MAXIMUM ENTROPY BEAM DIAGNOSTIC TOMOGRAPHY*

C. T. Mottershead, AT-6, MS H829†
Los Alamos National Laboratory, Los Alamos, NM 87545

Summary

This paper reviews the formalism of maximum entropy beam diagnostic tomography as applied to the Fusion Materials Irradiation Test (FMIT) prototype accelerator. The same formalism has also been used with streak camera data to produce an ultrahigh speed movie of the beam profile of the Experimental Test Accelerator (ETA) at Livermore.

Introduction

Intense particle beams require noninterceptive diagnostics. One of these is the light emitted from interaction of the beam with residual gas. If the light is produced by a first-order process linear in the beam density, its profile measured across the beam may be interpreted as a tomographic projection of that density distribution.[1-4] With a small number of such projections, and appropriate transfer matrices connecting them, Minerbo's[5,6] maximum entropy algorithm may be used to construct an estimate of the beam density distribution in both coordinate and phase space. The objective of this paper is to provide a concise review of this formalism and some of its applications to accelerator diagnostics.

Formulation of Problem

The problem is defined in a Cartesian coordinate system (x, y, z) with the z axis in the beam direction. The quantities being sought are the two-dimensional functions f(x,y), representing the density of the beam in the plane z = 0, or the phase-space density distribution f(x,x') where x' = dx/dz is the slope of the trajectory and is proportional to the transverse momentum.

The observed data is a set of N different projection integrals, defined by

$$p_n(s) = \int_{-\infty}^{\infty} f(x_n(s,t), y_n(s,t))\, dt \quad , \qquad n = 1, 2, \ldots N \quad . \tag{1}$$

Each projection is specified by a different pair of coordinate transformation functions

$$x = x_n(s,t) \quad \text{and} \quad y = y_n(s,t) \tag{2}$$

that give the mapping between the (x,y) plane and the (s,t) plane with s the projection sample coordinate, and t the transverse integration coordinate. Each projection integral provides a different "view" of the same (unknown) function f(x,y). This concept is easily visualized in the case of spacial reconstruction, where the views are taken at different angles about the beam axis (see Fig. 1). The nth pair of transformed coordinates are then specified by a simple rotation matrix

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{bmatrix} \cos\theta_n & \sin\theta_n \\ -\sin\theta_n & \cos\theta_n \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad , \tag{3}$$

from which we obtain

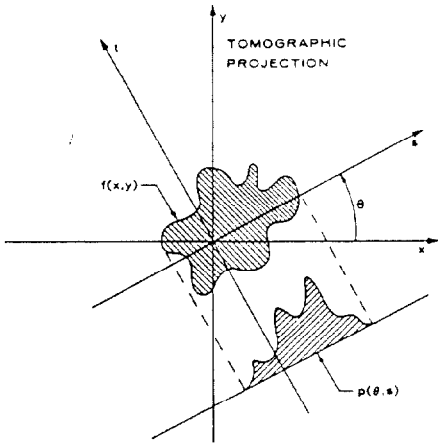

Fig. 1. The geometry of tomographic projection for spacial reconstruction (+ signs).

$$x_n(s,t) = s\cos\theta_n - t\sin\theta_n \quad \text{and} \quad y_n(s,t) = s\sin\theta_n + t\cos\theta_n \quad . \tag{4}$$

For the emittance (phase-space) reconstruction, observations are taken from a series of stations at different z-coordinates down the beamline. The projection sample coordinate at the nth station is just s = x, the spacial coordinate across the beam. The coordinate being integrated over is the trajectory slope t = x'.

We assume these (s,t) coordinates for the nth station are related to the (x,x') coordinates in the reconstruction plane at z = 0, by an arbitrary linear transport matrix A_n:

$$\begin{pmatrix} s \\ t \end{pmatrix} = \left[A_n\right]\begin{pmatrix} x \\ x' \end{pmatrix} = \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix}\begin{pmatrix} x \\ x' \end{pmatrix} . \tag{5}$$

The Jacobian of the transformation is $J_n = \det|A_n| = a_nd_n - b_nc_n$. Because J_n is the area of the (s,t) plane corresponding to a unit area in the (x,x') plane, it specifies the emittance change between z = 0 and station n ($z = z_n$). From the inverse matrix, we obtain in general

$$x = x_n(s,t) = (d_ns - b_nt)/J_n \quad \text{and} \quad x' = y_n(s,t) = (a_nt - c_ns)/J_n \quad . \tag{6}$$

Note that the s axis maps into a line through the origin with slope $-c_n/d_n$. The integration direction for the nth view is parallel to the image of the t-axis, namely, a family of lines with slope $-a_n/b_n$. The simplest example is a linear drift for which

$$\left[A_M\right] = \begin{bmatrix} 1 & z_n \\ 0 & 1 \end{bmatrix} \quad . \tag{7}$$

In this case, the s-axis maps onto the x-axis, while the t-axis projection integration lines have a slope $-1/z_n$. Note again that we are treating each observed projection as a view in a different coordinate system of the same, fixed, phase-space distribution function f(x,x'). But in the (x,x') plane, the projection and sample axes for the other views are in general not orthogonal. This interpretation is in contrast to the more usual one of the distribution function executing a kind of generalized rotation and distortion, in local (x,x') coordinates, as the particles move down the beamline.

*Work supported by the US Department of Energy.
†Most of this work was performed while the author was at EG&G, Los Alamos.

0018-9499/85/1000-1970$01.00© 1985 IEEE

The Maximum Entropy Principle

The FMIT prototype accelerator at Los Alamos is equipped with eight data-collecting stations capable of measuring such projections (three using mirrors and TV cameras, the rest using fiber optics and Reticon linear arrays. Given a set of projection data of the type described by Eq. (1), we would like to invert it to find f(x,y) or f(x,x'). If, in the rotation case, we knew the projection as a continuous function of angle, the Radon transform would give a unique inverse. In medical tomography, hundreds of data views are taken to approximate this result. For accelerator applications, however, we can have only a few views, usually 3 or 4. In this case the inversion is not unique, and we need a mechanism for constructing an estimate of f(x,y) that incorporates everything we know, and nothing else. The maximum entropy principle offers a natural way for doing this.[7-9] It argues that, of all the possible distribution functions f(x,y) that satisfy the observed constraints [Eq. (1)], the most reasonable one to choose is that one that nature can produce in the greatest number of ways, namely, the distribution having maximum entropy.

The entropy of a distribution f(x,y) may be defined as

$$H(f) = -\iint dxdy\, f(x,y)\, \ln f(x,y) \quad . \tag{8}$$

This formula is the unique measure of the multiplicity of the microstates consistent with a given distribution. It may be derived from the functional equations expressing the logic of combining probabilities, or from a direct counting of the number of independent ways of distributing N particles over M cells of phase space. This number is given by

$$W = \frac{N!}{n_1!\, n_2! \cdots n_m!} \quad , \tag{9}$$

where n_i is the number of particles in the ith cell. If we set $n_i = Nf_i$, so that f_i = the fraction of particles in the ith cell, and use Stirling's approximation, neglecting terms of order 1/N, we get

$$1/N \ln W = -\sum_{i=1}^{M} f_i \ln f_i \quad , \tag{10}$$

which is the discrete analog of Eq. (8).

Equation (10) implies that an N particle distribution with entropy H corresponds to a multiplicity W = exp NH. This means that, if the a priori probability of a given distribution is proportional to the number of ways it can be produced (that is, its multiplicity), nature strongly favors the maximum entropy distribution, in the sense that a unit increase in entropy corresponds to a factor of e^N in multiplicity.

Therefore, of all the distributions that satisfy our data constraints [Eq. (1)], we want the one having maximum entropy, as defined by Eq. (8). Its construction is a straightforward variational calculus problem outlined in the next section. Its nature may be surmised by noting that the unconstrained maximization of entropy always leads to the uniform distribution $f_i = 1/M$. For the constrained problem then, the distributions of lesser entropy are presumably not as smooth as the favored one, and contain various oscillatory terms that integrate to zero in the directions of observation, leaving no evidence in the projection data.

The Maximum Entropy Solution

The method of Lagrange multipliers is used to maximize the entropy of Eq. (8) subject to the constraints of Eq. (1). Form the Lagrangian functional

$$\psi(f,\lambda) = H(f) + \sum_{n=1}^{N} \int ds\, \lambda_n(s) \left[\int dt\, f(x_n, y_n) - p_n(s)\right] \quad , \tag{11}$$

and demand that it be stationary with respect to variations in both the unknown two-dimensional function f(x,y) and the N unknown one-dimensional Lagrange multiplier functions $\lambda_n(s)$. By construction, the condition $\delta\psi = 0$ under the variation $\lambda_n(s) \to \lambda_n(s) + \delta\lambda_n(s)$, for arbitrary $\delta\lambda_n(s)$, only reproduces the constraint Eqs. (1). What remains is to demand $\delta\psi = 0$ under the variation $f(x,y) \to f(x,y) + \delta f(x,y)$, for arbitrary $\delta f(x,y)$. The variation in the entropy term is immediately

$$\delta H = -\iint dxdy\, [1 + \ln f(x,y)]\delta f(x,y) \quad . \tag{12}$$

To collect the coefficients of $\delta f(x,y)$ from the summation term is a bit more difficult. It requires the observation that each of the double integrals on s and t is over the entire plane, and can be transformed back to the (x,y) coordinate system through the inverses of the N different mappings specified by Eq. (2). For the nth mapping, we denote by $s_n(x,y)$ the function giving the value of s corresponding to the point (x,y). In the case of simple rotations connecting the views, this is

$$s_n(x,y) = x \cos\theta_n + y \sin\theta_n \quad . \tag{13a}$$

In the emittance case, y is replaced everywhere by x', and Eq. (5) gives

$$s_n(x,x') = a_n x + b_n x'. \tag{13b}$$

After these coordinate transformations, the variation of the sum term may be written

$$\sum_{n=1}^{N} \iint \lambda_n(s_n(x,y))\, \delta f(x,y)\, J_n dxdy \quad , \tag{14}$$

where J_n is the Jacobian of the transformation, a constant for the linear transformations considered here. For the rotational case, J_n is always unity, whereas for the phase-space reconstruction, J_n gives the change in emittance (phase-space area scale) between the reconstruction plane and the nth data station. Normally $J_n = 1$ in this case also. The terms involving the data $p_n(s)$ may be ignored because they do not depend on f.

We may now collect all the coefficients of $\delta f(x,y)1$ from Eqs. (12) and (14) to find the condition that ψ be stationary under variations in f:

$$\ln f(x,y) = \sum_{n=1}^{N} J_n \lambda_n(s_n(x,y)) - 1 \quad . \tag{15}$$

Because at this point, the Lagrange multipliers $\lambda_n(s)$ are still unknown, it is convenient to replace them by equally unknown "Lagrange factors" $h_n(s) = \exp(J_n\lambda_n(s) - 1/N)$ in terms of which the maximum entropy distribution takes on a simple product form

$$f(x,y) = \prod_{n=1}^{N} h_n[s_n(x,y)] \quad . \tag{16}$$

Substitution of this form into the constraint Eq. (1), results in a set of N simultaneous nonlinear equations for the N unknown one dimensional functions $h_n(s)$. Minerbo[5] first noted the remarkable fact that, at least for linear transformations, $h_n(s)$ always factors out of its own constraint integral

$$p_n(s) = h_n(s) \int dt \prod_{k \neq n} h_k[s_k(x_n,y_n)] \quad . \tag{17}$$

He was therefore able in his MENT algorithm to use a very fast Gauss-Seidel iteration technique to solve for the $h_n(s)$. A modified solver has since been written by the author to improve the stability of the algorithm in the presence of noise.

In summary then, the maximum entropy distribution is favored as the one most easily produced by nature, and its form is always a simple product of Lagrange factors $h_n(s)$, one for each view. The arguments of the one-dimensional functions $h_n(s)$ are completely specified by the geometry, and their shape is adjusted to make the projections of the two-dimensional product function agree with the given data. The simplicity and generality of this result make it useful for many purposes, including noninterceptive beam diagnostics.

In the FMIT system at Los Alamos, this MENT algorithm is running as part of an integrated software system on an LSI 11/23 mounted in the same diagnostic node where the data is recorded. The solution usually converges in about 5 iterations, each of which takes a few seconds in the typical case of 3 or 4 views of 25 samples each. Figure 2 shows normalized projection data (+ signs) and the reprojected solution at 3 locations on the FMIT beam line. Figures 3 and 4 show contour and isometric plots of this emittance distribution.

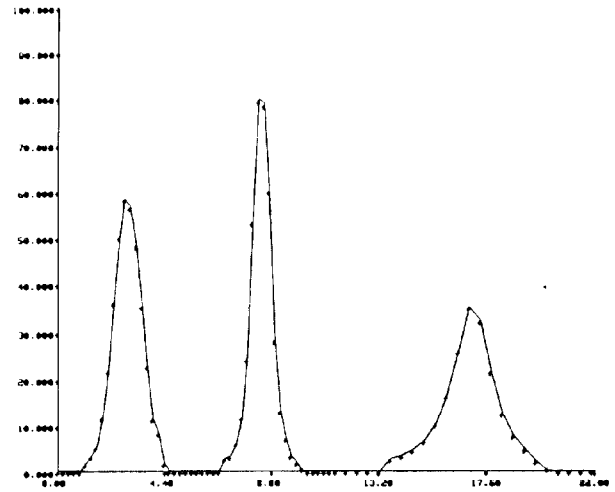

Fig. 2. Three views of profile data and maximum entropy fit.

Streak tomography is another application of this algorithm in which the time dependence of the instantaneous optical projection is recorded by a streak camera.[10] The streak image is raster scanned, and this algorithm is used to construct a two-dimensional frame of an output movie from the projection data stored in each scan line. In an experiment at Livermore, three views of the cross section of the electron beam pulse in the ETA were recorded with a streak camera.[11] The resulting time-slice profiles were processed at EG&G, Los Alamos, into a 700-frame video movie representing 42 ns of real time, thus demonstrating the feasibility of the technique.

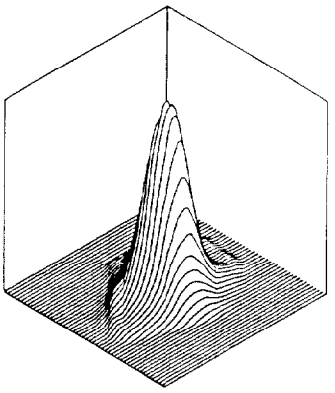

Fig. 3. Isometric display of maximum entropy solution for phase-space distribution function.

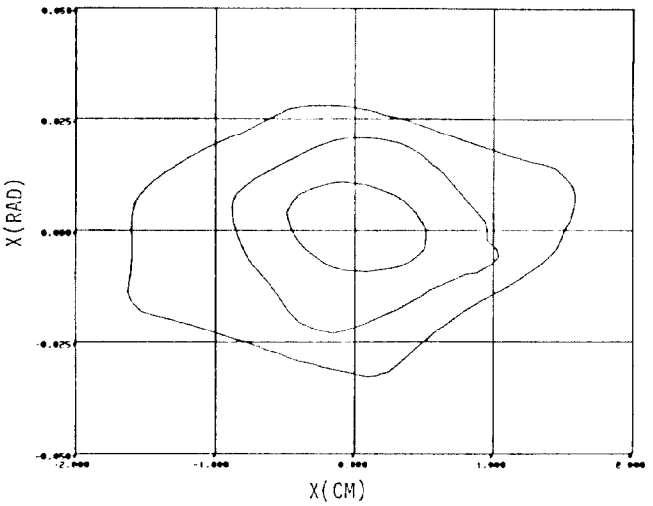

Fig. 4. Contour plot of maximum entropy solution for phase-space distribution function.

Acknowledgments

The FMIT beam diagnostic software, and the ETA movie were done at EG&G Los Alamos with the collaboration of D. Johnson. I would like to thank D. Chamberlin of Los Alamos National Laboratory for long-term support, G. Minerbo for copies of his code and many helpful discussions, H. Kohler and B. Jacoby of Livermore for use of their data in advance of publication, and J. Fraser and W. Cornelius of Los Alamos for help in preparing this paper.

References

1. D. D. Chamberlin, G. N. Minerbo, L. E. Teel, and J. D. Gilpatrick, "Non-Interceptive Transverse Beam Diagnostics," IEEE Trans. Nucl. Sci. 28, (3) (1981), 2347.
2. J. S. Fraser, "Developments in Non-Destructive Beam Diagnostics," IEE Trans. Nucl. Sci. 28, (3) (1981), 2137.
3. D. D. Chamberlin, G. N. Minerbo, and C. T. Mottershead, "Noninterceptive Transverse Beam Measurements," Proc. 1981 Linac Conf., Los Alamos National Laboratory report LA-9234-C (1981), 159.
4. D. D. Chamberlin, J. S. Hollabaugh and C. J. Stump, Jr., "Image Scope to Photodiode Beam-Profile Imaging System," IEEE Trans. Nucl. Sci. 30, (4) (1983), 2201.
5. G. N. Minerbo, "MENT: A Maximum Entropy Algorithm for Reconstructing a Source from Projection Data," Comp. Graphics Image Proc. 10 (1979), 48.
6. O. R. Sander, G. N. Minerbo, R. A. Jameson, and D. D. Chamberlin, "Beam Tomography in Two and Four Dimensions," Proc. 1979 Linac Conf., Brookhaven National Laboratory report BNL-51134 (1960), 314.
7. R. E. Turner and D. S. Betts, Introductory Statistical Mechanics, (Sussex University Press, 1974) Chap. 3.
8. E. T. Jaynes, "On the Rationale of Maximum-Entropy Methods," Proc. IEEE 70 (1982), 939.
9. E. T. Jaynes, "Prior Probabilities," IEEE Trans. Systems Sci. and Cybernetics, SSC-4 (1968), 227.
10. C. T. Mottershead, "A Tomographic Streak Camera for Ultrahigh-Speed Movies," IEEE Trans. Nucl. Sci. 29 (1) (1982), 900.
11. H. H. Koehler, B. Jacoby, and M. Nelson, "Time Resolved Tomographic Images of a Relativistic Electron Beam," Lawrence National Laboratory report UCRL-90692.