Open access • Journal Article • DOI:10.1109/78.127968

Maximum entropy regularization in inverse synthetic aperture radar imagery

B. Borden

Published on: 01 Apr 1992 - IEEE Transactions on Signal Processing (IEEE)

Topics: Inverse synthetic aperture radar, Principle of maximum entropy, Maximum a posteriori estimation, Maximum entropy spectral estimation and Least squares

Based on (13) and (19), one obtains

$$\nabla_{w_{nm}^{(l)}}(e^2) = -2(\delta_{Rn}^{(l)} x_{Rm}^{(l-1)} + \delta_{In}^{(l)} x_{Im}^{(l-1)}) - 2j(-\delta_{Rn}^{(l)} x_{Im}^{(l-1)} + \delta_{In}^{(l)} x_{Rm}^{(l-1)}) = -2\,\Delta_n^{(l)} X_m^{*(l-1)} \quad (22)$$

as reported in (17).

From (20), we need now to find the recusion for $E_n^{(l)}$. For the last layer ($l = M$), expression (18) particularizes as (see (11) and (12)):

$$E_n^{(M)} = (d_{Rn} - x_{Rn}^{(M)}) + j(d_{In} - x_{In}^{(M)}) = E_n. \quad (23)$$

For a hidden layer ($0 < l < M$), expression (18) can be expanded using the chain rule as follows:

$$E_n^{(l)} = -\frac{1}{2}\sum_{q=1}^{N_{l+1}} \left\{ \left(\frac{\partial e^2}{\partial s_{Rq}^{(l+1)}} \frac{\partial s_{Rq}^{(l+1)}}{\partial x_{Rn}^{(l)}} + \frac{\partial e^2}{\partial s_{Iq}^{(l+1)}} \frac{\partial s_{Iq}^{(l+1)}}{\partial x_{Rn}^{(l)}} \right) + j\left(\frac{\partial e^2}{\partial s_{Rq}^{(l+1)}} \frac{\partial s_{Rq}^{(l+1)}}{\partial x_{In}^{(l)}} + \frac{\partial e^2}{\partial s_{Iq}^{(l+1)}} \frac{\partial s_{Iq}^{(l+1)}}{\partial x_{In}^{(l)}} \right) \right\}. \quad (24)$$

Using definition (19) and relationship (13) gives

$$E_n^{(l)} = \sum_{q=1}^{N_{l+1}} \{ (\delta_{Rq}^{(l+1)} w_{Rqn}^{(l+1)} + \delta_{Iq}^{(l+1)} w_{Iqn}^{(l+1)}) + j(-\delta_{Rq}^{(l+1)} w_{Iqn}^{(l+1)} + \delta_{Iq}^{(l+1)} w_{Rqn}^{(l+1)}) \} \quad (25)$$

which coincides with (15).

REFERENCES

[1] *Proc. IEEE* (Special Issue on Neural Networks), vol. 78, no. 9 (part I) and no. 10 (part II), Sept. and Oct. 1990.

[2] D. E. Rumelhart and J. L. McClelland, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, vol. I. Cambridge, MA: M.I.T. Press, 1986.

[3] N. Benvenuto, M. Marchesi, F. Piazza, and A. Uncini, "Nonlinear satellite radio links equalized using blind neural networks," in *Proc. Int. Conf. ASSP* (Toronto, Canada), May 1991, pp. 1521-1524.

[4] N. Benvenuto, M. Marchesi, F. Piazza, and A. Uncini, "A comparison between real and complex-valued neural networks in communication applications," in *Proc. Int. Conf. Artificial Neural Networks* (Espoo, Finland), June 1991, pp. 1177-1180.

[5] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline, and backpropagation," *Proc. IEEE*, vol. 78, no. 9, pp. 1415-1442, Sept. 1990.

[6] T. M. Apostol, *Mathematical Analysis*, second ed. Addison-Wesley, 1974.

[7] M. S. Kim and C. C. Guest, "Modification of back propagation networks for complex-valued signal processing in frequency domain," in *Proc. Int. Joint Conf. Neural Networks* (San Diego, CA), June 1990, pp. II 27-31.

[8] T. L. Clarke, "Generalization of neural networks to the complex plane," in *Proc. Int. Joint Conf. Neural Networks* (San Diego, CA), June 1990, pp. II 435-440.

[9] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proc. IEEE*, pp. 719-720, Apr. 1975.

Maximum Entropy Regularization in Inverse Synthetic Aperture Radar Imagery

Brett Borden

***Abstract*—The method of maximum entropy is applied to the regularization of inverse synthetic aperture radar (ISAR) image reconstructions. This is accomplished by considering an ensemble of images with associated "allowed" probability density functions. Instead of directly considering the "solution" to be an image, we take it to be the *a posteriori* probability density found by minimizing a regularization functional composed of the usual "least squares" term and a Kullback (cross-entropy) information difference term. The desired image is then found as the expectation of this density. The basic model of this approach is similar to that used in usual maximum *a posteriori* analysis and allows for a more general relationship between the image and its "configuration entropy" than is usually employed. In addition, it eliminates the need for inappropriate nonnegativity constraints on the (generally complex-valued) image.**

I. INTRODUCTION

In the weak scatterer approximation, the far-field response E, resulting from a harmonic excitation of a "target," can be modeled by a superposition of plane waves [1]

$$E(\boldsymbol{k}) = \int_D \mathcal{A}(\boldsymbol{k}, \boldsymbol{r})\epsilon(\boldsymbol{r})\, d^3r \quad (1)$$

where $\mathcal{A}(\boldsymbol{k}, \boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, $\epsilon(\boldsymbol{r})$ is proportional to the local (possibly complex) scatterer strength at the position $\boldsymbol{r}$, $\boldsymbol{k}$ is the propagation vector with magnitude $2\pi/\lambda$, and D is the support of the target. (The $e^{-i\omega t}$ time dependence has been suppressed.)

Since any practical measurement will yield only discrete and finite measurements $E_i \equiv E(\boldsymbol{k}_i)$, (1) becomes

$$E_i = \int_D \mathcal{A}(\boldsymbol{k}_i, \boldsymbol{r})\epsilon(\boldsymbol{r})\, d^3r, \qquad i = 1, \cdots, N. \quad (2)$$

When D is finite, $E(\boldsymbol{k})$ spans k-space and (2) cannot be uniquely inverted to obtain the "image" $\epsilon(\boldsymbol{r})$ from the limited set of scattered field measurements. More generally, if $\mathcal{A}(\boldsymbol{k}_i, \boldsymbol{r})$ and $\epsilon(\boldsymbol{r})$ are considered to belong to a Hilbert vector space, then we can write $\boldsymbol{E} = \boldsymbol{A}\epsilon$, where $\boldsymbol{A}$ is the linear operator corresponding to $\mathcal{A}$ which maps the function $\epsilon(\boldsymbol{r})$ to the discrete measurements E_i. Typically, ϵ will belong to a space of infinite dimension and will have a component that lies in the null space of $\boldsymbol{A}$ [2], [3]. In such cases it will be impossible to uniquely determine ϵ from (2), and we may estimate the null space component only by including some form of extra or "prior" information.

When the data are noisy so that

$$\boldsymbol{E} = \boldsymbol{A}\epsilon + \boldsymbol{n} \quad (3)$$

we are confronted with the additional complication of trying to devise a stable reconstruction technique. This is because the inverse problem associated with (1) is ill posed [4], [5]. As a result, small variations in the data may be mapped to large variations in ϵ.

Manuscript received February 27, 1991; revised September 29, 1991. This work was supported by the Office of Naval Research.

The author is with the Research Department, Naval Air Warfare Center, China Lake, CA 93555-6001.

IEEE Log Number 9105998.

In the following, we briefly review the principle of maximum entropy regularization and show how it can be used to select that image which is known to be maximally noncommittal with regard to measurement limitations and which is consistent with the measured data and noise [11]. Usual maximum entropy methods require the image to be composed of nonnegative values and impose a simple relationship between the image space and its ensemble properties. This requirement is not well suited to data defined by (1), and we explore an alternate method which uses a more general image/ensemble mapping. Then we apply the method to the problem of inverse synthetic aperture array (ISAR) imagery.

II. Regularization

When the data are merely limited (and not noisy) we often expect least squares solutions to (2) to yield adequate results. Letting $\mathfrak{X}$ denote the set of all solutions which are consistent with the prior information, a least squares solution ϵ will satisfy

$$\|\boldsymbol{E} - \boldsymbol{A}\epsilon\|^2 = \min_{u \in \mathfrak{X}} \|\boldsymbol{E} - \boldsymbol{A}\boldsymbol{u}\|^2. \tag{4}$$

Similarly, for a solution considered to be a member of an ensemble of possible solutions, with P the set of all allowed ensemble probability density distributions, the least squares density distribution p of interest is defined to obey

$$\|\boldsymbol{E} - \boldsymbol{A}\langle \epsilon \rangle_p\|^2 = \min_{q \in P} \|\boldsymbol{E} - \boldsymbol{A}\langle \epsilon \rangle_q\|^2 \tag{5}$$

where the expected image $\langle \epsilon \rangle_p \equiv \int \epsilon p(\epsilon)\, d^N\epsilon$ is determined uniquely by the "solution" density distribution $p(\epsilon)$.

When the data are noisy the mechanics of determining the set P can be complicated by the requirement that the allowed solutions be properly smoothed. This restriction may run contrary to the information implied by the measured data and functional regularization methods are frequently employed. If p_0 denotes a smooth *a priori* distribution then, in terms of (5), our regularized solution p will minimize

$$J(p) = \|\boldsymbol{E} - \boldsymbol{A}\langle \epsilon \rangle_p\|^2 + \alpha J_2(p, p_0), \qquad p \in P \tag{6}$$

where $J_2(p, p_0)$ is a regularization functional measuring the "distance" between p and p_0, and the regularization parameter $\alpha > 0$ controls the extent to which J_2 dominates J.

Often, regularization methods concentrate directly on smoothing the least squares solution ϵ and many possible appropriate regularization functionals have been examined [4], [5]. Maximum entropy methods exploit the "information" difference between p and p_0 as the "distance" measure and set

$$J_2(p, p_0) = \int p(\epsilon) \ln\left(\frac{p(\epsilon)}{p_0(\epsilon)}\right) d^N\epsilon. \tag{7}$$

(This is the "Kullback distance" or "cross entropy" and is the negative of the entropy of $p(\epsilon)$ when the "prior" $p_0(\epsilon)$ is a uniform distribution [6], [7].) It has been demonstrated that the choice (7) assures that solutions which minimize (6) will be least biased by unwarranted assumptions about the data [8], [9].

In this formalism, the prior $p_0(\epsilon)$ holds the burden of introducing both smoothness and null space information. From a Bayesian view, the measurement process serves only to refine our prior knowledge and so $p_0(\epsilon)$ also helps restrict the set of allowed solutions to be considered. The *a posteriori* density $p(\epsilon)$ is related to the *a priori* density by Bayes' rule [10], [11]

$$p(\epsilon \mid \boldsymbol{E}) = \frac{p_{E|\epsilon}(\boldsymbol{E} \mid \epsilon) p_0(\epsilon)}{p_E(\boldsymbol{E})} \tag{8}$$

where $p_0(\epsilon)$ is the *a priori* density of ϵ, $p_E(\boldsymbol{E})$ is the *a priori* density of the data $\boldsymbol{E}$, and $p_{E|\epsilon}(\boldsymbol{E} \mid \epsilon)$ is the *a posteriori* density of $\boldsymbol{E}$ given ϵ.

The most common maximum entropy approach takes p_0 as constant, and $p_{E|\epsilon}$ as proportional to ϵ (which are further constrained to be real and nonnegative). The resulting J_2 from (7) is then the negative of the so-called configuration entropy of the image and the noisy data problem is treated by adding additional (or modified) contraints (cf. [12], [13]). In [14] more general p_0 are used in (7) in a maximum entropy analysis of the problem of spectral estimation (which relates p to the autocovariance of ϵ). Bayesian approaches have also been applied (often without entropy considerations [11]) but typically require good estimates of the properties of p_0. (A more complete survey of these issues and related work may be found in [3].)

ISAR imagery data consist of complex scattered field measurements and so the usual (simple proportional) relationship between ϵ and $p(\epsilon)$ is inappropriate. Moreover, we frequently have only minimal information about p_0 which, in principle, we would nevertheless like to be more general than a simple uniform distribution. Since our resulting model may depend in a sensitive way on p_0, we seek to make our results maximally noncommittal to our prior assumptions.

III. The Algorithm

Our basic problem is to determine the ensemble density distribution p which is consistent with any prior knowledge and which minimizes (6) over any unknowns. The connection between this solution p and the resulting image is given as the expectation $\langle \epsilon \rangle_p$.

We model the probability density of the noise by a Gaussian:

$$p_n(\boldsymbol{n}) = \frac{1}{(2\pi)^{N/2}(\det \boldsymbol{R}_n)^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{n}^\dagger \boldsymbol{R}_n^{-1}\boldsymbol{n}\right) \tag{9}$$

where $\boldsymbol{R}_n$ is the (assumed known) noise correlation matrix, $(^\dagger)$ denotes complex-conjugate transposition, and the mean noise is assumed to be zero. This gives the measurement probability density as

$$p_{E|\epsilon}(\boldsymbol{E} \mid \epsilon) = \frac{1}{(2\pi)^{N/2}(\det \boldsymbol{R}_n)^{1/2}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{E} - \boldsymbol{A}\epsilon)^\dagger \boldsymbol{R}_n^{-1}(\boldsymbol{E} - \boldsymbol{A}\epsilon)\right). \tag{10}$$

The *a priori* density for the solution ensemble is here assumed to also obey a Gaussian law:

$$p_0(\epsilon) = \frac{1}{(2\pi)^{N/2}(\det \boldsymbol{R}_\epsilon)^{1/2}} \exp\left(-\frac{1}{2}(\epsilon - \epsilon_0)^\dagger \boldsymbol{R}_\epsilon^{-1}(\epsilon - \epsilon_0)\right) \tag{11}$$

where ϵ_0 is the prior expectation.

The prior $p_E(\boldsymbol{E})$ is independent of ϵ and may be absorbed into the normalization coefficient. Application of Bayes' rule yields

$$p(\epsilon \mid \boldsymbol{E}) = \frac{(\det(\boldsymbol{R}_\epsilon^{-1} + \boldsymbol{A}^\dagger \boldsymbol{R}_n^{-1}\boldsymbol{A}))^{1/2}}{(2\pi)^{N/2}} \cdot \exp\left(-\frac{1}{2}(\epsilon - \langle \epsilon \rangle_p)^\dagger (\boldsymbol{R}_\epsilon^{-1} + \boldsymbol{A}^\dagger \boldsymbol{R}_n^{-1}\boldsymbol{A})(\epsilon - \langle \epsilon \rangle_p)\right) \tag{12}$$

where $\langle \epsilon \rangle_p$ obeys

$$(R_\epsilon^{-1} + A^\dagger R_n^{-1} A)(\langle \epsilon \rangle_p - \epsilon_0) = A^\dagger R_n^{-1}(E - A\epsilon_0). \quad (13)$$

Maximum *a posteriori* image restoration (MAP) maximizes (12) with respect to ϵ [10], [11]. In the present case, this leads to a recursive relation that must be satisfied by the image $\langle \epsilon \rangle_p$:

$$\langle \epsilon \rangle_p = \epsilon_0 + R_\epsilon A^\dagger R_n^{-1}(E - A\langle \epsilon \rangle_p). \quad (14)$$

(Note that this is equivalent to (13).)

In MAP analysis, the smoothing and null space information are imposed through the selection of ϵ_0 and R_ϵ. However, often the details of these ensemble quantities are not correctly known and must be guessed. The practical result is that either ϵ_0 or R_ϵ (sometimes both) are treated as tuning parameters and adjusted toward some final desired image [3]. Incorrect or trivial prior assumptions (particularly on R_ϵ) may lead to unwarranted features in the resultant image.

In our present model ϵ_0 and R_ϵ are independent. We choose to introduce the *a priori* image information through ϵ_0 alone, and to select the R_ϵ which are maximally noncommittal in the sense that they are determined by minimizing (6). Substituting (12) into (7), the total cost functional (6) becomes

$$\begin{aligned} J(p) = \|E - A\langle \epsilon \rangle_p\|^2 &+ \frac{\alpha}{2} \ln (\det (R_\epsilon^{-1} + A^\dagger R_n^{-1} A)) \\ &- \frac{\alpha}{2} \ln (\det R_\epsilon^{-1}) + \frac{\alpha}{2} \operatorname{tr} (R_\epsilon^{-1}(R_\epsilon^{-1} + A^\dagger R_n^{-1} A)^{-1}) \\ &+ \frac{\alpha}{2} (\langle \epsilon \rangle_p - \epsilon_0)^\dagger R_\epsilon^{-1}(\langle \epsilon \rangle_p - \epsilon_0) + \text{constant}. \end{aligned} \quad (15)$$

With the ϵ_0 fixed (and because of the relation (13)), the functional $J(p)$ depends completely on the unknowns R_ϵ^{-1}. Its minimization is determined by requiring

$$\frac{\partial J}{\partial R_\epsilon^{-1}} = 0. \quad (16)$$

A straightforward calculation yields

$$R_\epsilon[A^\dagger R_n^{-1} A R_\epsilon A^\dagger R_n^{-1} - A^\dagger R_n^{-1}(E - A\epsilon_0)D^\dagger]A = 0 \quad (17)$$

where $D \equiv (4I/\alpha - R_n^{-1})(E - A\langle \epsilon \rangle_p) + R_n^{-1} A(\langle \epsilon \rangle_p - \epsilon_0)$. (This result has been significantly simplified by applying the relation (13) whenever possible.)

By construction, $A^\dagger$ is of full column rank and so

$$A^\dagger R_n^{-1} A R_\epsilon A^\dagger R_n^{-1} = A^\dagger R_n^{-1}(E - A\epsilon_0)D^\dagger. \quad (18)$$

Operating on $(E - A\langle \epsilon \rangle_p)$ and simplifying yields the weighted "normal equations"

$$A^\dagger R_n^{-1} A(\langle \epsilon \rangle_p - \epsilon_0) = Q_\alpha(\langle \epsilon \rangle_p) A^\dagger R_n^{-1}(E - A\epsilon_0) \quad (19)$$

where $Q_\alpha(\langle \epsilon \rangle_p) \equiv D^\dagger(E - A\langle \epsilon \rangle_p)$.

Because $A^\dagger$ is of full column rank, $A^\dagger R_n^{-1} A$ will not have an inverse (except in the trivial case when A is square). However, since we have chosen to introduce the null space information through ϵ_0, we can work with generalized inverses [15]. Here, $(AA^\dagger)^{-1}$ does exist and $A^\dagger(AA^\dagger)^{-1} R_n (AA^\dagger)^{-1} A$ is a generalized inverse of $A^\dagger R_n^{-1} A$. Applying this inverse yields

$$\langle \epsilon \rangle_p = \epsilon_0 + Q_\alpha A^\dagger(AA^\dagger)^{-1}(E - A\epsilon_0) \quad (20)$$

which is easily seen to satisfy (19). Moreover, since ϵ_0 is presumed to contain all the null space information, (20) is the unique choice of solution to (19).

Since (19) can be rewritten in the form

$$A^\dagger R_n^{-1}[(1 - Q_\alpha(\langle \epsilon \rangle_p))(E - A\epsilon_0) - (E - A\langle \epsilon \rangle_p)] = 0 \quad (21)$$

it is easy to see that $E - A\langle \epsilon \rangle_p$ is parallel to $E - A\epsilon_0$ (when both are nonvanishing). This allows us to express Q_α in terms of $E - A\epsilon_0$. Substituting $E - A\langle \epsilon \rangle_p = (1 - Q_\alpha(\langle \epsilon \rangle_p))(E - A\epsilon_0)$ into the definition of Q_α results in a quadratic relation with roots given by

$$Q_\alpha = 1 - \frac{\beta + 1 \pm [(\beta + 1)^2 - 4(2\beta - 4\gamma/\alpha)]^{1/2}}{2(2\beta - 4\gamma/\alpha)} \quad (22)$$

where $\beta \equiv (E - A\epsilon_0)^\dagger R_n^{-1}(E - A\epsilon_0)$ and $\gamma \equiv (E - A\epsilon_0)^\dagger(E - A\epsilon_0)$.

Comparison of (20) with (13) allows us to identify

$$R_\epsilon = \frac{Q_\alpha}{1 - Q_\alpha} A^\dagger(AA^\dagger)^{-1} R_n (AA^\dagger)^{-1} A. \quad (23)$$

Equation (23) can be substituted into (15) to select for the Q_α (determined by (22)) which yields the smaller value of $J(p)$. This minimizing Q_α, together with (20), represents our regularized solution.

IV. A Method for Selecting the Prior Image

So far we have assumed that the prior image ϵ_0 is sufficiently complete that a solution based on (20) will be an improvement over the ordinary inversion of (3). However, usually we will have only partial *a priori* information upon which to base our solution. A common prior assumption is that the target is composed of point scatterers superimposed on a flat background and so

$$\epsilon_0(r) \equiv \sum_{j=1}^{M} a_j\, \delta(r - r_j) \quad (24)$$

where M is the (unknown) number of scatterers, the jth having complex amplitude a_j and location r_j.

If we require ϵ_0 in (20) to be of the form (24) then we can systematically determine the locations and strengths of the scatterers. The algorithm is as follows:

- Step 0. Set $\epsilon_0 = \mathbf{0}$.
- Step 1. Calculate $y \equiv A^\dagger(AA^\dagger)^{-1}(E - A\epsilon_0^{(i)})$.
- Step 2. Calculate $Q_\alpha(\epsilon_0^{(i)})$ and check against threshold τ; if $Q_\alpha < \tau$ then go to step 5.
- Step 3. Replace y by $y' \equiv a_j\, \delta(r - r_j)$ where a_j is the complex amplitude of max (y) and r_j is its location.
- Step 4. Set $\epsilon_0^{(i+1)} = \epsilon_0^{(i)} + Q_\alpha y'$ and go to step 1.
- Step 5. Done; set $\langle \epsilon \rangle_p = \epsilon_0 + Q_\alpha y$.

Additional firm prior information may be incorporated by altering the "seed" in step 0 ($\epsilon_0 = \mathbf{0}$ is interpreted to mean that scatterer strength and location are completely unknown).

This iterative refinement is Bayesian in intent and results in an approach which is similar to that of the CLEAN technique [16] except in the way in which noise is dealt with and convergence is determined. Here we halt the iterations when $Q_\alpha = Q_\alpha(\epsilon_0)$ becomes sufficiently small. Convergence is guaranteed since $0 < Q_\alpha < 1$.

V. ISAR Imagery from a Rotating Target

For a target with rotational position $\theta(t)$ in the far field of the transmitter/receiver, (1) becomes [1]

$$E(k) = \int_D \epsilon(x, y) \exp\, [i2k(y \cos\theta - x \sin\theta)]\, dx\, dy \quad (25)$$

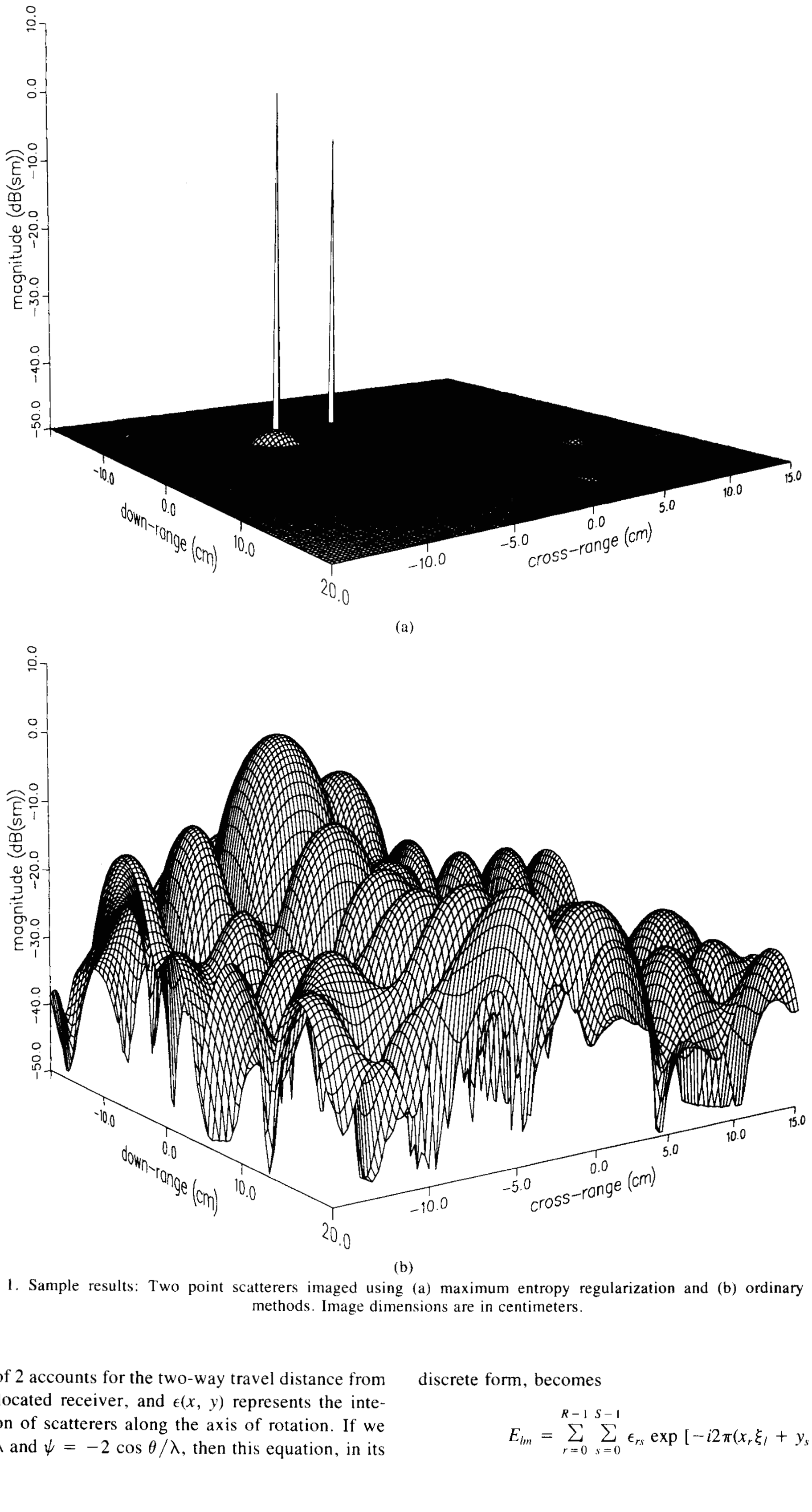

Fig. 1. Sample results: Two point scatterers imaged using (a) maximum entropy regularization and (b) ordinary Fourier methods. Image dimensions are in centimeters.

where the factor of 2 accounts for the two-way travel distance from transmitter to colocated receiver, and $\epsilon(x, y)$ represents the integrated contribution of scatterers along the axis of rotation. If we set $\xi = 2 \sin \theta/\lambda$ and $\psi = -2 \cos \theta/\lambda$, then this equation, in its discrete form, becomes

$$E_{lm} = \sum_{r=0}^{R-1} \sum_{s=0}^{S-1} \epsilon_{rs} \exp\left[-i2\pi(x_r \xi_l + y_s \psi_m)\right]. \tag{26}$$

TABLE I
ITERATION DATA FOR THE RECONSTRUCTION OF FIG. 1.
$\sigma^2 = 0.0225$ AND $\alpha = 0.095$.

Iteration #	γ	Q_α
1	86.06	0.99974
2	17.98	0.99875
3	2.750	0.05041

In the small angle approximation ($\sin\theta \approx \theta$ and $\cos\theta \approx 1$) (26) allows the usual discrete Fourier transform approximation in which $AA^\dagger = RSI$. If, in addition, we take $R_n = \sigma^2 I$, the equations simplify significantly and

$$\langle \epsilon \rangle_p = \epsilon_0 + Q_\alpha A^\dagger/(RS)(E - A\epsilon_0). \tag{27}$$

Fig. 1(a) illustrates a reconstruction based on this analysis. For comparison purposes, the usual inverse Fourier transform (DFT) reconstruction (Fig. 1(b)) has also been included. Synthetic data appropriate to a target consisting of two ideal point scatterers of amplitudes 1.0 and 0.5 were constructed for each of 8 angles and 8 frequencies. The angles were uniformly distributed over 20° of aspect and the frequencies were uniformly distributed over a 4-GHz bandwidth centered on 12 GHz. In addition, the data were contaminated with additive Gaussian noise of zero mean and standard deviation $\sigma = 0.1 \times$ maximum value of the data. For this reconstruction the standard deviation σ of the noise was set to its true value. The image space was 128×128 elements and the regularization parameter was set to $\alpha = 0.095$.

To better understand the regularization process we have included the corresponding iteration data in Table I. The two scatterers were reconstructed with $Q_\alpha \approx 1$ while the remaining iteration (associated with spurious, noise induced, scatterers) uses $Q_\alpha \approx 0.05$.

VI. DISCUSSION AND CONCLUSION

A method for applying maximum entropy analysis to ISAR image reconstruction when the data are incomplete and noisy has been developed and briefly examined. These data can be quite general, and need not be assumed to be positive definite. The method yields a nonrecursive solution which readily allows *a priori* image information to be included (in fact, it requires it). However, it requires less prior information than the corresponding MAP approach and typically should also offer a reduced computational burden. We have also shown how prior information may be recursively included into the algorithm.

By example, we have demonstrated that the technique depends sensitively upon the information contained in the *a priori* image—the more completely this information can be determined, the more accurately and free from artifact will be the regularized solution image. The algorithm is rather sensitive to the choice of regularization parameter α, however, which we typically set to $\alpha \approx 4\sigma^2$.

REFERENCES

[1] D. L. Mensa, *High Resolution Radar Imaging*. Dedham, MA: Artech House, 1981.
[2] K. Tanabe, "Projection method for solving a singular system of linear equations and its applications," *Numer. Math.*, vol. 17, pp. 203–214, 1971.
[3] K. M. Hanson, "Bayesian and related methods in image reconstruction from incomplete data," in *Image Recovery: Theory and Practice*, Henry Stark, Ed. Orlando, FL: Academic, 1987.
[4] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: V. H. Winston and Sons, 1977.
[5] M. Zuhair Nashed, "Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 220–231, 1981.
[6] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.
[7] Arthur Hobson, "A new theorem of information theory," *J. Stat. Phys.*, vol. 1, pp. 383–391, 1969.
[8] Edwin T. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939–952, 1982.
[9] E. T. Jaynes, "Information theory and statistical mechanics I," *Phys. Rev.*, vol. 106, pp. 620–630, 1957.
[10] H. Hurwitz, Jr., "Entropy reduction in Bayesian analysis of measurements," *Phys. Rev. A*, vol. 12, pp. 698–706, 1975.
[11] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Comput.*, vol. C-26, pp. 219–229, 1977.
[12] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686–690, 1978.
[13] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.*, vol. 62, pp. 511–518, 1972.
[14] J. E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 230–237, 1981.
[15] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*. New York: Wiley, 1974.
[16] J. Tsao and B. D. Steinberg, "Reduction of sidelobe and speckle artifacts in microwave imaging: The CLEAN technique," *IEEE Trans. Antennas Propagat.*, vol. 36, pp. 543–556, 1988.

Integral Inequality Bounding the Weighted Absolute Deviation of an n-Dimensional Function

Alan C. Bovik

***Abstract*—We state and prove an integral inequality that bounds the weighted integrated absolute deviation of a differentiable n-dimensional real function over an interval, relative to any value the function takes within the interval. Examples illustrate the utility of the inequality. In particular, the inequality is shown to be applicable to certain set-theoretic signal restoration algorithms, which project an observed (degraded) signal onto a closed, convex prototype set defined by a linear filter and a suitable bound.**

We denote points in $\boldsymbol{R}^n$ by $\boldsymbol{t} = (t_1, t_2, \cdots, t_n)$, $d\boldsymbol{t} = dt_1, dt_2 \cdots dt_n$, and the partial derivatives of $v: \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ differentiable at $\boldsymbol{t}$ by $(\partial/\partial t_i)v(\boldsymbol{t}) = v^{(i)}(\boldsymbol{t})$, unless $n = 1$, in which case $\partial v(t)/\partial t = v'(t)$.

Theorem: Let $w: \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ be any function for which $|t_i w(\boldsymbol{t})|^p$ is integrable on $\boldsymbol{R}^n$ for $i = 1, \cdots, n$, and let $v: \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ be any continuously differentiable function for which $|v^{(i)}(\boldsymbol{t})|^q$ is integrable on $\boldsymbol{R}^n$ for $i = 1, \cdots, n$, where $q > n$ and $(1/p) + (1/q) = 1$. Then

$$\int_{\boldsymbol{R}^n} |w(\boldsymbol{t} - \boldsymbol{c})|\,|v(\boldsymbol{t}) - v(\boldsymbol{c})|\, d\boldsymbol{t} \le \left(\frac{q}{q-n}\right) \cdot \Delta_p(w) \cdot \delta_q(v) \tag{1}$$

Manuscript received January 15, 1991; revised September 9, 1991.
The author is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712-1084.
IEEE Log Number 9105999.

1053-587X/92$03.00 © 1992 IEEE