Tomographic Reconstruction from Noisy Data

Amos Golan* and Volker Dose†

*Department of Economics, American University, 4400 Massachusetts Ave., NW Washington, DC 20016-8029, USA
†Center for Interdisciplinary Plasma Science Max-Planck-Institut für Plasmaphysik, EURATOM Association, D-85748, Garching b. München, Germany

Abstract. A generalized maximum entropy based approach to noisy inverse problems such as the Abel problem, tomography, or deconvolution is discussed and reviewed. Unlike the more traditional regularization approach, in the method discussed here, each unknown parameter (signal and noise) 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. We use this method for tomographic reconstruction of the soft x-ray emissivity of hot fusion plasma.

INTRODUCTION

The objective of this short paper is to summarize, explore and review a generalized inversion procedure for the tomography problem. This generalized inversion procedure, call it Generalized Maximum Entropy (GME), extends the classical Maximum Entropy (ME) formalism of Jaynes [1,2] and the information theoretical approach of Levine [3] by directly adjusting for the noise in the observed data. Therefore, this approach provides a more conservative inference from noisy data, does not employ the traditional regularization parameter, and is free of distributional assumptions.

In the next section we start by formulating the tomography problem, describe the detailed estimation method, and then provide the necessary diagnostics and inferential statistics. In Section 3 we discuss the continuous limit of this method. In Section 4, the experiment and results are presented. Some concluding remarks are then discussed.

THE BASIC TOMOGRAPHY PROBLEM

Consider the following discretized tomography model

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

where **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 $E_{ij} \geq 0$, and ε is a vector of independently and identically distributed noise.

Our objective is to recover E from the aggregated noisy data $\boldsymbol{s}$. If 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 (ME) formalism provides such an inversion procedure where one maximizes the entropy of E subject to the available information. See for example [4] for applying the ME method to analyzing the tomography problem.

However, for noisy data, if we want to avoid distributional assumptions, then regardless of the number of observations (data points) the problem is always ill-posed or underdetermined. The GME method we now describe, conforms to the above requirement of no a-priori specification of the underlying distribution/s.

As a first step, both E_{ij} and ε_k*'s* need to be transformed into proper probabilities. This transformation is necessary in order to build an estimation method within the philosophy of the classical ME [1,2,3] and information theory. Following [5], the model is reparametrized 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 $\mathbf{q}_{ij}$ is an M-dimensional vector of $(I \times J)$ proper probability distributions satisfying

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

and each one of the point estimates E_{ij} is now defined as

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

For now we define the vector $\mathbf{z}$ as an M-dimensional discrete support space with $M \geq 2$ specified such that $\mathbf{z}' = (z_1, ..., z_M)$ where *'$\prime$'* stands for "transpose". The continuous support case is discussed in Section 3. If for example due to physical arguments $E_{ij} \geq 0$, then all z_m must be non-negative. In a similar way, $\mathbf{w}_k$ is viewed as an L-dimensional vector of K proper probability distributions satisfying

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

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

With the above specification, in the GME framework, one maximizes the joint entropies of the signal, $\{\mathbf{q}_{ij}\}$, and the noise, $\{\mathbf{w}_k\}$, subject to the available data (Equation 2) and the requirements of proper distributions. Since one usually have some prior information in such problems, we formulate the estimation problem in terms of the Cross Entropy (CE) formalism. Let the prior information for the set $\{\mathbf{q}_{ij}\}$, defined on the support $\boldsymbol{z}$**,** be $\{\mathbf{q}_{ij}^0\}$. Similarly, within the support space $\boldsymbol{v}$, the priors for the noise components are $\mathbf{w}_k^0$, and are always taken to be *uniform* over the symmetric support $\boldsymbol{v}$. This generalized cross entropy (GCE), or generalized inversion procedure, is

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

subject to

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

$$\sum_m q_{ijm} = 1 \quad \sum_l w_{kl} = 1$$

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

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

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

Building on the Lagrangian, the dual, concentrated, unconstrained problem (as a function of λ) is

$$L(\lambda) = \sum_k s_k \lambda_k - \sum_{i,j} \log \Omega_{ij}(\lambda) - \sum_k \log \Psi_k(\lambda) \ . \tag{8}$$

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

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. In other words, regardless of the number of points in the support spaces (including the continuous case) there are K Lagrange multipliers and both $\{\mathbf{q}_{ij}\}$ and $\{\mathbf{w}_k\}$ 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. 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. These weights can be changed if one wishes to impose higher relative weights on the precision or the prediction parts respectively. Equivalently, this method can be viewed as a shrinkage inversion procedure that simultaneously shrinks the data to the priors and the errors to zero. For a detailed discussion and comparison with other methods see [6].

ENTROPY, INFORMATION AND DIAGNOSTICS

Within the GME-GCE approach used here, the amount of information in the estimated coefficients is captured via the normalized entropy:

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

This measure, which reflects the information in the whole system, is between zero and one with one reflecting complete ignorance and zero reflecting perfect knowledge. If the GCE is used, the divisors should be the entropy of the relevant priors $\{\mathbf{q}^0_{ij}\}$. Finally, a similar normalized measures reflecting the information in each one of the $i, j's$ distributions are easily defined.

For inferential purposes it would be of interest to relate the estimates with some well defined distribution. To do so, we use the Entropy-Ratio (ER) test (which is an analogue to the maximum likelihood ratio test). Let l_Ω be the constrained (by the data: Equation 2) GME (uniform priors case) with the optimal value of the objective function (5). Next, let l_ω be the value of the unconstrained objective function (5) where $\lambda = 0$ and therefore the maximum value of (5) is just $(I \times J)\log M + K \log L$ (e.g., optimizing with respect to no data). Hence, the ER statistic is just

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

Under the null hypothesis of λ=***0***, W(GME) converges in distribution to $\chi^2_{(K-1)}$ which enables us to perform any type of single or composite hypothesis test.

A "goodness of fit", or in-sample prediction, measure that relates the normalized entropy measure $S(Q)$ to the traditional R^2 is just

$$Pseudo - R^2 \equiv 1 - \frac{l_\Omega}{l_\omega} = 1 - S(\mathbf{q}^*). \tag{11}$$

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, in many circumstances, it may be of 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

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

$$\frac{\partial E_{ij}}{\partial p^k_{ij}} = \sum_m \frac{\partial q_{ijm}}{\partial p^k_{ij}} z_m = \sum_m q_{ijm}\left(\lambda_k z_m - H_{ijk}\right) z_m$$

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 $\Delta E^2_{ij} = \sum_k \left(\partial E_{ij}/\partial s_k\right)^2$ which reflects the variances of the reconstructed image. For further discussion and derivation of the covariance matrix see [6].

EXTENSIONS TO THE CONTINUOUS SUPPORTS

Recalling that $\boldsymbol{v}$ and $\boldsymbol{z}$ are closed, convex sets, we now investigate the behavior of the GME/GCE estimation rule at the limit of $M \to \infty$ and $L \to \infty$, while holding everything else unchanged. These limits are related to the notion of *super-resolution* and can also be viewed as some measures of *sufficient statistics*. For simplicity of exposition, we discuss here the case of $L \to \infty$ which relates to the errors' support $\boldsymbol{v}$. We demonstrate here two cases: uniform and Normal distributions.

Within our previous definition, assume $\underline{v}_k \leq \varepsilon_k \leq \overline{v}_k$ where $\underline{v}_k$ and $\overline{v}_k$ are the lower and upper bounds of $\boldsymbol{v}$ respectively, and let $V = \prod_{k=1}^{K} (\underline{v}_k, \overline{v}_k)$ be the joint space of all the K supports . Note that in many cases the choice $\underline{v}_k = \underline{v}$ and $\overline{v}_k = \overline{v}$ for all $k=1, 2, \ldots, K$ is appropriate. Assuming uniform a-priori information (within V) for each error, implies

$$dW^0(\zeta) = \prod_{k=1}^{K} \frac{d\zeta_k}{(\overline{v}_k - \underline{v}_k)} \tag{13}$$

where W^0 reflects our prior knowledge. To be consistent with the discrete version presented earlier, we choose to work with a continuous uniform distribution within the bounds $\underline{v}$ and $\overline{v}$. Building on this prior information the post-data W is $dW(\zeta) = \rho(\zeta) dW^0(\zeta)$. Maximizing the entropy (or similarly minimizing the cross-entropy)

$$I\left(W, W^0\right) = -\int_V \rho(\zeta) \log \rho(\zeta) dW^0(\zeta) \tag{14}$$

subject to the data and the normalization requirements yields

$$dW(\zeta) = \frac{e^{-\lambda'\zeta}}{\Psi(\lambda)} dW^0(\zeta) \tag{15}$$

Finally,

$$\Psi_k(\lambda^*) = \frac{e^{-\lambda_k^* \underline{v}} - e^{-\lambda_k^* \overline{v}}}{\lambda_k^* (\overline{v} - \underline{v})} \tag{16}$$

and

$$\varepsilon_k^* = \frac{1}{(\overline{v} - \underline{v})} \left\{ \frac{\underline{v} e^{-\lambda_k^* \underline{v}} - \overline{v} e^{-\lambda_k^* \overline{v}}}{\lambda_k^*} + \frac{e^{-\lambda_k^* \underline{v}} - e^{\lambda_k^* \overline{v}}}{\lambda_k^{*2}} \right\} \tag{17}$$

and the dual is just

$$L(\lambda) = \sum_k s_k \lambda_k - \sum_{i,j} \log \Omega_{ij}(\lambda) - \sum_k \log \Psi_k(\lambda) \tag{18}$$

where Ψ_k is defined by (15).

Next, we consider the normal distribution where, unlike the previous (continuous uniform) case, the noise vector ε is allowed to take any value. The prior distribution is then $W^0(d\eta^K) = \left(2\pi\sigma^2\right)^{-K/2} \exp(-\eta'\eta/2\sigma^2) d\eta^K$. If we allow the noise components to be correlated among the K degrees of freedom (observed data points), or similarly we do not exclude the possibility that the K observations are correlated via the vector ε, then,

$$\Psi(d\eta^K) = \left(\frac{\det\Sigma}{2\pi}\right)^{K/2} \exp\left[-\left(\eta'\Sigma^{-1}\eta\right)/2\right] d\eta^K \tag{19}$$

where Σ is the (known or unknown) covariance matrix.

Starting with the simpler case (no autocorrelations), the corresponding set of partition functions$\{\Psi_k\}$,Ψ, is

$$\Psi(\lambda) = \int \frac{e^{-\lambda'\eta} e^{-\eta'\eta/2\sigma^2}}{\left(2\pi\sigma^2\right)^{K/2}} d\eta^K =$$

$$\prod_{k=1}^{K} \left(\frac{1}{\left(2\pi\sigma^2\right)^{K/2}}\right) \int e^{-\lambda_k\eta_k - \eta_k^2/2\sigma^2} d\eta_k = \prod_{i=1}^{K} e^{\lambda_k^2\sigma^2/2}. \tag{20}$$

Substituting (20) into (8) yields

$$L(\lambda) = \sum_k s_k\lambda_k - \sum_{i,j} \log\Omega_{ij}(\lambda) - \sum_k \left(\sigma^2\lambda_k^2/2\right) \ . \tag{21}$$

Minimizing the dual (21) yields the desired λ^*, which, in turn, determines the estimated noise probabilities, where the mean of the post-data (or estimated) noise is just $-\sigma^2\lambda^*$. This means that, for finite samples, the posterior distribution of the noise may have a non-zero mean which in turn implies superior finite sample estimates.

The more general case where observations may be autocorrelated is a straight forward extension of the above where

$$\Psi(\lambda) = \int e^{-\lambda'\eta} \left(\frac{\det\Sigma}{2\pi}\right)^{K/2} e^{-\left(\eta'\Sigma^{-1}\eta\right)} d\eta^K = e^{(\lambda'\Sigma\lambda)/2} \tag{22}$$

Finally, substituting (22) for (20) within the dual, unconstrained model (21) yields the desired estimates that in turn yield $\varepsilon^* = -\Sigma\lambda^*$.

EXPERIMENT AND RESULTS

The method discussed earlier in this paper has been applied to the reconstruction of the soft x-ray emissivity of a hot fusion plasma. The experimental setup employed to collect the data is schematically shown in Figure 1. Soft x-rays emitted by the plasma

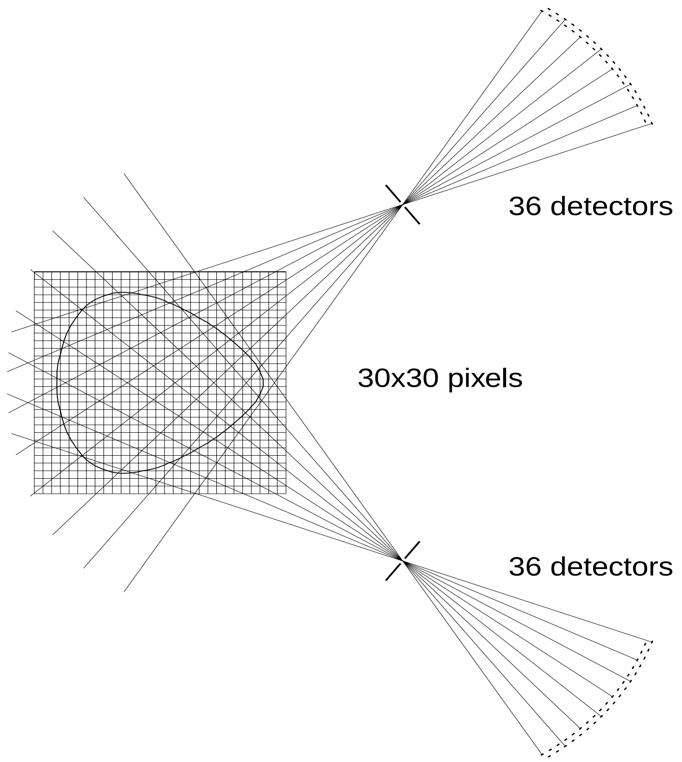

FIGURE 1. A sketch of the x-ray set up and pixel arrangement used for the emissivity reconstruction.

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 $i, j = 1, ..., 30$. 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 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 the one used earlier [4].

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 $L = 3$, $\mathbf{v}'=(-\alpha, 0, \alpha)$. If we knew the correct variance, a reasonable rule for α is the three-standard deviation rule. 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. The experimental input data were specified with an error of 5%. Therefore, we used $\nu'_k = (-0.15 s_k, 0, 0.15 s_k) = (-3\sigma_k, 0, 0.3\sigma_k) = (-\alpha_k, 0, \alpha_k)$ as the bounds for the support of ε. For a detailed set of experiments investigating the small sample behavior of the three-sigma rule see [7].

Finally, the specification of the support for E_{ij} requires the choice of a scale for the

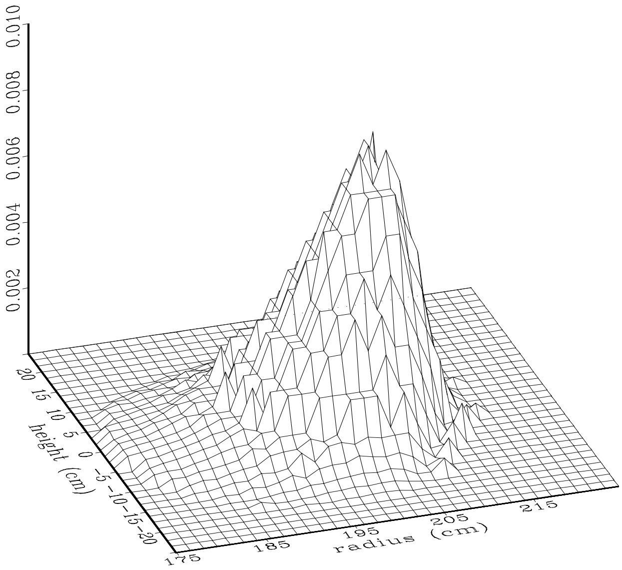

FIGURE 2. A reconstruction of the emissivity profiles using GME.

emissivities. Luckily, the researcher usually knows this choice from the physics of the problem. In the present case we draw on existing results and a-priori knowledge and chose $\mathbf{z}'=(0, 0.01, 0.02)$ 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 choice is consistent with the physics of the problem.

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 [4,8]. 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.

TABLE 1. The Different Statistics Resulting From the GME and The Quantified ME.

	GME (Fig. 2)	**Quantified ME (Fig. 3)**	**GME – CASE B (Fig. 5)**
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 second on a PC, Pentium III	20 seconds on an IBM 370 Workstation	1 second on a PC, Pentium III

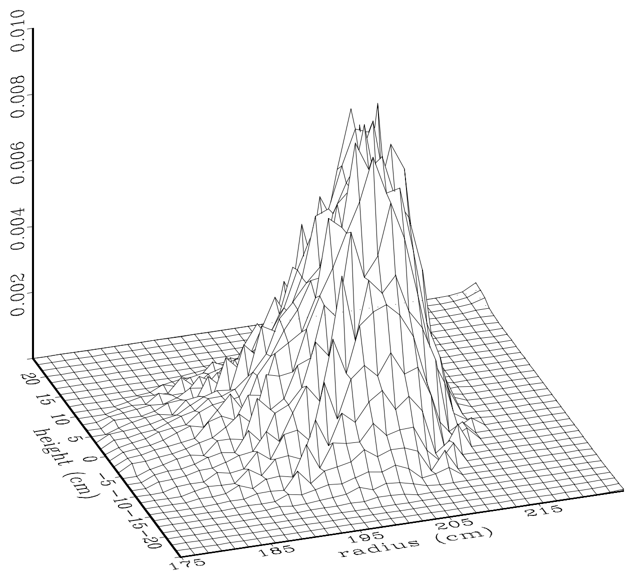

FIGURE 3. A reconstruction of the emissivity profiles using the quantified ME method.

DISCUSSION AND SUMMARY

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. Further, 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. Further, 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

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

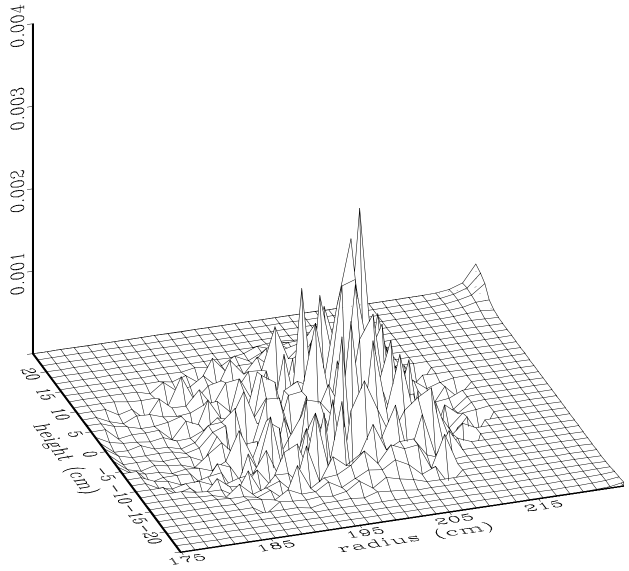

FIGURE 4. The difference between the estimates of the quantified entropy approach(Figure 3) and the GME Figure 2)

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 [9]. In the analysis done here ϕ is smaller by a factor of 4×10^7 for the GME reconstruction as compared with the quantified ME reconstruction. As we expect plasma emissivity reconstruction to yield a smooth surface, 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. By 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 ten to twenty iterations have to be performed for inverting the present data, the gain in computational speed is of the same order (e.g., in the present case, twenty iteration were needed). This is of course very welcome for visualizing plasma dynamics since the experimental system is capable of sampling at 4 microsecond intervals. See 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 (Fig. 2), we specify the model such that the support space for each pixel outside this regime is zero (see GME - Case B in Table 1 and Figure 5). The specification inside this regime remains as before. That is, inside the regime ***z'****=(0,*

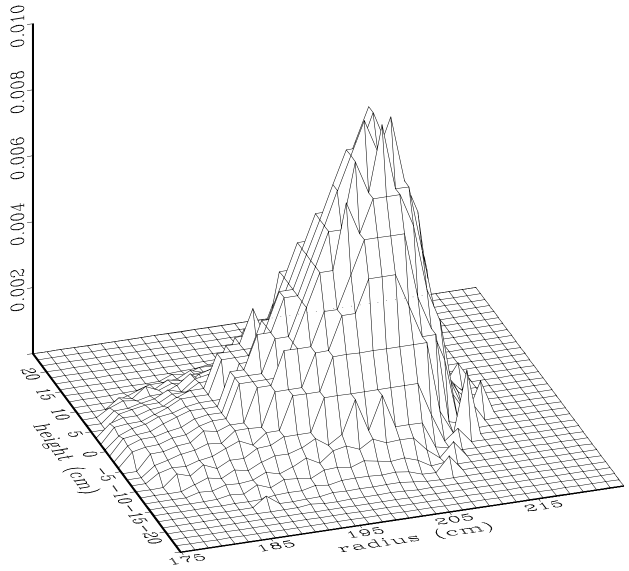

FIGURE 5. A reconstruction of the emissivity profiles using GME, but with support of zero outside the physically meaningful regime.

0.01, 0.02) and outside ***z=0***. Regardless of this different specification, the reconstructed image (Fig. 5) is practically equivalent to Figure 2. This similarity reveals the robust nature of this procedure. The statistics for this case (see Table 1) are very similar to those of column 2. But, as is expected, the goodness of fit measure (in-sample prediction) is slightly superior to this of Figure 2.

To summarize, in this paper we reviewed, extended and explored a generalized, informational based, inversion procedure for tomographic reconstruction. This approach is (i) easy to apply and compute, (ii) uses less a-priori assumptions on the underlying distribution that generated the data, (iii) accommodates for the noise in the data and thus yields good estimates, and (iv) contains the classical ME as special case.

The experimental results, of reconstructing the soft x-ray emissivity of hot fusion plasma, provided here demonstrated the advantages discussed above.

REFERENCES

1. Jaynes, E. T., *Phys. Rev*. **106**, 620-630 (1957).
2. Jaynes, E. T., *Phys. Rev*. **108**, 171-190 (1957).
3. Levine, R. D., *J. Phys. A*, **13** 91-108 (1980).
4. Ertl K, von der Linden, W., Dose, V., and Weller, A., *Nuclear Fusion **36***,11, 1477-1488 (1996).

5. Golan, A., Judge, G., and Miller, D., *Maximum Entropy Econometrics: Robust Estimation With Limited Data*, John Wiley & Sons, New York, 1996.
6. Golan, A., and Dose, V., *J. Physics A **34**, 1271-1283* (2001).
7. Golan, A., Judge, G., and Perloff. J., *J. Econometrics **79***, $23-51$(1997).
8. Skilling. J., "Classic Maximum Entropy," in *Maximum Entropy and Bayesian Methods, edited by j. Skilling*, Kluwer Academic Pub., Dordrecht, 1989, pp. 45-52.
9. Abramowitz, M., and Irene, A., *Handbook of Mathematical Functions*, Dover Pub. Inc., New York, 1965.