# A Unified Treatment of Remote Sensing Problems[+*]

R. P. Tewarson

CEAS #263

## 1. Introduction.

Remote sensing problems occur in many important practical applications. Some of these are: the determination of vertical temperature profiles by satellites for weather forecasting, inference from reflected energy (waves) to predict the rock structure in oil and coal exploration, reconstruction of pictures from their projections in x-ray image processing, and some inverse problems of compartmental analysis in Biology (e.g., parameter estimation in models of renal concentrating mechanisms).

In many cases, a common feature of the remote sensing problems is their underdetermined nature. The amount of data that can be observed is small due to the cost, instrumentation or experimental constraints. For example, a biological system can be subjected to only a small number of probes (experiments) before it degenerates and the number of parameters to be estimated may be much larger than the possible number of probes. In the caseofx-rays only a small number of projections can be taken due to the potential radiation damages to the patient. The satellites can have only a finite number of channels for observing the radiation, due to their weight and cost considerations.

Mathematically, the underdetermined nature of our problems can be described as follows. bt us assume that the relevant discretization and the numerical quadrature have already been applied, then we have a system of equations

---

[+] Presented at the "Image Processing for 2-D and 3-D Reconstruction from Projections-Theory and Practice in Medicine and the Physical Sciences" meeting, August 4-7, 1975, Stanford, California. Sponsored by Stanford university Institute for Electronics in Medicine and the Optical Society of America. A summary of this work was published in the Technical Digest of the meeting.

[*] This research was supported in part by the National Institute of Arthritis, Metabolic and Digestive Diseases, National Institutes of Health, Bethesda, Md. Grant No. 1 RM01 AM 17593-01-02.

$$F(y) = d, \qquad (1.1)$$

where d is an unknown m dimensional vector of observations, y is the unknown n element vector to be determined, and F is a column vector of dimension m. The number of observations m is much smaller than the number of unknowns n. In the applications mentioned earlier, d is a vector of, respectively, the radiance measurements made by the satellite, projection data suitably ordered, above ground sensor data, or input-output and probe data from a mammalian kidney.

Our problem is to determine y for a given value of d. Since the problem is underdetermined ($m < n$), viz., many y's lead to the same d, additional information has to be incorporated in the solution process to get a desirable unique solution y. The observed vector d is generally contaminated with noise (measurement errors etc.) and therefore an exact solution of (1.1) may not be as useful as an approximate solution which satisfies some constraints on y and d. These constraints on y and d can be obtained from the available additional information. This information can be of one or more of the following kinds: statistical, smoothing and structural. Furthermore, we may have some information to come up with a good initial guess for the solution, and/or know some basis functions in terms of which the solution can be expressed.

In the next section a method for solving the system of equation (1.1) is described. The subsequent section deals with the problem of incorporation of known information about the solution vector or the right handside in the solution process.

2. Solving the System of Equations.

The solution of (1.1) involves the determination of a vector $y^*$ such that $F(y^*) \equiv d$. Let $y^o$ be an initial approximation to $y^*$, then $F(y^o) \neq d$, otherwise $y^o = y^*$ and the solution process is complete. Let $x^o$ be the unknown correction such that $y^* = y^o + x^o$, then we have

$$F(y^o + x^o) - d = 0$$

and the use of Taylor's theorem leads to

$$F(y^o) - d + F'(y^o)\, x^o + \cdots = 0, \tag{2.1}$$

where $F'(y^o)$ is the derivative (Jacobian) of F with respect to y evaluated at $y^o$. If we let $b^o = d - F(y^o)$ and $A^o = F'(y^o)$ then (2.1) can be written as

$$A^o\, x^o = b^o + \cdots . \tag{2.2}$$

The exact solution of the above equation for $x^o$ may yield a value which may have a large norm and this is certainly undesirable because Taylor's expansion in (2.1) requires that the norm of $x^o$ be small. To alleviate this situation we proceed as follows. Let us consider the functional

$$\varphi(x^o) = \|b^o - A\, x^o\|_V^2 + \|x^o\|_W^2 ,$$

where V and W are symmetric positive semi-definite matrices and $\|z\|_G$ is the G-length of a vector z such that $\|z\|_G^2 = z^T G z$. Thus $\varphi(x^o)$ is the sum of the squares of the V-length of the residual and the W-length of the correction. Note that for G = I, the G-length is the usual Euclidean length of the vector. Minimizing $\varphi(x^o)$ will minimize the V-length of the residual in (2.2) as well as the W-length of the correction vector $x^o$. It should be observed that minimizing $\|x\|_W^2$ is a very desirable feature because the higher order terms in (2.1) can be neglected if $\|x^o\|_W^2$ is small. We have made the tacit assumption that the second derivatives (Hessians) of F in the neighborhood of $y^*$ are bounded. The matrices V and W are generally chosen to be positive definite and their choice is based on statistical and/or smoothing conditions. This is described in the next section.

The value of $x^o$ which minimizes $\varphi(x^o)$ can be found as follows (Golub and Businger, 1965). We first use the Cholesky factorization (this requires that V and W be positive definite or semi-definite (see Lawson and Hanson, 1974; page 124) to compute R and S such that

$$R^T R = V \text{ and } S^T S = W \tag{2.3}$$

and then apply an orthogonal triangularization to the linear system

$$\begin{bmatrix} RA^o \\ S \end{bmatrix} x^o = \begin{bmatrix} Rb^o \\ 0 \end{bmatrix} \tag{2.4}$$

such that

$$Q \begin{bmatrix} RA^o \\ S \end{bmatrix} x^o = Q \begin{bmatrix} Rb^o \\ 0 \end{bmatrix} \tag{2.5}$$

or

$$\begin{bmatrix} U \\ 0 \end{bmatrix} x^o = \begin{bmatrix} d \\ f \end{bmatrix} \tag{2.6}$$

where U is an upper triangular matrix and d and f are column vectors of dimension n and m respectively. It can be easily shown that the value of $x^o$ which minimizes $\varphi(x^o)$ is the least squares solution of (2.4) or (2.5), this in turn implies that $x^o = U^{-1}d$ and $\min_{x^o} \varphi(x^o) = f^T f$. For additional details see Tewarson (1972). If the matrix $\begin{bmatrix} RA^o \\ S \end{bmatrix}$ is large and sparse, then it is generally desirable to use sparse matrix methods to save storage and computing time (Tewarson, 1973).

Once $x^o$ is determined the new approximation for $y^*$ is $y^1 = y^o + x^o$, and we can then iterate until $y^k$ (the value of y after k such steps) is close to the desirable solution $y^*$.

We conclude this section with the observation that instead of directly solving (2.2) for $x^o$, the minimization of the functional $\varphi(x^o)$ leads not only to a desirable solution for proper choices of V and W, but also improves the conditioning of the problem (Tewarson and Narain, 1974b).

**3.** Incorporation of Known Information in the Solution Process.

The known information about the solution vector can be incorporated in the solution process in several ways. One of these is the use of matrices V and W that were introduced in the last section. Three types of information can be utilized to construct V and W: .(a) statistical, (b) smoothing and (c) structural. A brief description of each follows:

(a) Statistical Information. If the given system of equation (1.1) is linear, then **Ay** = b, since F(y) = **Ay** and $F'(y^o) = A$. Let us assume that from previous experiments of a similar nature a mean value $\bar{y}$ for y is available. Thus $\bar{y}$ may denote a mean (average) picture or a mean vertical temperature

profile. Now if we let $\bar{d} = A\bar{y}$ then it follows that $A(y - \bar{y}) = d - \bar{d}$ .or $Ax = b$, where $x = y - \bar{y}$ and $b = d - \bar{d}$. The vectors x and b are respectively the deviations of y and d from their mean values $\bar{y}$ and $\bar{d}$. If a large number of x values and the the corresponding b value are available from model calibration etc. and we let X and B denote the matrices with columns $x^{(1)}, \cdots, x^{(k)}$ and $b^{(1)}, \ldots, b^{(k)}$ respectively, and let

$$E = B - AX,$$

then the covariance matrices for y and the measurement noise E in d can be approximated by

$$S_y = \frac{XX^T}{k} \text{ and } S_\epsilon = \frac{EE^T}{k}.$$

We now take $W = S_Y^+$ and $V = S_\epsilon^+$, wher'e + denotes the Moore-Penrose generalized inverse. The reason for the above choice of V and W, the relationship with the best linear unbiased estimate, and a detailed description can be found in Strand and Westwater (1968), Fleming and Smith (1972), Tewarson (1972), Tewarson and Narain (1974a,b) and Mascarenhas and Pratt (1975). The special case $V = I$ and $W = \lambda I$ is described in Levenberg (1944), Marquardt (1963), Twomey (1965) and Tihonov (1965). Methods for computing generalized inverses are given in Tewarson (1969a, 1971).

(b) Smoothing. It is sometimes known that the vector x (or vector y) is a function of the distance from some origin and is reasonably smooth. Then $x = x(u)$, where u is the distance from the origin and therefore the functional

$$\int_\alpha^\beta [p(u)x(u)^2 + q(u) (\frac{dx(u)}{du})^2] \, du$$

should be minimized instead of just the norm of $x(u)$,(p(u) and q(u) are suitable weights). On discretization this integral leads to $x^T Wx$, where W is a tridiagonal matrix (Tihonov 1963a,b, 1965, Tewarson 1972).

(c) Structural. Once again we assume that F is linear, then in many cases it is known that the main contribution to a particular element of b comes from a small subset of the elements of x. Therefore, when predicting x from b, this relationship should be preserved. This leads to a structure-oriented

generalized inverse $A^s$ (Tewarson and Narain 1974a) such that in each row of $A^sA$ the elements get progressively smaller as one gets away from the main diagonal. See also Backus (1970), Conrath (1972), and Chen and Surmont (1975).

Having made the choice of V and W in one or more of the above mentioned ways we now focus our attention to the other methods for incorporating additional information in the solution process.

In some cases the constraints on y may be known as $H(y) = 0$ and it may be more convenient to consider solving the augmented system $W(y) = \hat{d}$, where $W(y) = \begin{bmatrix} F(y) \\ H(y) \end{bmatrix}$ and $\hat{d} = \begin{bmatrix} d \\ 0 \end{bmatrix}$. This, of course involves salving a larger system of equations than (1.1) and requiresadditional conditions for the convergence of the Newton-Raphson method. A survey of these and related techniques is given in Tewarson (1975). General theory for the solution of non-linear equations is discussed in Ortega and Rhienboldt (1970).

In all of the above mentioned methods the choice of the initial approximation $y^o$ is very critical. In any iterative method for solving non-linear (or linear) equations, $y^o$ must lie within the domain of attraction of $y^*$; this often requires $y^o$ to be very close to $y^*$ and such a choice is generally hard to make. In underdetermined problems, without any smoothing and other constraints the component of $y^*$ in the null space of the relevant iteration operator remains the same e.g., when using the Kaczmarz method (Tewarson, 1969b). Therefore this component should be reasonable in $y^o$ in order for the method to converge to a desirable solution.

We will now briefly mention some of the other methods that have been found useful in picture reconstruction and satellite temperature profile determination.

If it is known that each $x = B\gamma$, where B is some known basis (e.g., exponential functions in renal transport) and $\gamma$ is a suitable unknown, then from (2.2), neglecting the small terms, we have

$$A^oB\,\gamma = b$$

or

$$\gamma = (A^oB)^+ b + N_{A^oB}\lambda, \text{ where } N_{A^oB} \text{ is the}$$

orthogonal projector on the null space of $A^{\circ}B$ and $\lambda$ is arbitrary. Note that $(A^{\circ}B)^{+}$ can be determined by either a direct (Tewarson 1969a) or an iterative method (Tewarson 1971). Detailed descriptions of these techniques are given in Gordon and Herman (1974), Smith, Wolfe and Jacob (1970).

In the case of picture reconstruction, analytical solutions and convolution methods are given in Radon (1917), Bracewell and Riddle (1969), Herman and Rowland (1973). Fourier transform mehtods, which involve taking the Fourier transform of the projections, interpolating and then inverting the results, are discussed in Bracewell (1956), Hoppe et. al. (1968), DeRosier and Klug (1968), Crowther et. al. (1970), DeRosier (1971).

In conclusion, we again restate the fact that all possible information about the problem must be incorporated in the solution process when underdetermined problems are being solved. It is of course possible that some of this information may be of qualitative nature and hard to incorporate in a quantitative manner. The technique used to incorporate any information must be tailored to suit the particular problem one wants to solve. Being aware of the various successful methods in diverse application areas will make it easy for the user to choose the method best suited to his particular application.

References


1. Backus, G. E. (1970). Inference from inadequate and inaccurate data I, II and III. Proc. Nat. Acad. Sci. 65, 1-7, pp. 281-289.

2. Bracewell, R. N. (1956). Strip integration in radio astronomy. Aust. J. Phys. 9(2), pp. 198-217.

3. Bracewell, R. N., and Riddle, A, C. (1967). Inversion of fan-beam scans in radio astronomy. Astrophys. J. 150(2), pp. 427-434.

4. Chen, Y. M.,and Surmont, J. (1975). Iterative algorithms for non-linear problems in remote sensing. Applied Math. and Computation, (to appear).

5. Conrath, B. J. (1972). Vertical resolution of temperature profiles obtained from remote radiation measurements, J. Atmospheric Sc. 29, pp. 1262-1271.

6. Crowther, R. A., DeRosier, D. J. and Klug, A. (1970). The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Froc. Roy. Soc., Ser. A 317, pp. 319-340.

7. DeRosier, D. J. (1971). The reconstruction of three-dimensional images from electron micrographs. Contemp. Phys. 12(5), pp. 437-452.

8. DeRosier, D. J., and Klug, A. (1968). Reconstruction of three-dimensional structures from electron micrographs. Nature (London) 217, pp. 130-134.

9. Fleming, H. E., and Smith, W. L. (1972). Inversion techniques for remote sensing of atmospheric temperature profiles. Proceedings of the Fifth Symposium on Temperature, its Measurement and Control in Science and Industry, Washington, D. C.

10. Golub, G. H., and Businger, P. A. (1965). Linear least squares solutions by Householder Transformations. Numer. Math. 7 H. B. Series Linear Algebra, pp. 269-276.

11. Gordon, R., and Herman, G. T. (1974). Three-dimensional reconstruction from projections: A Review of Algorithms, Int. Rev. of Cytology 38, pp.111-151.

12. Herman, G. T., and Rowland, S. (1973). Three methods for reconstructing objects from x-rays: a comparative study. Comput. Graphics Image Process 2, pp. 151-178.

13. Hoppe, W., Langer, R., Knesch, G., and Poppe, C. (1968). Protein-Kristallstrukturanalyse mit Elektronenstrahlen. Naturwissenschaften 55, pp. 333-336.

14. Lawson, C. L., and Hanson, R. J. (1974). Solving Least Squares Problems. Prentice Hall, Englewood Cliffs, N. J.

15. Levenberg, K. (1944). A method for solution of certain non-linear problems in least squares. Q. Appl. Math. 5, pp. 164-168.

16. Marquardt, D. W. (1963). An algorithm for least squares estimation of non-linear parameters, J. Soc. Ind. Appl. Math. 11, pp. 431-441.

17. Mascarenhas, N.D.A., and Pratt, W. K. (1975). Digital image restoration under a regression model. IEEE Tran. on Circuits & Systems, CAS-22, pp. 252-266.

18. Ortega, J. M., and Rheinboldt, W. C. (1970). iterative Solution of Non-linear Equations. Academic Press, New York, N. Y.

19. Radon, J. (1917). Uber die Bestimmung von Funktionen durch ihre integralwerte längs gewisser Mannigfaltigkeiten. (On the determination of functions from their integrals along certain manifolds.) Ber.Saechs. Akad. Wiss. Leipzig, Math-Phys. Kl. 69, pp. 262-277.

20. Smith, W. L.,Woolf, H. M., and Jacob, W. J. (1970). A regression method for obtaining real-time temperature and geopotential height profiles from satellite spectrometer measurements and its application to Nimbus 3 "SIRS" Observations. Mo. Weather Rev. 98, pp. 582-603.

21. Strand, O. N., and Westwater, E. R. (1968). The statistical estimation of the numerical solution of Fredholm integral equation of the first kind. J. ACM 15, pp. 100-114.

22. Tewarson, R. P. (1969a). On computing generalized inverses, Computing 4, pp. 139-152.

23. Tewarson, R. P. (1969b). On projection methods for solving linear systems, Computer J.12, pp. 78-81.

24. Tewarson, R. P. (1971). An iterative method for computing generalized inverses, Int. J. Comp. Math. 3, pp. 65-74.

25. Tewarson, R. P. (1972). Solution of linear equations in remote sensing and picture reconstruction. Computing 10, pp. 221-230.

26. Tewarson, R. P. (1973). Sparse Matrices. Academic Press, New York, N.Y.

27. Tewarson, R. P., and Narain, P. (1974a). Generalized inverses and resolution in the solution of linear equations, Computing 13, pp. 81-88.

28. Tewarson, R. P., and Narain, P. (1974b). Solution of linear equations resulting from satellite remote soundings, J. Math. Anal. and Appl. 47, pp. 1-14.

29. Tewarson, R. P. (1975). Use of smoothing and damping techniques in the solution of non-linear equations. (submitted for publication)

30. Tihonov, A. N. (1963a). Solution of incorrectly formulated problems and the regularization method. Soviet Math. Dokl. 4, pp. 1035-1038.

31. Tihonov, A. N. (1963b). Regularization of incorrectly posed problems, Soviet Math. Dokl. 4, pp. 1624-1627.

32. Tihonov, A. N. (1965). The stability of algorithsm for the solution of degenerate systems of linear algebraic equations.USSR Computational Math. and Math. Physics 5, pp. 181-188.

33. Twomey, S. (1965). The application of numerical filtering to the solution of integral equations encountered in direct sensing measurements. J. Franklin Inst. 279, pp. 95-109.