INTERNAL REPORT #85

THE MAXIMUM SMOOTHNESS METHOD

OF SPECTRUM RECONSTRUCTION

by

J. Zmuidzinas

Space Radiation Laboratory

California Institute of Technology

Pasadena, California

9-16-81

I. Introduction

The Maximum Smoothness method was originally created to solve a frequently ocurring problem in space physics, the problem of reconstructing a spectrum from an instrument's (TET, in our case) response functions and counting rates. The spectrum, response functions, and counting rates are all related by the following equation:

$$(1) \qquad r_i = \int_a^b s(E) f_i(E) dE \qquad i = 1, n$$

where r_i is the counting rate of type i events (events/sec), $f_i(E)$ is the instrument's effective geometry factor (cm^2–sr) for type i events as a function of energy (the response functions), and $s(E)$ is the isotropic particle spectrum (particles/cm^2–sr–MeV–sec). In the case of TET, the subscript i labels events according to their range. The functions $f_i(E)$ are usually determined experimentally and are assumed to be completely known in this report. The measurement of these functions for TET is described in Zmuidzinas and Gehrels, [1981]. Note that eqns. (1) do not determine $s(E)$ uniquely ; in fact, there is an uncountable number of functions $s(E)$ which satisfy eqns. (1) . The way information about a spectrum $s(E)$ is usually presented is by specifying the averages of $s(E)$ for n energy intervals (n = number of response functions) :

$$(2) \qquad s_j = \frac{1}{(E_j - E_{j-1})} \int_{E_{j-1}}^{E_j} s(E) dE \quad .$$

Just as eqns. (1) do not uniquely specify $s(E)$, they do not uniquely specify the s_j 's . In order to solve for the s_j 's, one has to make some assumptions about the spectrum $s(E)$.

II. Some Methods

Perhaps the simplest and most straightforward way of approaching this problem is by assuming the spectrum s(E) is a step function :

$$s(E) = s_j, \quad E_{j-1} \leq E < E_j, \quad j = 1, n \quad .$$

One must choose E_0 and E_n so that the entire region where the response functions are nonzero (call this region [a,b]) is covered, i.e. $E_0 \leq a$ and $E_n \geq b$. In this case, eqns. (1) can be written

$$r_i = \int_a^b s(E) f_i(E) dE$$

$$= \sum_{j=1}^n \int_{E_{j-1}}^{E_j} s(E) f_i(E) dE$$

$$= \sum_{j=1}^n s_j \int_{E_{j-1}}^{E_j} f_i(E) dE$$

With the definition

$$A_{ij} = \int_{E_{j-1}}^{E_j} f_i(E) \, dE$$

we get the following matrix equation :

$$(4) \qquad r_i = \sum_{j=1}^n A_{ij} s_j, \qquad i = 1, n \quad .$$

The A_{ij} matrix can then be inverted to yield the s_j 's in terms of the r_i 's. This method will be referred to as the MI method in this report. A way to improve this method is to try to take the

gross spectral shape into account. In space physics, spectra often have a power law form:

$$s(E) = A E^{-\gamma} \ .$$

Eqns. (1) are rewritten as

$$r_i = \sum_{j=1}^{n} \int_{E_{j-1}}^{E_j} s(E) f_i(E) dE$$

$$= \sum_{j=1}^{n} \frac{\int_{E_{j-1}}^{E_j} s(E) f_i(E) dE}{\frac{1}{\Delta E_j} \int_{E_{j-1}}^{E_j} s(E) dE} \times \frac{1}{\Delta E_j} \int_{E_{j-1}}^{E_j} s(E) dE$$

which gives

$$(5) \qquad r_i = \sum_{j=1}^{n} A_{ij} s_j$$

where

$$A_{ij} = \frac{\int_{E_{j-1}}^{E_j} s(E) f_i(E) dE}{\frac{1}{\Delta E_j} \int_{E_{j-1}}^{E_j} s(E) dE}$$

and $\Delta E_j = E_j - E_{j-1}$. The A_{ij} is calculated with a power law spectrum $s(E) = A E^{-\gamma}$. The matrix is then inverted to calculate the s_j 's. The A_{ij} 's are independent of A; the γ is chosen such that the calculated values of s_j are consistent with this γ. This is the method presented in Hoyng and Stevens, [1973] and will be referred to as the RMI (refined matrix inversion) method in this report.

Other methods solve for the s_j 's by calculating a specific $s(E)$ that satisfies eqns. (1) and

then integrating this s(E) over the various energy intervals. One way to calculate an s(E) is to assume a specific functional form. For example, assume that s(E) is a linear combination of a set of basis functions:

$$s(E) = \sum_{j=1}^{n} \alpha_j y_j(E) \quad .$$

Inserting this form in eqns. (1) gives a set of linear equations for the α_j 's :

$$r_i = \sum_{j=1}^{n} A_{ij}\alpha_j$$

where $A_{ij} = \int_{a}^{b} f_i(E)y_j(E)dE$. Once the α_j 's have been calculated, the spectrum s(E) and the s_j 's can easily be calculated. Note that both the MI and RMI methods are special cases of this method. The choices for the $y_j(E)$'s that correspond to the two methods are:

(MI) $\quad y_j(E) = 1 \quad E_{j-1} \le E < E_j$

(RMI) $\quad y_j(E) = E^{-\gamma} \quad E_{j-1} \le E < E_j$.

Note, however, that the γ in the RMI method changes.

The Maximum Smoothness method also calculates a specific s(E) ; however, the s(E) is chosen by an extremum principle with no explicit assumptions about the functional form. The next section describes this method.

III. The Maximum Smoothness Method

The Maximum Smoothness method (the MS method) calculates a specific s(E) by finding the s(E) which satisfies eqns. (1) and which minimizes

$$(6) \qquad I = \int_a^b \left[\frac{d^2 \log(s(E))}{d(\log(E))^2} \right]^2 d(\log(E))$$

The reasons for minimizing this integral will now be discussed. We would like the s(E) that we choose to have the following properties :

1) s(E) is a smooth, continuous function

2) If a power law is consistent with the counting rates r_i,

we would like s(E) to be that power law.

First of all, we must define what we mean by "smooth". After defining "smooth", we can always satisfy requirement 1) by choosing the "smoothest" possible s(E). Requirement 2) will automatically be satisfied if power laws are the "smoothest" possible functions according to our definition. Our intuitive notion of "smoothness" tells us that straight lines are the "smoothest" functions - not power laws. However, power laws are straight lines on a log-log plot. This can be expressed as

$$\frac{d(\log(s(E)))}{d(\log(E))} = \text{constant} \quad .$$

We shall therefore consider s(E) to be smooth if its derivative is nearly constant; i.e. its second derivative, $\frac{d^2 \log(s(E))}{d(\log(E))^2}$ is small. The integral

$$I = \int_a^b \left[\frac{d^2 \log(s(E))}{d(\log(E))^2} \right]^2 d(\log(E))$$

is thus a measure of the "smoothness" of s(E) on the interval [a,b]. The second derivative is squared to make both positive and negative values count equally.

IV. Mathematics of the Maximum Smoothness Method

This section treats the mathematical problem of finding a spectrum which satisfies eqns. (1) and minimizes the integral given in eqn. (6). We first treat a simpler problem whose solution is needed to solve the problem above. The problem: Find the $y(x)$ which satisfies

$$d_i = \int_a^b g_i(x) y(x) dx$$

and which minimizes

$$I = \int_a^b \left[\frac{d^2 y}{dx^2}\right]^2 dx \quad .$$

This can be solved using variational calculus with Lagrange multipliers (see Mathews and Walker, [1970]). We need to consider the following integral:

$$J = \int_a^b \left[\left[\frac{d^2 y}{dx^2}\right]^2 - \sum_{i=1}^n 2\lambda_i g_i(x) y(x)\right] dx \quad .$$

The λ_i 's are the Lagrange undetermined multipliers; the factor of 2 is included for convenience. To find a differential equation for $y(x)$, we perform a variation in y and set the variation in J to zero. We get:

$$y \rightarrow y + \delta y$$

$$\delta J = 2\frac{d^2 y}{dx^2}\frac{d\delta y}{dx}\bigg]_a^b - 2\frac{d^3 y}{dx^3}\delta y\bigg]_a^b + \int_a^b \left[2\frac{d^4 y}{dx^4} - \sum_{i=1}^n 2\lambda_i g_i(x)\right]\delta y dx$$

$$= 0 \quad .$$

First consider variations δy that satisfy

$$\delta y(a) = \delta y(b) = \frac{d\delta y}{dx}(a) = \frac{d\delta y}{dx}(b) = 0 \quad .$$

The terms in δJ which are left over from the partial integrations vanish and we are left with the integral term equal to 0. From this, we conclude that

$$(7) \qquad \frac{d^4 y}{dx^4} = \sum_{i=1}^{n} \lambda_i g_i(x) \quad .$$

Now, consider variations δy which vanish at the endpoints a and b but whose derivatives are arbitrary at these endpoints. Since the integral term has already been established to be 0, we are left with

$$0 = \delta J = 2 \frac{d^2 y}{dx^2} \frac{d\delta y}{dx} \Bigg]_a^b$$

from which we conclude that

$$(8) \qquad \frac{d^2 y}{dx^2}(a) = \frac{d^2 y}{dx^2}(b) = 0 \quad .$$

Similarly,

$$(9) \qquad \frac{d^3 y}{dx^3}(a) = \frac{d^3 y}{dx^3}(b) = 0 \quad .$$

We must now integrate the differential equation for y. First, we need notation for the integrals of $g_i(x)$:

$$G_i^{(n)}(x) = \int_a^x G_i^{(n-1)}(x) dx$$

and

$$G_i^{(0)}(x) = g_i(x) \ .$$

With this notation, we have (from eqns. 7, 8, and 9)

$$(10) \qquad y(x) = \alpha + \beta x + \sum_{i=1}^{n} \lambda_i G_i^{(4)}(x)$$

along with

$$(11) \qquad \sum_{i=1}^{n} \lambda_i G_i^{(1)}(b) = 0$$

$$(12) \qquad \sum_{i=1}^{n} \lambda_i G_i^{(2)}(b) = 0 \ .$$

Inserting (10) into the counting rate equations (eqns. 1) gives

$$(13) \qquad d_i = \alpha \int_a^b g_i(x)dx + \beta \int_a^b x g_i(x)dx + \sum_{j=1}^{n} \lambda_j \int_a^b G_j^{(4)}(x)g_i(x)dx \qquad i = 1, n \ .$$

Equations 11, 12, and 13 are n+2 equations for α, β, and λ_j, j = 1, n . Figure 1 shows these equations in a matrix form. This matrix is inverted to calculate α, β, and the λ_i's , which can then be used to calculate y(x) according to (10).

We will now return to the original problem. First of all, we make the following substitutions:

$$x = \log(E) , \quad \alpha = \log(a) , \quad \beta = \log(b)$$

$$y(x) = \log(s(E))$$

$$f'_i(x) = E\, f_i(E) \quad.$$

With these substitutions, the problem becomes: Find the $y(x)$ which satisfies

$$(14) \qquad r_i = \int_\alpha^\beta \exp(\, y(x)\,) f'_i(x) dx$$

and minimizes

$$(15) \qquad I = \int_\alpha^\beta \left[\frac{d^2 y}{dx^2}\right]^2 dx \quad.$$

If we apply variational calculus to this problem, we obtain a nonlinear differential equation which cannot be solved numerically since it contains the Lagrange multipliers which are unknown. We will instead develop an iterative scheme to calculate $y(x)$. Assume we have an approximation to $y(x)$, say $y^{(n-1)}(x)$. From $y^{(n-1)}(x)$ we would like to calculate an improved approximation $y^{(n)}(x)$, and then $y^{(n+1)}(x)$, etc. As this procedure converges, we will have $\left|y^{(n)}(x) - y^{(n-1)}(x)\right| \ll 1$. In this case,

$$\exp(\, y^{(n)}(x)\,) = \exp(\, y^{(n-1)}(x)\,) \cdot \exp(\, y^{(n)}(x) - y^{(n-1)}(x)\,)$$

$$\approx \exp(\, y^{(n-1)}(x)\,) \left[1 + y^{(n)}(x) - y^{(n-1)}(x)\right] \quad.$$

Using this approximation in the counting rate equations as a strict equality, we obtain

$$r_i = \int_\alpha^\beta \exp(\, y^{(n-1)}(x)\,) \left[1 + y^{(n)}(x) - y^{(n-1)}(x)\right] f'_i(x) dx$$

or

$$(16) \qquad d_i^{(n)} = \int_\alpha^\beta y^{(n)}(x) g_i^{(n)}(x) dx$$

where

$$d_i^{(n)} = r_i - \int_\alpha^\beta g_i^{(n)}(x) \left[1 - y^{(n-1)}(x) \right] dx$$

$$g_i^{(n)}(x) = \exp(y^{(n-1)}(x)) f'_i(x) \quad .$$

Note that $g_i^{(n)}(x)$ and $d_i^{(n)}$ can be calculated entirely from r_i, $f'_i(x)$, and $y^{(n-1)}(x)$. We must therefore find the $y^{(n)}(x)$ which satisfies eqns. (16) and which minimizes

$$I = \int_\alpha^\beta \left[\frac{d^2 y^{(n)}}{dx^2} \right]^2 dx \quad .$$

This is simply the problem solved in the beginning of this section. To start this iteration process, we use a $y^{(1)}(x)$ of the form $y^{(1)}(x) = a + \gamma x$. The initial a and γ are not critical ; better choices might require one less iteration for convergence. In our case, a and γ are chosen to minimize

$$\chi^2(a,\gamma) = \sum_{i=1}^n \left[\frac{\mu_i - c_i}{\sqrt{c_i}} \right]^2$$

where $r_i = c_i / \tau$, τ = livetime (c_i events of type i are observed in time τ), and

$$\mu_i / \tau = \int_\alpha^\beta \exp(a + \gamma x) f'_i(x) dx \quad .$$

V. Comparison of Methods

A comparison of the Maximum Smoothness (MS), the Matrix Inversion (MI), and the Refined Matrix Inversion (RMI) methods was made. The following steps were taken:

1) A spectrum s(E) was selected - for example, a power law spectrum, an exponential spectrum, etc.

2) The averages of the selected spectrum s(E) in each of the energy bins was calculated. These are the so-called "true fluxes".

3) The spectrum s(E) was multiplied by the response functions and integrated to yield the count rates r_i.

4) The fluxes in the energy bins were calculated with each of the three methods - MS, RMI, and MI.

5) The fluxes calculated in step 4) were compared to the "true" fluxes calculated in step 2). The average absolute error (defined below) was used as a figure of merit in this comparison.

Average absolute error: For each energy bin, the relative error of the calculated flux (step 4) as compared to the true flux (step 2) was calculated (for each method). The absolute values of these relative errors were then averaged over all of the energy bins to yield the average absolute error for each method on the particular trial spectrum.

Both the MS and RMI methods calculate fluxes consistent with a power law spectrum if some power law is consistent with the counting rates. The MI method is therefore the only method with a nonzero error for power law trial spectra. Figure 2 shows the average absolute

error of the MI method as a function of γ ($s(E) = AE^{-\gamma}$). As γ increases, the approximation that $s(E)$ is constant in each energy bin becomes worse, hence the error gets larger.

The average absolute error is plotted against spectrum number in Figure 3. The spectrum number simply labels the different trial spectra that were used in the comparison. They are plotted, along with the Maximum Smoothnes reconstruction of that trial spectrum, in Figures 4 through 10. The trial spectra were:

1) Exponential, $s(E) = \exp(-E/10\text{Mev})$. In this case, both the MS and RMI methods give similar errors, while the MI method is much worse.

2) Gaussian, $s(E) = \exp\left[-\frac{1}{2}\left[\frac{\log E - \log 20}{0.5}\right]^2\right]$ The MS method is much better than the other two in this case.

3) These spectra are all power laws with a break - i.e. there are two γ 's, one for low energies and one for high energies. The following functional form was used:

$$s(E) = A\left[\frac{1}{1 + (E/E_0)^n}\left(\frac{E}{E_0}\right)^{\gamma_1} + \frac{1}{1 + (E/E_0)^{-n}}\left(\frac{E}{E_0}\right)^{\gamma_2}\right]$$

for $E \ll E_0$, $s(E) = A(E/E_0)^{\gamma_1}$

for $E \gg E_0$, $s(E) = A(E/E_0)^{\gamma_2}$

In all cases, A was chosen to be 1 (this factor is unimportant since all methods are homogeneous - if the spectrum is multiplied by a factor of 2, so are the fluxes that these three methods calculate), and n (which determines how sharp the break is) was chosen to be 5.

a) $\gamma_1 = -1$, $\gamma_2 = -3$, and $E_0 = 20$ MeV . MS and RMI are fairly close in this case.

b) $\gamma_1 = -1$, $\gamma_2 = -3$, and $E_0 = 70$ MeV. RMI is much worse than MS in this case.

c) $\gamma_1 = -3$, $\gamma_2 = -1$, and $E_0 = 20$ MeV. As before, MS is much better than either of the other two methods.

4) Step function spectrum. This is an example of a spectrum that the MI method calculates exactly. It was tried mainly to compare the MS and RMI methods. The MS method is much better than the RMI method in this case.

5) Gaussian + Power Law, $s(E) = 500E^{-3} + \exp\left[-\frac{1}{2}\left(\frac{\log(E) - \log(16.6)}{0.2}\right)^2\right]$. Although none of the methods come very close, the MS method is by far the best. This example also serves to show how different spectra can be and yet give the same counting rates.

VI. Other Applications

The Maximum Smoothness method can also be used to find a smooth function $y(x)$ which passes through a set of points (x_i, y_i), $i = 1, n$, i.e. $y(x_i) = y_i$. We need to write these constraints in an integral form:

$$y_i = \int_a^b y(x)\delta(x-x_i)dx \quad .$$

These are analogous to the counting rate equations; the "response" functions in this case are Dirac delta functions. We then find the $y(x)$ which satisfies these equations and which minimizes

$$I = \int_a^b \left(\frac{d^2y}{dx^2}\right)^2 dx \quad .$$

The solution $y(x)$ turns out to be a cubic spline interpolation of the points (x_i, y_i) with the boundary conditions

$$\frac{d^2y}{dx^2}(a) = \frac{d^2y}{dx^2}(b) = \frac{d^3y}{dx^3}(a) = \frac{d^3y}{dx^3}(b) = 0 \quad .$$

This method was used to interpolate the response functions for TET. This method can be generalized to match derivatives of $y(x)$ in addition to values of $y(x)$ by using derivatives of the Dirac delta functions as "response functions".

FIGURE CAPTIONS

Figure 1   The matrix equation which occurs in the solution of the Maximum
Smoothness method.

Figure 2   Average absolute error of the MI method as a function of
power law gamma.

Figure 3   Average absolute error of the three methods (MI, RMI, and MS)
for all of the different trial spectra.

Figures 4-10   Plots of the trial spectra along with the Maximum Smoothness
reconstruction. Horizontal bars represent the average fluxes in the
energy bins of the plotted spectrum.

Figure 11   Plot of the response functions used (TET response functions;
D1-3 to D1-7; 0.5 MeV lower limit on D1 and D2 energy loss, 2.5 MeV
upper limit). Only the first four (D1-3 to D1-6) were used in the
spectrum reconstruction since the last one (D1-7) is not determined
well enough in the high energy region.

REFERENCES

Hoyng, P. and Stevens, G. A., On the Formation and Unfolding of Pulse Height Distributions, Astrophysics and Space Science, 27 (1974) 307.

Mathews, J. and Walker, R.L., Mathematical Methods of Physics, W. A. Benjamin, California, 1970.

Zmuidzinas, J. and Gehrels, N., The August, 1976 Electron Calibration of TET, SRL Internal Report #79, 1981

Matrix Equation

$$
\begin{bmatrix} \alpha & \beta & \gamma & \delta & \cdots & \lambda_n \end{bmatrix}
\times
\begin{bmatrix}
0 & 0 & G_1^{(1)}(b) & G_2^{(1)}(b) & \cdots & \cdots & G_n^{(1)}(b) \\
0 & 0 & G_1^{(2)}(b) & G_2^{(2)}(b) & \cdots & \cdots & G_n^{(2)}(b) \\
xg_1 & xg_2 & G_1^{(4)}g_1 & G_2^{(4)}g_1 & \cdots & \cdots & G_n^{(4)}g_1 \\
 & & G_1^{(4)}g_2 & G_2^{(4)}g_2 & \cdots & \cdots & G_n^{(4)}g_2 \\
\vdots & \vdots & \vdots & \vdots & & & \vdots \\
 & & G_1^{(4)}g_n & G_2^{(4)}g_n & \cdots & \cdots & G_n^{(4)}g_n
\end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}
$$

$$xg_i \equiv \int_a^b x\, g_i(x)\, dx$$

$$G_j^{(4)} g_i \equiv \int_a^b G_j^{(4)}(x)\, g_i(x)\, dx$$

Fig. 1

MI METHOD

$s(E) = A\bar{E}^{-\gamma}$

Average Absolute Error (%)

Fig. 2

Fig. 3

Average Absolute Error (%)

Spectrum #

+ — MI
× — RMI
⊢ — MS

(1712) (452)

Fig. 4

22

Fig. 5

Fig. 6

Fig. 7

Fig. 6

Fig. 9

Fig. 10

Fig. 11

Energy (MeV)

Response (Dimensionless)