INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **34** (2001) 1271–1283 www.iop.org/Journals/ja PII: S0305-4470(01)19639-3

A generalized information theoretical approach to tomographic reconstruction

Amos Golan[1,3] **and Volker Dose**[2]

[1] American University, Roper 200, 4400 Massachusetts Ave, NW Washington, DC 20016-8029, USA
[2] Centre for Interdisciplinary Plasma Science, Max-Planck-Institut für Plasmaphysik, EURATOM Association, D-85748 Garching b. München, Germany

E-mail: agolan@american.edu

Received 16 November 1999, in final form 4 January 2001

Abstract

A generalized maximum-entropy-based approach to noisy inverse problems such as the Abel problem, tomography or deconvolution is presented. In this generalized method, instead of employing a regularization parameter, each unknown parameter is redefined as a proper probability distribution within a certain pre-specified support. Then, the joint entropies of both, the noise and signal probabilities, are maximized subject to the observed data. After developing the method, information measures, basic statistics and the covariance structure are developed as well. This method is contrasted with other approaches and includes the classical maximum-entropy formulation as a special case. The method is then applied to the tomographic reconstruction of the soft x-ray emissivity of the hot fusion plasma.

PACS numbers: 4230W, 0250F, 0250P

1. Introduction

A common problem in analysing physical systems is limited data. In most cases, these data are in terms of a small number of expectation values, or moments, representing the distribution governing the system or process to be analysed. Examples of limited data problems include: reconstructing an image from incomplete and noisy data; reconstruction problems in emission tomography such as estimating the spatial intensity function based on projection data; estimation of the local values of some quantity of interest from available chordal measurements of plasma diagnostics (the discrete Abel inversion problem); and simply reconstructing a discrete probability distribution from a small number of observed moments.

When working with such limited data, an inversion procedure is required in order to obtain estimates of the unobserved parameters. But with limited data the problem may be

[3] NSF/ASA/Census Senior Research Fellow.

underdetermined, or ill-posed. Thus, given this underdetermined inversion problem, the estimation problem can be thought of as choosing a criterion that picks out a unique distribution out of the infinitely many (distributions) satisfying the data. The maximum-entropy (ME) regularization provides an approach for analysing such noisy inverse problems. The ME formalism was introduced by Jaynes (1957, 1963) and is based on Shannon's entropy measure. This formalism was extended and applied by a large number of researchers (e.g. Levine 1980, Levine and Tribus 1979, Tikochinsky *et al* 1980, Skilling 1989, Hanson and Silver 1995, Golan *et al* 1996). Axiomatic foundations for this approach were developed by Shore and Johnson (1984), Skilling (1989) and Csiszar (1991).

In most of the above work, the objective is to estimate the most conservative distribution given some limited number of moments or expectation values. These moments are assumed to be *perfect* in the sense that what is measured or observed in an experiment must be the correct value. In other words, the sample and population (correct but unknown) moments are taken to be identical. Consider, for example, an experiment where the data are imperfectly measured, or similarly, a set of experiments where the observed data may be noisy. In these examples the observed moments may differ from the exact (unknown) moments. Therefore, what is really observed in these cases is the set of *noisy data* or *noisy moments*.

The objective of this paper is to introduce, explore and test a generalized inversion procedure for the tomography problem that adjusts for the noise in the observed data. This generalized inversion procedure is an extension of the ME formalism of Jaynes (1957) and the information theoretical approach of Levine (1980). It provides a more conservative inference from noisy data, does not employ the traditional regularization parameter, and is free of distributional assumptions.

The tomography problem formulation and suggested inversion procedure are discussed in section 2. In section 3, the experiment and results are presented. Some concluding remarks are discussed in section 4.

2. Reconstruction of the tomography problem

2.1. The tomography problem setting and solution

Consider the following discretized tomography model:

$$s_k = \sum_{i,j} p_{ij}^k E_{ij} + \varepsilon_k \tag{1}$$

where $\boldsymbol{s}$ is a K-dimensional vector of observed signals recorded by detector k, P is a $(J \times I)$ known matrix of the proportion of the emission E_{ij} accumulated in detector k, E_{ij} is a $(J \times I)$ matrix of unknowns to be recovered with the property that $E_{ij} \geqslant 0$, and $\boldsymbol{\varepsilon}$ is a vector of independently and identically distributed noise.

The objective is to recover E from the aggregated noisy data $\boldsymbol{s}$. However, since the number of unknowns outnumbers the number of data points this problem is ill-posed (underdetermined) and one needs to choose a certain criterion to reduce the problem to a well posed problem. The classical maximum-entropy formalism provides such an inversion procedure where one maximizes the entropy of E subject to the available information. See, for example, Ertl *et al* (1996) for applying the ME method to analysing the tomography problem.

In this paper, we modify the above approach in two basic ways. First, we directly accommodate for the noise in the data. Second, we view the constraints in a more relaxed way. That is, the constraints represent a unique relationship between the observed sample's moments and the population's moments, but the two sets of moments may be slightly different. Thus, unlike other theoretical approaches, the constraints here do not represent exact conservation laws; they represent noisy and aggregated data, or a relaxed conservation law. To construct this type of generalized model within the ME formalism and with minimum distributional assumptions, both E_{ij} and ε_ks need to be transformed into proper probabilities. Following Golan *et al* (1996) and Golan (1998) the model is represented as

$$s_k = \sum_{i,j} p_{ij}^k E_{ij} + \varepsilon_k = \sum_{i,j,m} p_{ij}^k q_{ijm} z_m + \sum_l w_{kl} v_l \tag{2}$$

where $\boldsymbol{q}_{ij}$ is an M-dimensional vector of weights satisfying

$$\sum_m q_{ijm} = 1 \tag{3a}$$

and where

$$\sum_m z_m q_{ijm} \equiv E_{ij}. \tag{3b}$$

The vector $\boldsymbol{z}$ is an M-dimensional support space with $M \geqslant 2$ specified such that $\boldsymbol{z}' = (z_1, \ldots, z_M)$ where the prime denotes a 'transpose'. If for physical arguments $E_{ij} \geqslant 0$, then all z_m must be non-negative. Similarly, $\boldsymbol{w}_k$ is an L-dimensional vector of weights satisfying

$$\sum_l w_{kl} = 1 \qquad \sum_l v_l w_{kl} \equiv \varepsilon_k. \tag{4}$$

The vector $\boldsymbol{v}$ is an L-dimensional support space with $L \geqslant 2$ and is symmetric around zero. Specifically, $\boldsymbol{v}' = (v_1, \ldots, 0, \ldots, v_L)$ where $v_1 = -v_L$.

Building on the classical ME formalism, under the new generalized ME (GME) formalism one maximizes the joint entropies of the signal, $\{\boldsymbol{q}_{ij}\}$, and the noise, $\{\boldsymbol{w}_k\}$, subject to the available data and the proper distributions' requirements. For generalization, the model is formulated in terms of the cross entropy (CE) which uses Kullback's (1959) divergence measure that takes into account the possible prior information for the set $\{\boldsymbol{q}_{ij}\}$, $\{\boldsymbol{q}_{ij}^0\}$, if such information exists. The priors $\boldsymbol{w}_k^0$ for the noise are taken to be *uniform* over the symmetric support $\boldsymbol{v}$. This generalized inversion procedure is

$$\min_{\boldsymbol{q},\boldsymbol{w}} I\left(\boldsymbol{q}, \boldsymbol{w}; \boldsymbol{q}^0, \boldsymbol{w}^0\right) = \left\{ \sum_{i,j,m} q_{ijm} \log(q_{ijm}/q_{ijm}^0) + \sum_{k,l} w_{kl} \log(w_{kl}/w_{kl}^0) \right\} \tag{5}$$

subject to (2), $\sum_m q_{ijm} = 1$ and $\sum_l w_{kl} = 1$.

The optimization yields

$$q_{ijm}^* = \frac{q_{ijm}^0 \exp\left(\sum_k \lambda_k^* p_{ij}^k z_m\right)}{\sum_m q_{ijm}^0 \exp\left(\sum_k \lambda_k^* p_{ij}^k z_m\right)} \equiv \frac{q_{ijm}^0 \exp\left(\sum_k \lambda_k^* p_{ij}^k z_m\right)}{\Omega_{ij}\left(\lambda^*\right)} \tag{6}$$

and

$$w_{kl}^* = \frac{w_{kl}^0 \exp\left(\lambda_k^* v_l\right)}{\sum_l w_{kl}^0 \exp\left(\lambda_k^* v_l\right)} \equiv \frac{w_{kl}^0 \exp\left(\lambda_k^* v_l\right)}{\Psi_k\left(\lambda^*\right)} \tag{7}$$

where λ^* are the optimal (estimated) Lagrange multipliers associated with (2).

Given this solution (6) and (7) and building on the Lagrangian, the dual, concentrated, unconstrained problem is

$$
\begin{aligned}
L(\lambda) = \sum_k s_k \lambda_k - \sum_{i,j} \log \left[\sum_m q^0_{ijm} \exp\left(\sum_k \lambda^*_k p^k_{ij} z_m \right) \right] \\
- \sum_k \log \left[\sum_l w^0_{kl} \exp\left(\sum_l \lambda^*_k v_l \right) \right] \\
= \sum_k s_k \lambda_k - \sum_{i,j} \log \Omega_{ij}(\lambda) - \sum_k \log \Psi_k(\lambda). \qquad (8)
\end{aligned}
$$

Taking the first derivatives of (8) with respect to λ and equating to zero, yields the optimal λ's which in turn yield $\{\boldsymbol{q}^*_{ij}\}$ and $\{\boldsymbol{w}^*_k\}$.

A main feature of this generalized method, is that by introducing the noise and the additional set of probability distributions $\{\boldsymbol{w}\}$, the level of complexity and computation time (relative to the classical ME or any other inversion procedure) does not increase. This is because the basic number of the parameters of interest, the Lagrange multipliers, remains unchanged. That is, regardless of the number of points in the support spaces (including the continuous case) there are K Lagrange multipliers and both $\{\boldsymbol{q}_{ij}\}$ and $\{\boldsymbol{w}_k\}$ are unique functions of these parameters. This follows directly from (8).

Under this generalized criterion function (the primal 5 or the dual 8), the objective is to minimize the joint entropy distance between the data and the priors, or the data and the uniform distributions if no priors are used. As such, the noise is always pushed toward zero but is not forced to be exactly zero for each data point. It is a 'dual-loss' objective function that assigns equal weights to prediction and precision. Equivalently, it can be viewed as a shrinkage inversion procedure that simultaneously shrinks the data to the priors (or uniform distributions in the ME case) and the errors to zero. Before continuing, a short comparison with other inversion procedures is briefly discussed.

Within the more traditional maximum-entropy approach, the above noisy problem is handled as a regularization method (e.g. Donoho *et al* 1992, Csiszar 1991, 1996, Bercher *et al* 1996, Ertl *et al* 1996, Gzyl 1998). One approach to recovering the unknown probability distribution/s is to subtract from the entropy objective, or from the negative of the cross-entropy objective, $I(\boldsymbol{q}, \boldsymbol{w}; \boldsymbol{q}^0, \boldsymbol{w}^0)$, a regularization term penalizing large values of the errors. Typically, the regularization term is chosen to be $\delta \sum_k \varepsilon^2_k$, where ε_k is defined by (1) and $\delta > 0$ is the regularization parameter to be determined prior to the optimization (e.g. Donoho *et al* 1992), or is iteratively determined during repeated optimization cycles (e.g. Ertl *et al* 1996). This approach, or variations of it, works fine when we know the underlying distribution of the random variables and the exact variances (the population variances). For example, the assumption of Gaussian noise is frequently used. This assumption leads to the χ^2 distribution with a number of (random) constraints minus one degree of freedom. In that case, the regularization problem is reduced to a reconstruction of the $\boldsymbol{q}_{ij}$'s for say, some $\chi^2_{(K-1)} < \chi^2_{(K-1),0.95}$.

The solution to this class of noisy models was also approached via the maximum entropy in the mean formulation (e.g. Navaza 1986, Bercher *et al* 1996). In that formulation the underlying philosophy is to specify an additional convex constraint set that bounds the possible solutions. Each observation s_k is considered to be a mean of the process $\boldsymbol{q}_{ij}$ under some distribution defined over this set. However, distributional assumptions are still necessary and a possible knowledge of the constraint set is essential. The maximum entropy in the mean approach was generalized further (e.g. Gzyl 1998) for finding the ME distribution/s from noisy data with no assumption on the nature of the noise. But even under this innovative approach, a regularization term is still needed and some assumptions on the true value of the variances are used. In the approach taken here, no such assumptions are used. The only assumption used is that the sample errors are bounded (in some interval) within the set $\boldsymbol{v}$. That is, the upper and lower bounds on the errors' supports replace the more 'traditional' regularization parameter. In that view, this different and more relaxed regularization is not as restrictive since it is done simultaneously for each observation individually where the values for these bounds are taken directly from the sample. This is shown explicitly in the next section. To demonstrate the performance of this generalized method a small number of numerical simulations are discussed in the appendix.

In a different class of estimation approaches, such as those discussed by Csiszar (1991, 1996), the main idea is to convert the noisy system into a system of inequalities. This can be done if the bounds on the magnitude of the errors are known, but increases the complexity level of the solution. Other studies suggest the more traditional regularization approach, but not within the ME formalism. Again, these different approaches require the specification of some 'smoothing' parameter to reflect the available non-sample information for each particular problem (e.g. Donoho *et al* 1992). To summarize, under the assumption of normally distributed errors with a certain and known variance, the regularization approach works well for a large enough data set. However, if the noise process is unknown, or if one is faced with a small data set or a small number of experimental data, this assumption may be incorrect and hard to confirm and may result in inconsistent estimates as well as high variances.

Finally, to see the relationship with the more traditional (and parametric) inversion approaches, assume that both $\{\boldsymbol{q}_{ij}\}$ and $\{\boldsymbol{w}_k\}$ are characterized by the parametric exponential (logit) distribution. (Note that for computational reasons, it is impossible to characterize the distributions for $\{\boldsymbol{q}_{ij}\}$ and $\{\boldsymbol{w}_k\}$ by the normal distribution.) Then, substitute these parametric specifications in the discrete likelihood (multiplicity factor) objective to obtain a 'generalized log-likelihood' function of the form

$$L(\boldsymbol{\beta}) = \sum_k s_k \beta_k - \sum_{i,j} \log \Omega_{ij}(\boldsymbol{\beta}) - \sum_k \log \Psi_k(\boldsymbol{\beta}) \tag{8a}$$

with $\beta = -\lambda$ and with uniform priors for both $\boldsymbol{q}_{ij}$ and $\boldsymbol{w}$. Now, as the end points of the support space $\boldsymbol{v}$ approach zero ($\boldsymbol{v} \to \mathbf{0}$) the generalized maximum likelihood (logit) converges to the 'traditional' maximum likelihood model, which is equivalent (in this case), to the classical ME. This happens as the number of observations $(K) \to \infty$. However, since the whole logic here is to accommodate for the noise while estimating the parameters of a small and finite data set, this argument is just a consistency one.

2.2. Information, sensitivity, diagnostics and inference

Within the entropy approach used here, one can investigate the amount of information in the estimated coefficients. Define the normalized entropy (information) measure as

$$S(Q^*) \equiv \frac{-\sum_{i,j,m} q^*_{ijm} \log q^*_{ijm}}{(I \times J) \log M} \tag{9}$$

where this measure is between zero and one with one reflecting complete ignorance and zero reflecting perfect knowledge. If the CE is used, the divisors should be the entropy of the relevant priors $\{\boldsymbol{q}^0_{ij}\}$. This measure reflects the information in the whole system. Similar normalized measures reflecting the information in each one of the i, j distributions are easily defined.

Next, following our derivation of the generalized ME formalism, it is possible to construct the entropy-ratio test (which is an analogue to the maximum-likelihood ratio test). Let l_Ω be the constrained (by the data) ME with the optimal value of the objective function (5). Next, let l_ω be the unconstrained one where $\lambda = 0$ and therefore the maximum value of the objective function is $(I \times J)\log M + K \log L$ (e.g. optimizing with respect to no data). Hence, the entropy ratio statistic is just

$$W(\text{ME}) = 2l_\omega - 2l_\Omega = 2\big[(I \times J)\log M + K \log L\big]\big[1 - S(\boldsymbol{q}^*, \boldsymbol{w}^*)\big]. \qquad (10)$$

Under the null hypothesis of $\lambda = \mathbf{0}$, $W(\text{ME})$ converges in distribution to $\chi^2_{(K-1)}$.

Finally, a 'goodness of fit' measure that relates the normalized entropy measure to the traditional R^2 is

$$\text{pseudo} - R^2 \equiv 1 - \frac{l_\Omega}{l_\omega} = 1 - S(\boldsymbol{q}^*). \qquad (11)$$

These measures are easily adjusted for the CE case where $(I \times J)\log M$ is substituted for the total entropy of the priors.

Finally, we can derive the covariance matrix, which allows us to investigate the sensitivity of the estimates. There are two ways to derive the variances in this approach. We start with the traditional way and then discuss a different way. Under the more traditional way, the variance-covariance matrix $V(\lambda^*)$ of the real unknown parameters λ, given the data, is the inverse of the Hessian of the dual formulation (8), evaluated at λ^* (and at the mean values of E^*_{ij}). Therefore, the covariance elements of $\boldsymbol{q}_{ij}$ are

$$\text{Cov}(\boldsymbol{q}^*) = \left(\frac{\partial \boldsymbol{q}^*}{\partial \lambda^*}\right)' V(\lambda^*) \left(\frac{\partial \boldsymbol{q}^*}{\partial \lambda^*}\right).$$

Similarly,

$$\text{Cov}(E^*) = \left(\frac{\partial E^*}{\partial \lambda^*}\right)' V(\lambda^*) \left(\frac{\partial E^*}{\partial \lambda^*}\right)$$

where $E^*_{ij} = \sum_m q^*_{ijm} z_m$.

As a closing note to this derivation, it is emphasized that the covariance elements, and especially the variances of interest, are smaller than those of the ML or other regularization approaches. This is due to the more 'relaxed' constraints of the approach taken here. These relaxed constraints imply the relevant multipliers (in the primal or dual approach) are more stable. This is also seen in the additional noise terms in (2).

A different and a more natural way of constructing the variances is $\text{Var}(E^*_{ij}) = \sum_m q^*_{ijm} z_m^2 - \left(\sum_m q^*_{ijm} z_m\right)^2$. In that case, the variance of each point estimate E^*_{ij} can be interpreted as the estimated variance of the underlying estimated probability within the support $\boldsymbol{z}$. This is the largest possible variance of each point estimate, given the support space and the data. But, as this variance is unique for this approach, in order to compare estimates with the more traditional (Bayes or non-Bayes) methods we also developed the above variance-covariance matrix. Note that this variance is sensitive to the number of points (M) in the support space. For example, for uniform priors, as $M \to \infty$ the variance converge to the variance of the continuous uniform distribution.

Finally, using the dual approach, the objective here is to recover the K unknowns, which in turn yield the $J \times I$ matrix E. However, it may be of our interest to also evaluate the impact of the elements of the known P on each E_{ij} (or q_{ijm}). That is, many times the more interesting parameters are these marginal effects. These marginal effects are just

$$\frac{\partial q_{ijm}}{\partial p_{ij}^{k}} = q_{ijm}\left(\lambda_k z_m - H_{ijk}\right) \tag{12a}$$

$$\frac{\partial E_{ij}}{\partial p_{ij}^{k}} = \sum_m \frac{\partial q_{ijm}}{\partial p_{ij}^{k}} z_m = \sum_m q_{ijm}\left(\lambda_k z_m - H_{ijk}\right) z_m \tag{12b}$$

where $H_{ijk} \equiv \sum_m q_{ijm}\lambda_k z_m$. These effects can be evaluated at the mean values of the sample, or individually for each observation and may be used to evaluate the optimal experimental design. For the tomography problem, however, the more interesting values may be $\partial E_{ij}/\partial s_k$ or the simpler quantity $\Delta E_{ij}^2 = \sum_k (\partial E_{ij}/\partial s_k)^2$ which reflects the variances of the reconstructed image.

3. Experiment and results

3.1. Experiment and estimation results

The formalism developed in section 2 of this paper has been applied to the reconstruction of the soft x-ray emissivity of a hot fusion plasma. The experimental set-up employed to collect the data is shown schematically in figure 1. Soft x-rays emitted by the plasma are detected by two pin hole cameras each equipped with an array of 30 surface barrier detectors. Each detector signal, s_k, recorded by diode k depends linearly on the unknown local emissivity E_{ij} defined on the square mesh, shown in figure 1, with $i, j = 1, \ldots, 30$. The physically meaningful support for the reconstruction is, however, not the whole grid. It is restricted rather to the area within the grid where viewing chords from the two cameras cross. Fortunately, this area is

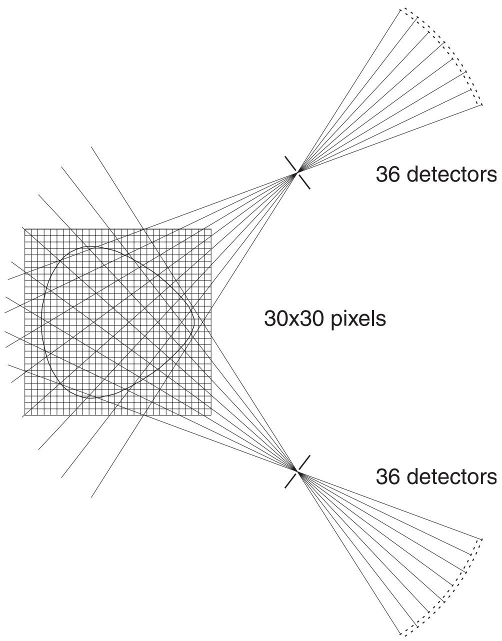

Figure 1. A sketch of the soft x-ray set-up and pixel arrangement used for the emissivity reconstruction.

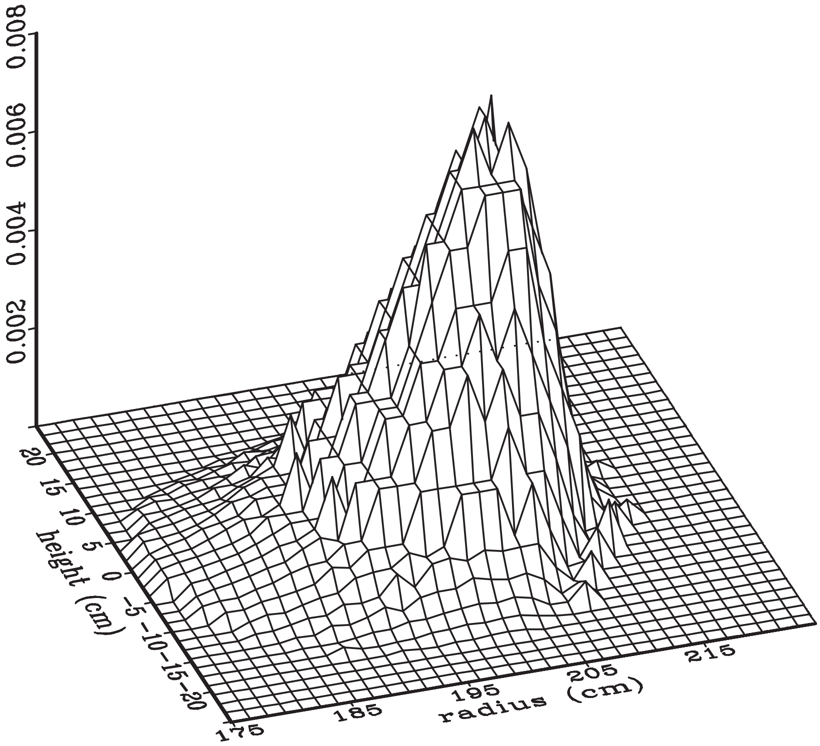

Figure 2. A reconstruction of the emissivity profiles using the generalized inversion procedure developed here.

larger than the plasma cross section as obtained from equilibrium calculations and shown in figure 1 as the closed triangularly shaped curve. Inside the physically meaningful region we choose a flat default model for E_{ij} ($E^0_{ij} = 10^{-2}$) and outside this region we take $E^0_{ij} = 10^{-5}$, a choice which serves to push the emissivity towards zero in the region which does not contribute information to the data. The pixel matrix P^k_{ij} used in this work is identical to that used earlier (Ertl *et al* 1996).

Keeping our objective of minimal distributional assumptions in mind, the support space for each error term is chosen to be uniformly symmetric around zero. For example, for $L = 3$, $\boldsymbol{v}' = (-\alpha, 0, \alpha)$ or for $L = 5, \boldsymbol{v}' = (-\alpha, -\alpha/2, 0, \alpha/2, \alpha)$. If we knew the correct variance, a reasonable rule for a is the three-standard deviation rule (Pukelsheim 1994). However, given our incomplete knowledge of the correct value of the scale parameter σ, we use the sample variance, given to us by the experimentalists, as an estimate for α. Specifically, since the experimental input data were specified with an error of 5%, we used $\boldsymbol{v}'_k = (-0.15s_k, 0, 0.15s_k) = (-3\sigma_k, 0, 3\sigma_k) = (-\alpha_k, 0, \alpha_k)$ as the bounds for the support of ε. For a detailed set of experiments investigating the small sample behaviour of the three-sigma rule see Golan *et al* (1997).

Finally, the specification of the support for E_{ij} requires the choice of a scale for the emissivities. Luckily, this choice is usually known to the researcher from the physics of the problem. In the present case we draw on existing results and *a priori* knowledge and chose $\boldsymbol{z}' = (0, 0.01, 0.02)$ which extends sufficiently beyond the maximum emissivity. It is important to note that the estimated values are not sensitive to this choice as long as the pre-determined

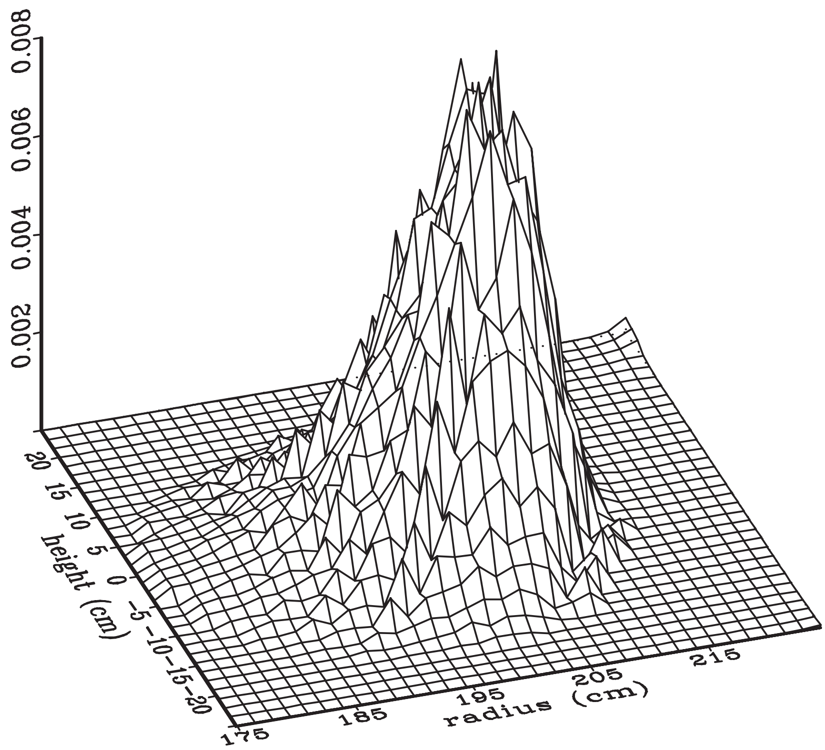

Figure 3. A reconstruction of the emissivity profiles using the quantified maximum-entropy method.

Table 1. The different statistics for the two methods compared.

	GME (figure 2)	Quantified ME (figure 3)	GME CASE B
Normalized entropy	0.171	0.157	0.123
Pseudo-R^2	0.829	0.843	0.877
Ent-ratio statistic	152.8	2140.0	121.4
Approximate computing time	1 s on a PC, pentium III	20 s on an IBM 370 workstation	1 s on a PC, pentium III

choice is consistent with the physics of the problem. For example, if instead of above support, we would specify $10z$, our results will remain practically unchanged.

Figures 2 and 3 show the reconstructed soft x-ray emissivity. The data in figure 2 result from estimation using the method described in this paper. In that case, the entropy-ratio statistic is 152.8, which is significant at the 1% level. Figure 3 was obtained using quantified maximum entropy (Skilling 1989, Ertl *et al* 1996). Table 1 summarizes the basic statistics of these estimates. To make these statistics comparable, the estimates of the quantified entropy method were transformed to be fully consistent with those of the GME.

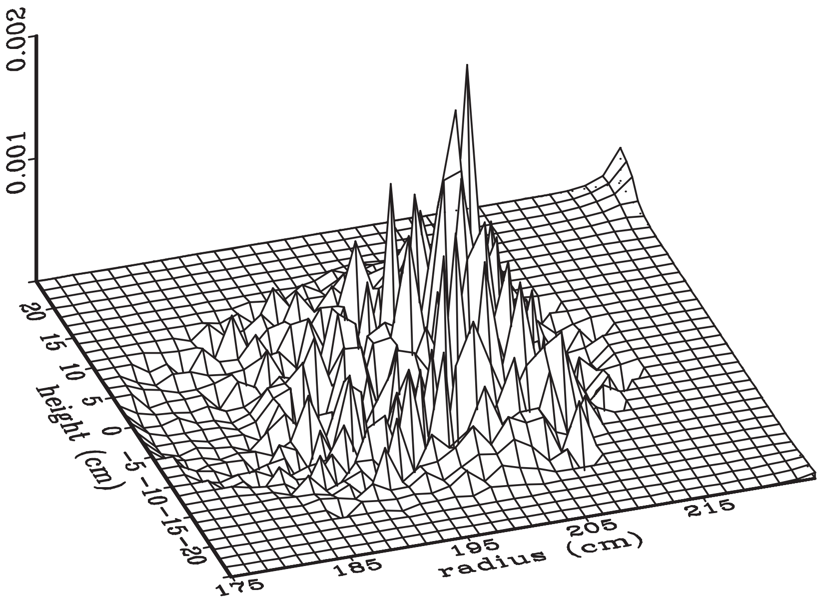

Figure 4. The difference between the estimates of the quantified entropy approach (figure 3) and the GME (figure 2).

3.2. Discussion

As is expected, at a first sight, the overall shape of the two results seems to be very much the same. But investigation of these two figures reveals that the main difference between the two profiles is the significantly more spiky surface (even outside the physically meaningful region) of the traditional quantified maximum-entropy reconstruction. This is not unexpected because, in the present treatment (figure 2), misfits between the data and the model (2) are absorbed in an individual and self-consistent way through the errors $\varepsilon_k \equiv \sum_l w_{kl} v_l$. In the quantified maximum-entropy treatment, on the other hand, the errors are fixed and misfits are reflected in a 'rough' reconstruction. Furthermore, in the quantified ME approach a quadratic loss function, that pushes the errors to zero in a faster rate, is employed, which is also reflected in the 'rough' reconstruction.

To demonstrate the qualitative differences of these two methods of estimation, figure 4 presents 'figure 3 minus figure 2'. The qualitative difference of these two estimation methods and the relative smoothness of the new method are quite obvious. Furthermore, this difference in smoothness can be quantified. A convenient measure of the curvature and hence the smoothness of a function $E_{(x,y)}$ of the two variables (x, y) is the global second derivative squared:

$$\phi = \int \left\{ \left(\frac{\partial^2 E}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 E}{\partial x \, \partial y} \right)^2 + \left(\frac{\partial^2 E}{\partial y^2} \right)^2 \right\} \mathrm{d}x \, \mathrm{d}y.$$

Note that the lower bound of ϕ is zero, which applies for a constant or a linear function in x and y. We approximate ϕ by substituting finite second differences for the partial derivatives (Abramowitz and Stegun 1965). In the analysis done here ϕ is smaller by a factor of 4×10^7 for the GME reconstruction as compared with the quantified MaxEnt reconstruction. As we expect plasma reconstruction to yield a smooth reconstruction, the results of our GME estimates seem to be much more convincing.

An additional advantage of the present method on the previous treatment lies in the fact that a single pass calculation provides the final results. In contrast, in the quantified ME method, in order to find the most probable value of the hyperparameter controlling the amount of regularization, repeated calculations are necessary for different values of this parameter. Since about 10–20 iterations have to be performed to invert the present data, the gain in computational speed is of the same order (e.g. in the present case, 20 iteration were needed). This is of course very welcome for visualizing plasma dynamics since the experimental system is capable of sampling at 4 μs intervals. See the details in the bottom row of table 1.

Finally, another advantage of this approach is its robustness to the prior and model specification. For example, instead of specifying the priors outside the physically meaningful regime to be very small (figure 2), we specify the model such that the support space for each pixel outside this regime is zero (call it GME, case B). The specification inside this regime remains as before. That is, inside the regime $\boldsymbol{z}' = (0, 0.01, 0.02)$ and outside $\boldsymbol{z} = \mathbf{0}$. Regardless of this different specification, the reconstructed image (not presented here) is practically equivalent to figure 2, indicating the robust nature of this procedure. The statistics for this case are represented in column 4 of table 1 and are quite similar to those of column 2 (figure 2). Note, that as is expected, the goodness of fit measure is slightly superior to this of figure 2.

Finally, a short discussion of the priors and supports is in place. For the tomography problem discussed here, the specification of the support spaces is quite natural. First, the support $\boldsymbol{v}$ for the noise was discussed earlier and is determined directly from the data or the experimentalists. The support space $\boldsymbol{z}$ must be positive. As long as this support is specified such that it spans the correct true values of the image, the solution is quite robust to this specification. Obviously, if this support is specified incorrectly the reconstructed image will be affected. But luckily, we know here that the support should be positive with a lower bound at zero so miss-specification seems illogical.

4. Concluding remarks

In this paper we develop and apply a generalized, informational-based, inversion procedure for tomographic reconstruction. This new approach is easy to apply and compute, uses less *a priori* assumptions on the underlying distribution that generated the data, accommodates for the noise in the data and yields good estimates. We compared our approach with other regularization methods and apply it to the reconstruction of the soft x-ray emissivity of a hot fusion plasma. Furthermore, the classical ME is just a special case of this new and generalized method.

This generalized inversion procedure can also be used to reconstruct the images of a large number of other noisy problems such as the Markov (or matrix reconstruction) problem.

Acknowledgments

We thank the referees for their valuable comments. Part of this research was done while AG was visiting the Max-Planck-Institut für Plasmaphysik (June–July 1999). AG would like to thank the Max-Planck-Institut for the support and hospitality.

Appendix. A simulated (Monte Carlo) example

To illustrate, evaluate and provide a flavour for the performance of the generalized inversion procedure discussed here a simple example is presented and contrasted with other procedures.

To keep these simulations as general as possible, instead of investigating only the case of a 34×34 matrix, we consider the four cases $I, J = 5, 10, 50, 75$. The correct values for the unknown $(I \times J)$ E matrix are chosen arbitrarily from a multi-normal distribution with four different means and variances. In this way we have four picks with different spread each. For each I (and J) and a noise level of about 5%, 100 samples were generated and estimated with three different methods. The results are summarized in terms of the mean squared errors (MSE) and the total variance (TV) where

$$\mathrm{MSE} \equiv \frac{1}{100} \sum_{\mathrm{samples}=1}^{100} \left\{ \sum_{i,j} (E_{ij}^* - E_{ij})^2 \right\} = \text{total variance} + \text{bias}^2.$$

The results are presented in table A1, where E^* is the matrix of estimated values and E is the matrix of correct values. Each row of table A1 presents estimates resulting from a different inversion procedure. The different procedures used are the classical ME (Jaynes 1957, Levine 1980, Skilling 1989), the RAS method of biproportional adjustments often used for matrix reconstruction (Bacharach 1970, Schneider and Zenios 1990) and the GME method discussed in section 2. In each case and method the correct priors were used. In all cases and dimensions of these simulated examples the GME is significantly more accurate and has a significantly lower variability as compared with the other methods. Finally, the same set of simulations were repeated for the incorrect uniform priors and for different noise levels. Since the results are qualitatively similar to the results shown in the table, they are not presented here.

Table A1. Summary of simulated sampling results.

	Estimation	MSE	TV
$I = J = 5$	ME	1.165	0.056
	RAS	2.391	0.135
	GME	0.265	0.009
$I = J = 10$	ME	0.976	0.025
	RAS	2.385	0.044
	GME	0.246	0.004
$I = J = 50$	ME	0.663	0.004
	RAS	1.396	0.005
	GME	0.172	0.0007
$I = J = 75$	ME	0.671	0.003
	RAS	1.413	0.003
	GME	0.171	0.0004

References

Abramowitz M and Stegun I A 1965 *Handbook of Mathematical Functions* (New York: Dover)
Bacharach M 1970 *Biproportional Matrices and Input–Output Change* (New York: Cambridge University Press)
Bercher J-F, Besnerais G L and Demoment G 1996 *Maximum Entropy and Bayesian Studies* ed J Skilling and S Sibisi (Dordrecht: Kluwer)
Csiszar I 1991 *Ann. Stat.* **19** 2039
——1996 *Maximum Entropy and Bayesian Studies* ed K M Hanson and R N Silver (Dordrecht: Kluwer)

Donoho D, Johnstone I, Hoch J and Stern A 1992 *J. R. Stat. Soc.* B **54** 41
Ertl K, von der Linden, Dose V and Weller A 1996 *Nucl. Fusion* **36** 11
Golan A 1998 *Maximum Entropy and Bayesian Studies* ed G Erickson (Dordrecht: Kluwer)
Golan A, Judge G and Miller D 1996 *Maximum Entropy Econometrics: Robust Estimation with Limited Data* (New York: Wiley)
Golan A, Judge G and Perloff J 1997 *J. Econometrics* **79** 23–51
Gzyl H 1998 *Maximum Entropy and Bayesian Studies* ed G Erickson (Dordrecht: Kluwer)
Hanson K M and Silver R N 1996 *Maximum Entropy and Bayesian Methods* (Dordrecht: Kluwer)
Jaynes E T 1957 *Phys. Rev.* **106** 620
——1963 *Statistical Physics* (New York: Benjamin) p 181
Kullback J 1959 *Information Theory and Statistics* (New York: Wiley)
Levine R D 1980 *J. Phys. A: Math. Gen.* **13** 91
Levine R D and Tribus M 1979 *The Maximum Entropy Formalism* (Cambridge, MA: MIT Press)
Navaza J 1986 *Acta Crystallogr.* 212
Pukelsheim F 1994 *Am. Stat.* **48** 88–91
Schneider M H and Zenios S A 1990 *Oper. Res.* **38** 439
Skilling J 1989 *Maximum Entropy and Bayesian Methods in Science and Engineering* ed J Skilling (Dordrecht: Kluwer)
——1989 *Maximum Entropy and Bayesian Methods in Science and Engineering* (Dordrecht: Kluwer)
Shore J E and Johnson R W 1980 *IEEE Trans. Inf. Theory* **26** 26
Tikochinsky Y, Tishby N Z and Levine R D 1984 *Phys. Rev. Lett.* **52** 1357