilizing ill-posed problems.

INTRODUCTION

The recovery or restoration of a signal that has been distorted is one of the most important problems in signal processing. An appropriate mathematical representation of the distorting process is

$$K x = y \tag{1}$$

where x is the unknown input signal, y is the known output signal, and **K** is a known linear distortion operator or transformation. The restoration problem is then the problem of solving the operator equation (1) for x given y and **K**. The signals x and y may be one or two dimensional (images) and the linear operator **K** may be shift-invariant or shift-varying. When x and y represent images **K** may represent the distortion introduced due to motion of the camera with respect to original image plane, defocussing, atmospheric turbulence, etc.

In infinite dimensional spaces the operator **K** is an integral operator and equation (1) is called a Fredholm integral equation of the first kind. Then the solution of equation (1) is almost always an ill-posed problem [1,2]. A problem is called ill-posed when the solution or the least-squares solution does not depend continuously on the data; or in other words, when the set of solutions of equation (1) is unbounded for a small perturbation in the data [1,3]. When the continuous problem is discretized (sampled) then **K** in equation (1) represents an n×n matrix and x and y represent n×1 vectors. The matrix K is now ill-conditioned [4]. The solution or the least-squares solution of the set of linear equations (1) is unstable, which means that a small perturbation in the data y results in a large perturbation in the solution x [2]. Phillips [5] demonstrated the instability of the restoration problem and the effect of the sampling interval on the solution.

Obviously, random noise (error) in the data is unavoidable. It may originate with the transmission medium, the recording process, the image formation process or any combination of these. Quantization and computational errors also represent considerable limitations in the solution of ill-posed problems. Therefore, methods are required that will provide stable solutions to the signal restoration problem. These methods are called regularization methods and they are different approaches that circumvent lack of continuous dependence or stability [1,3,6]. Roughly speaking, a regularization method entails an analysis of an ill-posed problem via an analysis of an associated well-posed problem, provided this analysis yields meaningful answers and approximations to the given ill-posed problem. From the large number of regularization methods [2,3], a method proposed by Tikhonov [1] and Miller [6] is used in this work.

A REGULARIZATION METHOD

The general approach for solving ill-posed problems is to incorporate knowledge about the solution and the error (noise) into the solution process. Prior knowledge is the remedy of ill-posedness. Therefore, we assume that the norm of the noise ϵ is known, that is

$$||y-y_T|| \le \epsilon \tag{2}$$

where y_T and y are the true and noisy data in equation (1), respectively. Denote by x′ the solution (or minimum norm least-squares solution) of equation (1) when the noisy data y are available. Then, since the problem is ill-posed, the set F of functions x′ that satisfy

$$||Kx'-y_T|| \le \epsilon \tag{3}$$

is unbounded for the infinite dimensional case, or very large along some directions (as will be shown in a later section) for the finite dimensional case

*This work was supported in part by the Joint Services Electronics Program under contract DAAG29-84-K-0024 and in part by the Netherlands Organization for the Advancement of Pure Science (ZWO).

[2]. The main underlying idea for most of the regularization methods is to restrict the set F of possible solutions.

In order to do this, Miller [6] formulates the original problem of solving equation (1), or minimizing the functional $||Kx-y||$, as follows: search for functions (vectors) x that satisfy both the constraints

$$||Kx-y||\leq\epsilon \tag{4a}$$

and

$$||Cx||\leq E \tag{4b}$$

where ϵ as before, is an estimate on the data accuracy, E is a prescribed constant and C is a linear operator. C is called the constraint operator and it incorporates the a priori information about the solution directly into the algorithm.

Many different choices are possible for C, depending on the available prior knowledge. The simplest one is C=I; then (4b) is a constraint on the norm of x. Another common choice is to let C be a differential operator and then (4b) expresses a smoothness requirement imposed on the solution. However, for the general formulation of the theory, it is not necessary to specify C. It is only required that C have a continuous inverse in the infinite dimensional case [6], or be a nonsingular matrix in the finite dimensional case. The specific desirable properties of C will be analyzed in a later section in this paper.

Miller shows [6] that the problem described by the constraints (4b), (4c) is equivalent to minimizing the functional

$$M(\alpha,x)=||Kx-y||^2+\alpha||Cx||^2 \tag{5}$$

where α is the regularization parameter which is set to be equal to $(\epsilon/E)^2$. When E is not known α is adjusted iteratively by the discrepancy, so that $||Kx_\alpha-y||=\epsilon$, where x_α is the function that minimizes $M(\alpha,x)$ for a specific value of α [1]. The function x that minimizes the functional $M(\alpha,x)$ is given by the expression [6]

$$(K^*K+\alpha C^*C)x=K^*y \tag{6}$$

where * denotes the adjoint operator (for real matrices it denotes the transpose of a matrix). The operator $(K^*K+\alpha C^*C)$ has a continuous inverse [6] (infinite dimensional case) or it is a better conditioned matrix than K or K^*K, assuming that the matrix C has been chosen properly.

Obviously, the set of x that satisfy both the constraints (4a) and (4b) must not be empty if a useful approximation is to be obtained. A couple $\{\epsilon,E\}$ that results in a nonempty set of solutions is called permissible. Miller [6] describes a method for constructing the set of permissible $\{\epsilon,E\}$. We suggest the use of a multiple of the signal norm for the bound E, i.e., $E=\gamma||x||$. The constant γ is a function of the norm of C. Then the regularization parameter α is the reciprocal of a multiple of the signal-to-noise ratio, which is assumed to be known in advance. Such a choice for E may lead to a smaller α (the effect of the constraint is reduced), and E may need to be corrected. The whole approach can be considered as an interactive approach; that is, some of the problem parameters can be adjusted iteratively so that some error criterion is satisfied [1,10]. When the signals x and y represent images, the visual quality of the restored image may be considered as the criterion to be satisfied.

GENERAL FORM OF CONSTRAINED ITERATIVE RESTORATION ALGORITHMS

The regularized solution given by equation (6) can be successively approximated by means of the iteration

$$x_0 = 0$$

$$x_{k+1} = x_k+\beta[K^*y-(K^*K+\alpha C^*C)x_k] \tag{7a}$$

$$= (I-\beta\alpha C^*C)x_k+\beta K^*(y-Kx_k) \tag{7b}$$

$$= C_N x_k+\beta K^*(y-Kx_k) \tag{7c}$$

where $C_N=(I-\beta\alpha C^*C)$ can be considered as a new constraint operator, and β is a multiplier that can be constant or it can depend on the iteration index [7]. When K^*y is in the range of the operator $K^*K+\alpha C^*C$ algorithm (7) converges to the minimum norm solution for $0<\beta<2||K^*K+\alpha C^*C||^{-1}$, according to Bialy's theorem [8]. Algorithm (7c) has a very similar form to the algorithm presented in the paper by Schafer et al. [9]. One difference between the two algorithms is that the constraint C_N is applied to x_k inside the parenthesis in equation (7c), in the algorithm by Schafer et al. In some cases it can be argued that this will not alter the behaviour of the algorithm; for instance if C_N and K represent low-pass filtering operations then KC_Nx_k might be very close to Kx_k, depending on the specific properties of C_N and K. The main difference though lies in the interpretation of the constraint operators. The constraints that are used in the paper by Schafer et al. [9], have the property of a projection operator; that is, if the constraint is applied to the signal more than once the signal remains unchanged. We interpreted these constraints as hard constraints [7]. Hard constraints can still be used in algorithm (7) above. The constraints C or C_N that are used in this paper do not have the property of a projection operator and their "power" depends on the amount of noise in the data. This is easily seen if we include the regularization parameter α in the constraint operator, since when the noise goes to zero, α goes to zero and $C_N=I$. Conditions for the regularized solution to approach the true solution when the noise goes to zero, as well as error estimates are given in [2,6]. We interpreted these constraints as "soft" statistical constraints [7]. Their main role is to stabilize an ill-posed problem. The predictor constraint that we proposed in [7] agrees with the formulation of the problem that is presented in this paper. A new form of the constraint C_N will be motivated and proposed in the next section.

PROPERTIES OF THE CONSTRAINT OPERATOR

In order to analyze the properties of the constraint operator and also get a better understanding of the role of the regularization we use the singular value decomposition of the distortion operator K. The operator **K** is represented by an n×n matrix K for this analysis. The same analysis holds for a compact operator **K** in an infinite dimensional space [3]. For the matrix K the following similarity transformation holds [4,10]

$$K = U\,\Lambda^{1/2}\,V' \tag{8a}$$

where

$$K\,K' = U\,\Lambda\,U' \tag{8b}$$

$$K'\,K = V\,\Lambda\,V' \tag{8c}$$

and

$$U'\,U = V\,V' = I \tag{8d}$$

The matrices U and V are n×n orthogonal matrices and they are the respective eigenvector matrices of KK′ and K′K, where ′ denotes the transpose of a matrix. Since KK′ and K′K are symmetric, positive semidefinite matrices, each has a complete set of eigenvectors and the same set of real nonnegative eigenvalues λ_i, $1 \le i \le n$. We assume that the λ_i's are ordered in decreasing order and they are the entries of the diagonal matrix Λ. The quantities $\mu_i = +\lambda_i^{1/2}$ are the entries of the matrix $\Lambda^{1/2}$ and they are called the singular values of K. Then if we denote by u_i, v_i, $1 \le i \le n$ the ordered column eigenvectors of KK′ and K′K, respectively, and if R is the rank of the matrix K

$$K = \sum_{i=1}^{R} \mu_i u_i v_i' \tag{9a}$$

and

$$K^{+} = \sum_{i=1}^{R} \mu_i^{-1} v_i u_i' \tag{9b}$$

where K^{+} is the generalized inverse of K.

According to equation (9b) the minimum norm least-squares solution of equation (1), when the noisy data $y = y_T + \delta y$ are available is

$$x = K^{+}y = \sum_{i=1}^{R} \mu_i^{-1}(u_i, y_T)v_i + \sum_{i=1}^{R} \mu_i^{-1}(u_i, \delta y)v_i \tag{10}$$

where (.,.) is the inner product in R^n. It is easy to see from equation (10) the difficulty in solving a system of linear equations when K is an ill-conditioned matrix. The ill-conditionedness of the matrix K means that some of the singular values μ_i are very small; even if we exclude the zero singular values (as we do in computing K^{+}), most probably there will still be singular values close to zero, since the finite dimensional problem resulted from the discretization of an ill-posed infinite dimensional problem. Therefore, some of the weights μ_i^{-1} are very large numbers. On the other hand, since y_T is in the range of K, $(u_i, y_T) = \mu_i (v_i, x_T)$ goes to zero when μ_i goes to zero, but $(u_i, \delta y)$ does not necessarily go to zero when μ_i goes to zero (this is always true if we assume that the noise is white). Therefore, the weight $\mu_i^{-1}(u_i, \delta y)$ can become a very large number for those μ_i that are close to zero, even though δy is a very small number.

From the above discussion it is easy to characterize or design the constraint C. C should increase the μ_i's for which (v_i, x_T) is small and leave the remaining μ_i's unchanged. In order for this to happen, C′C should be diagonalized by the same set of eigenvectors v_i, $1 \le i \le n$, that is

$$C'\,C = V\,\Sigma\,V' \tag{11}$$

where Σ is a diagonal matrix with entries σ_i, $1 \le i \le n$. Then the regularized solution obtained from equation (7b) after k iterations can be written as

$$x_k = \sum_{i=1}^{R} \mu_i(\mu_i^2 + \alpha\sigma_i)^{-1}(1 - (1 - \beta(\mu_i^2 + \alpha\sigma_i))^k)(u_i, y_T + \delta y)v_i \tag{12a}$$

and for $k \to \infty$

$$x = \sum_{i=1}^{R} \mu_i(\mu_i^2 + \alpha\sigma_i)^{-1}(u_i, y_T + \delta y)v_i \tag{12b}$$

provided that $0 < \beta < 2\|K'K + \alpha C'C\|^{-1}$. Clearly, for the singular values that are close to zero the weight $\mu_i(\mu_i^2 + \alpha\sigma_i)^{-1}$ goes to zero and the error is not amplified. Equations (11) and (12b) can be used in choosing the constraint operator C. Obviously, these equations are convienient for theoretical arguments, but are not useful as practical solution. This is so because the problem of determining the eigenvalues and eigenvectors of a large matrix is a computationaly difficult problem.

The problem of choosing the operator C is straightforward when both the distorting and the constraint operators are linear shift-invariant. Then the matrices K and C are Toeplitz and they are asymptotically equivalent in the sense of the Euclidean norm to circulant matrices [10]. The two basic properties of circulant matrices are [10]: a) the complete collection of eigenvectors of a circulant is obtained from the unitary transform matrix that represents the discrete Fourier transform (DFT); and b) the eigenvalues of a circulant are the DFT values of the cyclic sequence that makes up the circulant. Therefore, the values μ_i and σ_i in equations (12a-b) are the DFT values of the impulse response of the distorting system and the system that imposes the constraint, respectively. Our intuition of the property of the constraint operator is confirmed; when for example the distortion represents a low-pass filter the constraint C is desired to represent a high-pass filter, so that the inverse of K′K+αC′C does not blow up at high frequencies, and the energy of the restored signal at high frequencies (mainly due to the noise) is suppresed. In order to be successful in supressing the noise without severely distorting the original signal, information about the smoothness of the original signal or equivalently information about the highest frequency in the signal is necessary. This information can be incorporated into the algorithm by the signal and noise covariance matrices, assuming that the undistorted signal represents a sample from a stochastic process. The introduction of signal statistics is related to a class of regularization methods that extent an ill-posed problem to a stochastic well-posed problem [2,3].

Therefore, for the case that K and C represent linear shift-invarint operators, we propose as a constraint operator C_N in equation (7c) the use of a Wiener noise smoothing filter. This constraint incorporates knowledge about the "degree of smoothness" of the original signal. The constraint can be also considered as part of the regularization parameter. This can be understood if we observe that when the noise goes to zero $C_N=I$, that is $\alpha C'C=0$. Therefore if we define a new regularization parameter α' by $\alpha'=\alpha C'C$, then α' goes to zero when the noise goes to zero without the need to determine α by the discrepancy. The other choice for the constraint operator C that leads to a similar result is to set $C'C=\Phi_x^{-1}\Phi_n$, where Φ_x and Φ_n are the signal and noise covariance matrices, respectively. This case has been studied by Hunt [10] and results in the parametric Wiener filter. When the matrices Φ_x and Φ_n are also circulant (the signal and the noise are stationary), the Fourier version of the restoration filter becomes applicable. In this case the frequency response of the iterative restoration filter after an infinite number of iterations can be used in choosing the appropriate constraint operator and parameter α.

EXPERIMENTAL RESULTS

In our preliminary experiments equation (7c) was used to restore noisy blurred images. A Wiener noise smoothing filter was used for the constraint operator C_N. In figures (1a) and (1b) the noisy (SNR=20db), blurred (motion blur of nine samples) image, and the restored image after 50 iterations with $\beta=1$ are shown, respectively. The one-dimensional filter with impulse response $c(0)=.9$, $c(-1)=c(1)=.05$ was used as a constraint, and the image was processed line by line. The improvement in SNR in the restored image is 7.6 db. Comparable results were obtained by using a two-dimensional Wiener filter as a constraint. Better results are expected if a different Wiener filter is used for each image line.

DISCUSSION AND CONCLUSIONS

The iterative signal restoration algorithm that was developed in this paper generalizes the form of the iterative algorithms that were described in the paper by Schafer et. al. [9]. The available information about the signal (deterministic and/or statistical) and the noise (statistical) is directly incorporated into the iterative process. Two types of constraints can be combined in the algorithm; "soft" statistical constraints as well as "hard" constraints.

The development of the algorithm is based on a specific regularization method. In developing the algorithm, the constraint operator C as well as the regularization parameter α (or equivalently the bounds ϵ and E) need to be specified. Different constraint operators may be suitable for a specific problem. In choosing or designing a constraint operator the knowledge of the way its properties effect the solution is essential. The regularized solution is derived iteratively because of the advantages iterative methods offer over noniterative methods. Among these advantages we mention that it is easy to incorporate constraints and distortions, the solution process can be monitored as it progresses and it is easy to develop adaptive (spatially varying) algorithms. The development of spatially varying constrained iterative algorithms is the subject of another paper.

REFERENCES

[1] A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-Posed Problems*, Wiston, Wiley, 1977.

[2] M. Bertero, C. DeMol, and G.A. Viano, "The stability of inverse problems," in *Inverse Scattering Problems in Optics*, vol. 20 of Topics in Current Physics, H.P. Baltes, ed., pp. 161-214, Springer-Verlag, Berlin, 1980.

[3] M.Z. Nashed, "Aspects of generalized inverses in analysis and regularization," in *Generalized Inverses and Applications*, edited by M.Z. Nashed, Academic Press, New York, 1976.

[4] G. Strang, *Linear Algebra and Its Applications*, Academic Press, New York, 1976.

[5] D.L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *Assoc. Comput. Mach.*, vol.9, pp. 84-97, 1962.

[6] K. Miller, "Least-squares method for ill-posed problems with a prescribed bound," *SIAM J. Math. Anal.*, vol. 1, pp. 52-57, Feb. 1970.

[7] A.K. Katsaggelos, J. Biemond, R.M. Mersereau, and R.W. Schafer, "An iterative method for restoring noisy blurred images," *Circuits, Systems and Signal Processing*, vol. 3, no. 2, pp. 139-160, June 1984.

[8] J.L.C. Sanz and T.S. Huang, "Unified Hilbert space approach to iterative least-squares linear signal restoration," *J. Opt. Soc. Am.*, vol. 73, no. 11, pp. 1455-1465, Nov. 1983.

[9] R.W. Schafer, R.M. Mersereau, and M.A. Richards, "Constrained iterative restoration algorithms," *Proc. IEEE*, vol. 69, pp. 432-450, April 1981.

[10] H.C. Andrews and B.R. Hunt, *Digital Image Restoration*, Prentice Hall, 1977.

(a) (b)

Fig. 1: (a) noisy-blurred image: motion blur of nine samples, SNR=20dB; (b) restored image: improvement in SNR=7.6dB.