JOURNAL OF COMPUTATIONAL PHYSICS 110, 221–233 (1994)

Dual Principles in Maximum Entropy Reconstruction of the Wave Distribution Function

TORD OSCARSSON

Swedish Institute of Space Physics, University of Umeå, Umeå, Sweden

Received September 23, 1992; revised March 8, 1993

Maximum entropy methods are used for reconstructing the distribution of energy in wave vector space from frequency spectra observed on board satellites. The reconstruction scheme is based on a modified entropy function, and dual principles are used to solve the resulting optimization problem. Our scheme is not limited to reconstructions of wave distribution functions, but it should be useful also for solving other types of underdetermined inverse problems.

1. INTRODUCTION

Charged particles emerging from the sun form the so-called solar wind. The solar wind flows radially outwards from the sun into the interplanetary space. When the solar wind particles reach a distance of about 50000–75000 km from the Earth, they interact strongly with the planetary magnetic field. The interaction diverts the flow around the Earth in a collisionless shock, the bow shock. The region of space formed behind the shock front is called the Earth's magnetosphere (see Fig. 1). Satellite missions have revealed the existence of magnetospheres also around other planets, such as Jupiter, Saturn, Uranus, and Neptune. Magnetospheres are believed to exist around the sun and other stars, and even galaxies may have their own magnetospheres.

In the magnetosphere and the solar wind, as well as in the sun itself, matter is in the state of a highly ionized gas, a plasma. The plasma state dominates our solar system and is believed to prevail throughout the universe. As plasmas are formed by free charged particles, they are rich in electric and magnetic wave phenomena. Waves with frequencies ranging from below 1 Hz up to a few megahertz can be generated and absorbed spontaneously in the Earth's magnetosphere. These electric and magnetic wave fields play an active role in the dynamics of the magnetosphere. For instance, ion-related electromagnetic emissions close to the Earth are believed to heat ions at low altitudes, thereby contributing to the outflow of oxygen from the Earth's atmosphere.

Theoretical investigations of plasma waves are often performed by Fourier transforming the fields according to

$$\mathbf{E}(\mathbf{k}, \omega) = \int dt\, d\boldsymbol{\rho} \mathbf{E}(\boldsymbol{\rho}, t) \exp[i(\omega t - \mathbf{k} \cdot \boldsymbol{\rho})], \tag{1}$$

where $\mathbf{E}(\boldsymbol{\rho}, t)$ is the electric field as a function of the spatial coordinates $\boldsymbol{\rho}$ and time t. Furthermore, ω is the angular frequency and $\mathbf{k}$ is the wave vector. Transforming the Vlasov–Maxwells equations, which govern the dynamics of collisionless plasmas, results in a dispersion relation [1]

$$\det \mathbf{D}(\mathbf{k}, \omega) = 0, \tag{2}$$

where det $\mathbf{D}$ stands for the determinant of $\mathbf{D}$. The dispersion tensor $\mathbf{D}$ is related to the dielectric tensor $\boldsymbol{\varepsilon}(\mathbf{k}, \omega)$ through the relation

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon_0[(\mathbf{kk} - \mathbf{I}k^2)\, c^2 + \omega^2 \boldsymbol{\varepsilon}(\mathbf{k}, \omega)], \tag{3}$$

where ε_0 is the permittivity of free space, $\mathbf{I}$ is the unit tensor, c is the speed of light in vacuum, and $k = |\mathbf{k}|$. We denote the solution of the dispersion relation by $\omega = \Omega(\mathbf{k})$. In the dissipationless limit that we consider here, ω is always real. From the Vlasov–Maxwells equations we also obtain the polarization vectors $\mathbf{e}$ and $\mathbf{b}$ for the wave electric and magnetic field, respectively.

The key parameter for understanding the physical processes generating space plasma waves is the wave distribution function (WDF). The WDF is defined in terms of the dispersion tensor as [2]

$$\psi(\mathbf{k}) = \mathbf{E}^*(\mathbf{k}) \cdot \left[\frac{1}{\omega}\, \partial_\omega\, \mathbf{D}(\mathbf{k}, \omega)\right] \cdot \mathbf{E}(\mathbf{k}). \tag{4}$$

Here the bracket is evaluated at $\omega = \Omega(\mathbf{k})$, and $\mathbf{E}^*(\mathbf{k})$ is the complex conjugate of the spatial Fourier transform of the electric field. As discussed elsewhere [2, 3], ψ is in reality

0021-9991/94 $6.00

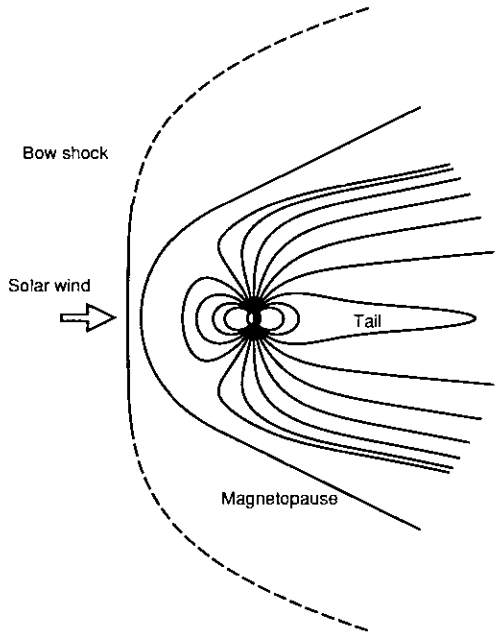

FIG. 1. Schematic picture of the Earth's magnetosphere. The dipolar magnetic field of the Earth is compressed by the solar wind on the side of the planet facing the sun and elongated in a long tail on the anti-sunward side. The boundary of the magnetosphere is called the magnetopause.

the energy density in configuration and wave vector space, having the dimension of energy. For our purpose here, the WDF can be thought of as the energy density in **k**-space. Being an energy density, $\psi(\mathbf{k})$ is strictly positive.

The central role played by the WDF makes it the crucial *quantity to measure. However, satellites can measure only* at one or at most a few points in space, making a spatial spectrum impossible to compute. Instead, wave observations are normally processed to produce frequency spectra. A frequency spectrum $C(\omega)$ can be either the power spectrum of a given electric or magnetic field component, or the cross spectrum between two different field components. Theoretically, frequency spectra are related to the WDF through an equation of the form [3, 4]

$$C(\omega) = S(\omega; \psi), \tag{5}$$

where

$$S(\omega; \psi) = \int d\mathbf{k}\, A(\mathbf{k})\, \psi(\mathbf{k})\, \delta(\omega - \Omega(\mathbf{k})). \tag{6}$$

Here $\delta(\ )$ denotes Dirac's delta function. The kernels $A(\mathbf{k})$ depend on the polarization vectors **e** and/or **b** and on the characteristics of the measuring antennas.

The measured frequency spectra contain vital information about the WDF, and it was Storey and Lefeuvre [4] who first suggested that frequency spectra could be used to reconstruct the WDF. Reconstruction of the WDF is a type of ill posed problem that occurs frequently in many research areas [5–7]. While these problems cannot be solved exactly, their importance has promoted the use of various reconstruction techniques for extracting as much information as possible from the available data. The purpose of this paper is to present a new version of the maximum entropy reconstruction method that we have found useful for resolving the WDF problem.

It should be emphasized that our paper presents a modification of the classic maximum entropy method. Recently developed inversion techniques use Bayesian statistics to perform reconstructions as well as for estimating the reliability of the solution [8]. Extending our method to include Bayesian statistics will be the subject of future work.

The method we use is based on a modification of the traditional entropy of a continuous function. The modified entropy function automatically compensates for the lack of information about the normalization of the WDF. By transforming the resulting optimization problem into a dual problem, a great reduction in dimensionality is achieved. Depending on the particular problem, the number of degrees of freedom can be reduced by a factor 100 or even more. These particular features do not pertain particularly to reconstruction of the WDF, but they should be of equal importance in many other situations where continuous distributions are reconstructed from sparse data. Here we demonstrate how the method works in the WDF context and hope that our examples will arouse interest in our method.

2. MAXIMUM ENTROPY RECONSTRUCTION OF THE WDF

Most physicists are acquainted with the entropy concept from statistical physics. The canonical distribution, for example, can be found by maximizing the entropy of a system at fixed density and temperature. During the last decades, starting with Jaynes [9–11], the entropy concept has found its way to other research areas with little or no connection to statistical physics. The maximum entropy (M-E) method is becoming one of the standard tools for solving ill-posed problems in, for instance, image decoding [12], radio astronomy [13], seismology [14], and computer tomography [15]. The method is used also in spectral analysis [16] and in the construction of expert systems [17]. However, even though M-E has been used for several decades, a completely unified implementation of the method still does not exist. Various experimental constraints are used [13, 18], and even the appropriate form of the entropy function is debated [13, 19]. Here we discuss the relevant entropy of the WDF and the constraints imposed by the observed frequency spectra.

2.1. *Conventional Definitions of Entropy*

The M-E method is often introduced by considering the loaded dice problem [20]. Given the average number of eyes turning up in a series of N tosses, how many times, N_i, did the ith face turn up? Alternatively, we can ask what is the probability p_i of turning up the ith face in the next toss? To give the most plausible answer to this question M-E advocates write down the multiplicity

$$M = \frac{N!}{\Pi_{i=1}^{6} N_i!} \tag{7}$$

and argue that the most likely numbers N_i of those consistent with the observed average are the ones maximizing M. Assuming N to be sufficiently large and using Stirlings approximation one can write

$$\log M = -\sum_{i=1}^{6} N_i \log(N_i/N) \tag{8}$$

or introduce the entropy H_0 by writing

$$H_0 = \frac{1}{N} \log M = -\sum_{i=1}^{6} p_i \log(p_i), \tag{9}$$

where $p_i = N_i/N$. When the total number of tosses is known, maximizing the multiplicity is equivalent to finding the probability distribution maximizing the entropy. Note that $p_i = \text{const} = \frac{1}{6}$ maximizes the entropy when no constraint other than $\sum p_i = 1$ is imposed.

The M-E resolution of the loaded dice problem can be transferred to other underdetermined inverse problems in what we may define as the M-E spirit—of all solutions consistent with the data, choose the one having the highest entropy. In contrast to the loaded dice problem, physical distributions are normally continuous rather than discrete. In accordance with (9), the entropy of a continuous function $f(y)$ assumed to be positive everywhere is often defined by

$$H_1(f) = -\int dy f(y) \log\left[\frac{f(y)}{\Delta(y)}\right]. \tag{10}$$

Jaynes [11] calls Δ an "invariant measure," which is introduced to ensure that the entropy scales properly under coordinate transformations. In practice, when it comes to numerical computations, the continuous problem is normally discretized and the integral replaced by a sum. The analogy between (10) and (9) is then clear, especially when $\Delta(y) = 1$. In any case, the global maximum of $H_1(f)$ is found for

$$f(y) = e^{-1} \Delta(y). \tag{11}$$

Some authors add $-f$ to the entropy to avoid the factor e^{-1} [21, 22]. The M-E method strives to satisfy the constraints imposed by the data while remaining as close as possible to the global optimum (11). Thereby Δ gives the solution a certain bias, and this creates a valuable flexibility. In practice one often uses $\Delta(y) = \text{const} = \Delta_c$.

2.2. *The Entropy of the WDF*

In many applications interest is focused on certain structures in the distribution being reconstructed. Astronomers performing image reconstruction may be interested in proving the existence of a double star and computerized tomography searches for structures inside objects using projection measurements. In space physics structures in the distribution of energy in **k**-space contain information about the underlying physical process generating the waves and about the location of the source region in space. In these cases it is essential that the reconstruction method only reproduces structures required by the data. Spurious peaks must be suppressed.

Having stressed the importance of structures in the WDF, it is clear that whenever the data are consistent with a completely structureless, constant WDF this must also be the solution produced by the reconstruction scheme. Otherwise, if our inversion technique fails to reproduce a flat WDF even in the case when $\psi(\mathbf{k}) = \text{const} = c$ is consistent with the data, we could easily be lead to conclusions not supported by the data in the general case. A M-E method based on (10) satisfies the above requirement, provided that we choose $\Delta_c = ec$. However, since no information about ψ is available at the outset of the analysis, it is not clear how the constant c should be chosen.

To illustrate the problems that the missing information about normalization can lead to, we consider a set of six scales. The scales are biased so that a load of w kg causes the ith scale to deviate by an amount $d_i = iw$. Let w_i be the unknown load on the ith scale. Furthermore, assume that the total load $W = \sum_i w_i$ is unknown, but that the averages

$$\langle d \rangle = \tfrac{1}{6} \sum_{i=1}^{6} d_i \tag{12}$$

$$\langle d^2 \rangle = \tfrac{1}{6} \sum_{i=1}^{6} d_i^2 \tag{13}$$

are given by $\langle d \rangle = 3.5$ and $\langle d^2 \rangle = 15.17$. Figure 2 shows the results obtained by reconstructing the load distribution using a discretized version of (10) with $e^{-1} \Delta_c = 0.5$, 1, and 2, respectively. With $\Delta_c = e$ the M-E method found the constant solution $w_i = 1$ consistent with the data. However, in the other two cases the reconstruction method imposes structures on the solutions that we would falsely credit inhomogeneities in the true load distribution had we no other information available.

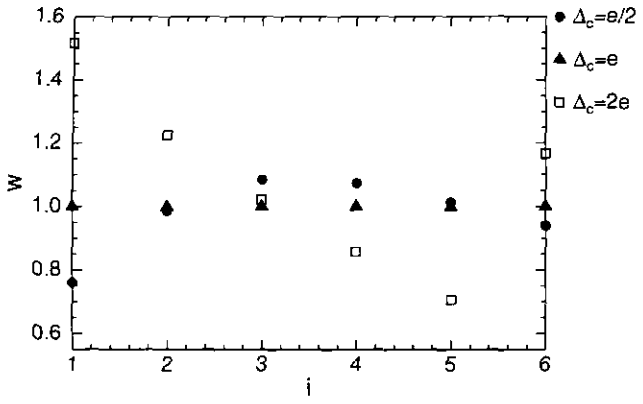

FIG. 2. Continuous "loaded dice" problem. The load distribution on six biased scales is reconstructed from observations of the averages $\langle d \rangle = 3.5$ and $\langle d^2 \rangle = 15.17$ using three different constant values $\Delta = \Delta_c$.

The solutions presented in Fig. 2 are of the form

$$w_i = \Delta_c \exp\left[\sum_{p=1}^{2} \lambda_p i p (i w_i)^{p-1} - 1\right], \tag{14}$$

where λ_1 and λ_2 are the Lagrange multipliers for the constraint in (12) and (13), respectively. The results in Fig. 2 can be verified by inserting $\lambda_1 = 0.470085$ and $\lambda_2 = -0.032424$ for the case $\Delta_c = e/2$, and $\lambda_1 = -0.328474$ and $\lambda_2 = 0.017046$ for the case $\Delta_c = 2e$.

The non-linear constraint (13) is used in order to simply the exposition. However, it should be noted that the results from the analysis do not depend crucially on the fact that one of the constraints is non-linear in w. Similar results are obtained when using (12), together with the linear constraint

$$\tfrac{1}{6} \sum_{i=1}^{6} i^2 w_i = 15.17. \tag{15}$$

This can be verified by applying the method of Lagrange multipliers using $\lambda_1 = 0.413846$ and $\lambda_2 = -0.052777$, and $\lambda_1 = -0.432553$ and $\lambda_2 = 0.055953$ for the cases $\Delta_c = e/2$ and $\Delta_c = 2e$, respectively. Inhomogeneities in the reconstruction shown in Fig. 2 do not depend on our particular choice of constraints, but on our failure to prescribe the proper entropy function.

Clearly, reconstructions based on (10) depend crucially on the choice of Δ. One value of Δ_c may produce a constant WDF while some other value may not. The problem would be resolved if the total energy density $\Psi = \int d\mathbf{k}\, \psi(\mathbf{k})$ was known beforehand. Then the only constant WDF allowed would be the one consistent with Ψ, and we could choose

$$\Delta_c = \frac{1}{V} e\Psi, \tag{16}$$

where V is the volume in $\mathbf{k}$-space. The problem of no a priori knowledge of the normalization occurs also in, for instance, image reconstruction. There one often circumvents the problem by using an estimate of the corresponding integrated quantity [13, 23]. A similar procedure could be used also when reconstructing the WDF. One can also envisage a scheme based on iterative reconstructions, where the Ψ value of the previous reconstruction is used in the next. However, neither of these methods is quite satisfactory. To first estimate Ψ from the observed frequency spectra and then reconstruct the WDF using the same data is not in line with the M-E philosophy. Nor does the use of iterative reconstructions seem to be the most consistent implementation of the M-E method. Instead we draw the conclusion that (10) does not define the appropriate entropy of the WDF.

To find a suitable form for the entropy of the WDF, we return to the loaded dice problem. The argument for resolving the loaded dice problem by maximizing the entropy in (9) is that this automatically maximizes the multiplicity. However, this is true only when the total number of tosses is known. When N is unknown, which is the analogue to the WDF case, the problem should be resolved by maximizing the multiplicity directly. Hence, in analogy with (8) rather than with (9), we define the entropy of the WDF as

$$H(\psi) = -\int d\mathbf{k}\, \psi(\mathbf{k}) \log\left[\frac{\psi(\mathbf{k})}{\Psi\, \Delta(\mathbf{k})}\right]. \tag{17}$$

This entropy has a global maximum when

$$\frac{\psi(\mathbf{k})}{\Psi} = \Delta(\mathbf{k}). \tag{18}$$

Note that the existence of this solution requires that $\int \Delta(\mathbf{k})\, d\mathbf{k} = 1$. For any Δ satisfying this condition we see that the WDF at the global maximum is determined only to within a multiplicative factor. In particular, for $\Delta = \Delta_c = 1/V$ the entropy is at maximum for any constant WDF. This means that whenever $\psi(\mathbf{k}) = c$ is consistent with the data, this solution maximizes the entropy irrespective of the value of the constant c.

The problem of using M-E reconstruction when the normalization $F = \int f(y)\, dy$ is unknown is not new. Surprisingly, although the entropy definition in (17) is a natural choice considering the multiplicity argument for M-E, it is not used to circumvent the problem. Instead, a common way of resolving the problem is to define the entropy as [24–27]

$$H_2(f) = -\int dy \frac{f(y)}{F} \log\left[\frac{f(y)}{F\, \Delta(y)}\right]. \tag{19}$$

The entropies $H(f)$ and $H_2(f)$ both attain maximum values for distributions satisfying $f(y)/F = \Delta(y)$. However, we believe that the entropy in (17) is more in line with the general M-E reasoning, and there seems to be nothing in the latter form justifying the extra complexity introduced in the computation of the gradient and the Hessian matrix. Hence, we define $H(\psi)$ to be the entropy of the WDF and base our reconstructions on this.

2.3. *Summary of the Reconstruction Scheme*

Experimental data are always afflicted by errors. Hence, WDFs reproducing the data exactly are not the only solutions consistent with the data. In fact, an exact reproduction of the observations may be impossible due to noise. Instead of demanding that the WDF should satisfy (5) exactly, we take the statistical approach used by Skilling and Gull and others [13, 28].

Before writing down the constraint function a few words on notations are required. Our data consist of a certain combination of power and cross-spectra measured at a finite number of discrete frequencies. To keep notations short we arrange these data in a single vector $\mathbf{C}$ of length m, where m is the total number of independent observations. This means that $\mathbf{C}$ does not depend on ω. Instead the frequency determines the position of a certain spectrum within the vector $\mathbf{C}$. A similar vector notation is used throughout also for $S(\omega;\psi)$. To formulate the constraint imposed on the solution by the data we define

$$\xi(\psi) = \frac{1}{2}\sum_{j=1}^{m} \frac{|C_j - S_j(\psi)|^2}{\sigma_j^2}, \tag{20}$$

where C_j is the jth element of $\mathbf{C}$. Assuming the errors to be random and normally distributed with known variance σ_j^2, we can use the χ^2 distribution to set an upper bound, ξ_{bound}, on $\xi(\psi)$. At the 95% level the upper limit is roughly equal to $m/2$. This is a satisfactory procedure even though in general the variancies are not known. We come back to this point later in Section 3.

Any WDF generating a ξ value below ξ_{bound} is considered a feasible solution to the inverse problem. In accordance with the M-E spirit discussed earlier, our reconstruction scheme sums up to

$$\begin{aligned} &\text{maximize} && H(\psi) \\ &\text{subject to} && \xi(\psi) \leqslant \xi_{\text{bound}}. \end{aligned} \tag{21}$$

In the following we consider the numerical solution of (21), and in particular we discuss a dual formulation of the problem.

3. NUMERICAL SOLUTION

3.1. *The Original Problem*

The numerical integration in (6) is facilitated by first averaging over a finite frequency interval $\Delta\omega$ centered on $\omega = \omega_0$. The $\mathbf{k}$-integral is then performed over the volume in wave vector space where the dispersion relation $\omega = \Omega(\mathbf{k})$ is satisfied within the frequency interval $[\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2]$. We introduce a mesh in $\mathbf{k}$-space and denote by δ_l the volume of the grid element centered on the lth grid point, $\mathbf{k}_l$. Defining

$$x_l = \psi(\mathbf{k}_l)\,\delta_l, \tag{22}$$

we can replace the $\mathbf{k}$-integral by the sum

$$\mathbf{S}(\mathbf{x}) = \sum_{l=1}^{n} x_l \mathbf{A}_l, \tag{23}$$

where n is the number of grid points in $\mathbf{k}$-space, and $\mathbf{x}$ contains the n values of $\psi(\mathbf{k}_l)\,\delta_l$. In practice, n is always much larger than m, the number of independent observations. The vector $\mathbf{A}_l$ contains the kernel values at the lth grid point and has the same dimension, m, as the vectors $\mathbf{C}$ and $\mathbf{S}$.

Similarly, we write the entropy as the sum

$$H(\mathbf{x}) = -\sum_{l=1}^{n} x_l \ln\left[\frac{x_l}{X}\Big/\frac{\delta_l}{V}\right], \tag{24}$$

where

$$X = \sum_{l'=1}^{n} x_{l'} \tag{25}$$

and

$$V = \sum_{l'=1}^{n} \delta_{l'}. \tag{26}$$

A more detailed description of the discretization procedure is given elsewhere [3, 29].

Introducing the negative of the entropy $H^-(\mathbf{x})$, the discretized reconstruction problem can be written as the minimization problem

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} && H^-(\mathbf{x}) \\ &\text{subject to} && \xi(\mathbf{x}) \leqslant \xi_{\text{bound}}, \end{aligned} \tag{27}$$

where the constraint function is given by

$$\xi(\mathbf{x}) = \tfrac{1}{2}\,|\bar{\mathbf{C}} - \bar{\mathbf{S}}(\mathbf{x})|^2. \tag{28}$$

Here we have included the standard deviations in the normalization of $\mathbf{C}$ and $\mathbf{S}$. Accordingly, the jth component of the vectors $\bar{\mathbf{C}}$ and $\bar{\mathbf{A}}_l$ are given by C_j/σ_j and $(\mathbf{A}_l)_j/\sigma_j$, respectively.

The optimization problem in (27) is a convex program [30]. As such it has several desirable qualities, the most important being that the solution is unique. Except for the trivial case when the data are consistent with a constant WDF, the solution of (27) satisfies $\xi(\mathbf{x}) = \xi_{bound}$. The solution can then be obtained by minimizing the Lagrangian function

$$L(\mathbf{x}) = \mu H^{-}(\mathbf{x}) + \xi(\mathbf{x}) \tag{29}$$

with respect to $\mathbf{x}$ for fixed values of μ. We know from the convex nature of the problem that $\mu > 0$ and that its value should be chosen so that $\xi(\mathbf{x}) = \xi_{\text{bound}}$ at the minimum of L. To begin, we solved this problem using Monte-Carlo techniques [3] which are simple to implement but slow and inefficient. In our later work we use an improved algorithm based on conjugate gradient methods and reconstruct WDFs from real satellite data [31, 32]. However, the large size of the problem also makes conjugate gradient algorithms slow and inefficient. The largest problem solved with conjugate gradient methods has $n \approx 1500$ and $m = 20$. Some interesting things can be done at this size, but a really useful WDF reconstruction scheme must be able to handle larger problems. For this purpose we develop a dual formulation, which exploits the much lower dimension m of the observed spectral densities.

3.2. *The Dual Problem*

Minimization of $L(\mathbf{x})$ for fixed μ gives

$$\frac{x_l}{X} = \frac{\delta_l}{V} \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right], \tag{30}$$

where

$$\mathbf{r} = \bar{\mathbf{C}} - \sum_{l=1}^{n} x_l \bar{\mathbf{A}}_l. \tag{31}$$

Equation (30), together with the constraint $\xi = \xi_{\text{bound}}$, defines $\mathbf{x}$ implicitly. However, these equations are just as difficult to solve as minimizing the Lagrangian function in (29). If we instead treat $\mathbf{r}$ and X as the unknowns, the solution of (27) can be obtained by solving the dual problem

$$\mathbf{r} = \bar{\mathbf{C}} - X \sum_{l=1}^{n} \frac{\delta_l}{V} \bar{\mathbf{A}}_l \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right] \tag{32}$$

$$1 = \sum_{l=1}^{n} \frac{\delta_l}{V} \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right], \tag{33}$$

where (33) comes from summing both sides of (30). Now μ should be adjusted so that

$$\xi(\mathbf{r}) = \tfrac{1}{2}\mathbf{r}\cdot\mathbf{r} = \xi_{bound}. \tag{34}$$

To solve the system of non-linear equations (32) and (33), we recast the problem by noting that (32) can be written as

$$\begin{aligned}\mathbf{r} - \bar{\mathbf{C}} + X \sum_{l=1}^{n} \frac{\delta_l}{V} \bar{\mathbf{A}}_l \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right] &= \frac{\partial}{\partial \mathbf{r}}\left\{\frac{1}{2}\mathbf{r}\cdot\mathbf{r} - \mathbf{r}\cdot\bar{\mathbf{C}} + \mu X \sum_{l=1}^{n} \frac{\delta_l}{V} \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right]\right\} \\ &= 0.\end{aligned} \tag{35}$$

Considering μ as fixed and letting the product μX play the role of a Lagrange multiplier, eq. (35) can also be phrased as the constrained optimization problem

$$\begin{aligned} &\underset{\mathbf{r}}{\text{minimize}} && f(\mathbf{r}) \\ &\text{subject to} && q(\mathbf{r}, \mu) = 1, \end{aligned} \tag{36}$$

where

$$f(\mathbf{r}) = \tfrac{1}{2}(\mathbf{r} - \bar{\mathbf{C}})\cdot(\mathbf{r} - \bar{\mathbf{C}}) \tag{37}$$

and

$$q(\mathbf{r}, \mu) = \sum_{l=1}^{n} \frac{\delta_l}{V} \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right]. \tag{38}$$

Note that the equality constraint ensures that (33) is satisfied. The optimization problem (36) can be solved by minimizing the dual Lagrangian function

$$L_D(\mathbf{r}, \mu, \lambda) = f(\mathbf{r}) + \lambda q(\mathbf{r}, \mu) \tag{39}$$

for fixed values of the Lagrange multipliers μ and λ. The simultaneous solution of (32) and (33) requires that $\lambda = \mu X$ when $q(\mathbf{r}, \mu) = 1$ at the minimum of L_D. Consequently, when a solution to the dual problem has been found, $\psi(\mathbf{k}_l)$ can be retrieved through eqs. (30) and (22), with $X = \lambda/\mu$.

Obviously $f(\mathbf{r})$ is convex and, since the Hessian of q is given by

$$\nabla^2 q(\mathbf{r}, \mu) = \sum_{l=1}^{n} \frac{\delta_l}{V} \frac{\bar{\mathbf{A}}_l}{\mu} \frac{\bar{\mathbf{A}}_l}{\mu} \exp\left[\frac{1}{\mu}\mathbf{r}\cdot\bar{\mathbf{A}}_l\right], \tag{40}$$

we have $\mathbf{v}\cdot\nabla^2 q\cdot\mathbf{v} \geqslant 0$ for any vector $\mathbf{v}$, and consequently, $q(\mathbf{r}, \mu)$ is also convex. Thus, the original convex optimization problem in n dimensions is replaced by another convex problem in m unknowns, and thereby, the dual problem has

the same nice qualities as the original problem. Furthermore, since in general $n \gg m$, solving the dual instead of the original problem leads to a tremendous reduction in computational cost.

3.3. *Optimization Techniques*

Our scheme for solving the dual problem can be split into two parts. One part consists of the subproblem of minimizing the dual Lagrangian function with respect to $\mathbf{r}$ for given values of μ and λ. The other part involves varying μ and λ until the constraints (33) and (34) are fulfilled at the minimum of L_D.

To minimize L_D for fixed values of μ and λ we use Newtons method [30]. Let $\mathbf{h}$ denote the Hessian of L_D so that

$$\mathbf{h} = \nabla_{\mathbf{r}}^2 L_D(\mathbf{r}, \mu, \lambda). \tag{41}$$

Starting the iteration from an arbitrary point $\mathbf{r}$, the step length α along the Newton direction

$$\mathbf{p}_N = -\mathbf{h}^{-1} \frac{\partial L_D}{\partial \mathbf{r}} \tag{42}$$

is determined through a line search along $\mathbf{p}_N$. The line search is performed with the Van Wijngaarden–Dekker–Brent method [33] until

$$\frac{\partial L_D}{\partial \mathbf{r}} (\mathbf{r} + \alpha \mathbf{p}_N, \mu, \lambda) \cdot \mathbf{p}_N = 0 \tag{43}$$

is satisfied to within desired accuracy. Once the step length is computed, $\mathbf{r}$ is updated according to $\mathbf{r} \to \mathbf{r} + \alpha \mathbf{p}_N$, and the iteration is continued until the minimum of L_D is found. The solution thus obtained is denoted by $\hat{\mathbf{r}}(\mu, \lambda)$.

To determine μ and λ we first treat μ as fixed and consider $q(\hat{\mathbf{r}}, \mu)$ as a function of λ. We then apply the Van Wijngaarden–Dekker–Brent method to determine the $\hat{\lambda}$ satisfying $q(\hat{\mathbf{r}}(\mu, \hat{\lambda}), \mu) = 1$. Finally, we treat $\xi(\hat{\mathbf{r}}(\mu, \hat{\lambda}))$ as a function of μ and use the Van Wijngaarden–Dekker–Brent method to determine the $\hat{\mu}$ causing $\hat{\mathbf{r}}(\hat{\mu}, \hat{\lambda})$ to satisfy (34). This is probably not the most efficient way of determining $\hat{\mu}$ and $\hat{\lambda}$, but it is robust.

To exploit possible linear dependencies among the kernels $\bar{\mathbf{A}}_l$ we introduce a suitable set of orthogonal basis vectors. The procedure, which is outlined in Appendix A, reduces the dimension of the problem when there are less than m linearly independent kernel vectors. Numerical problems sometimes occur due to the exponential in (38). Even though it should be well-behaved at the solution, the exponential may cause overflow during a specific iteration. To remedy this problem we use the logarithmic constraint $\ln[q(\mathbf{r}, \mu)] = 0$ rather than $q(\mathbf{r}, \mu) = 1$. The logarithm of q can be computed accurately while avoiding overflow. Furthermore, since $q = 1$ at the solution of (36), the Lagrange multiplier solving the subproblem with logarithmic constraint is identical to $\hat{\lambda}$.

In reality we establish convergence in four different ways. At the lowest level we use the convergence criterion described in [30, p. 306] to determine when the minimization of L_D for fixed μ and λ has succeeded. The subproblem must be solved accurately due to the exponential relationship between $\mathbf{x}$ and $\mathbf{r}$, which can blow up small errors in $\hat{\mathbf{r}}$. We perform the optimization in double precision and require that the constraint $\ln[q] = 0$ must not be violated by more than 10^{-7}. When the λ-iteration has converged to within this limit, we check that the solution has converged also in the sense that the resulting $\mathbf{x}$ computed from (30) is really the minimum of $L(\mathbf{x})$. To do this we compute

$$\beta = \left[\frac{1}{n} \sum_{l=1}^{n} \left(\frac{\partial L / \partial x_l}{\mu \, |\partial H^- / \partial x_l| + |\partial \xi / \partial x_l|} \right)^2 \right]^{1/2}, \tag{44}$$

where a value of $\beta < 0.01$ is considered acceptable. Finally, we make sure that the value of $\xi(\mathbf{r})$ is consistent with the value of ξ_{bound}. The requirement on this constraint is not as strict as on $q(\mathbf{r}, \mu)$, since we only have a rough estimate of ξ_{bound}. This fact greatly simplifies computations. On the other hand, each problem must be solved with several ξ-values to make sure that the solution does not depend strongly on the exact value of ξ_{bound}. In cases where the reconstructed WDF is unstable to small variations in ξ_{bound} the solution is not reliable and must be rejected. In all cases tried so far, the main features of the solutions have been stable to substantial variations in ξ_{bound}, and this has convinced us that the χ^2 estimate of ξ_{bound} can be used even though the variances are not known exactly.

The algorithms and numerical methods described above have been implemented in the FORTRAN-package MEREWASP (maximum entropy reconstruction of wave distribution functions in space plasmas). In the next section we demonstrate how the scheme operates by applying our method to synthetic as well as to real satellite data.

4. APPLICATIONS

In this section we reconstruct the WDF from observations made with the French Aureol-3 satellite [34]. The Aureol-3 satellite was launched on September 21, 1981, into a quasi-polar orbit with apogee at 2012 km and an orbital period of 1 h 50 min. The satellite carries wave experiments measuring three magnetic and two electric field components [35]. Here we concentrate on the magnetic field observations, and we use both real and simulated data.

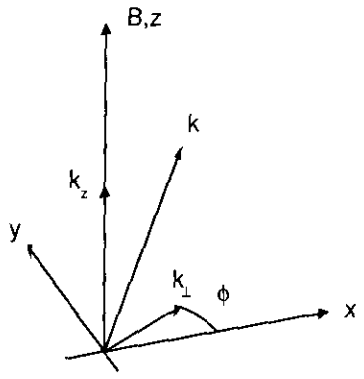

FIG. 3. Cylindrical wave vector coordinates k_z, $k_\perp$ and ϕ are often used when describing wave phenomena in magnetized plasmas. The z-axis is chosen in the direction of the background magnetic field.

4.1. *Solution of the Dispersion Relation*

In the geometry of magnetized plasmas it is natural to introduce cylindrical wave vector coordinates k_z, $k_\perp$, and ϕ, where k_z and $k_\perp$ are the wave vector components parallel and perpendicular to the background magnetic field, as shown in Fig. 3. To simplify analytical calculations it is often assumed that the gyration of plasma particles in the background magnetic field makes the dispersion relation independent of the azimuthal angle ϕ. The resulting dispersion relation can then be solved with numerical methods and displayed as dispersion surfaces [36–38]. Figure 4 shows such a dispersion surface for a plasma consisting of electrons, protons, and positively charged helium and oxygen ions. Here ρ_{H^+} is the proton gyroradius and Ω_{H^+} is the proton gyrofrequency. The model that we use is consistent with conditions in space near the poles at roughly 1500 km above the surface of the Earth.

The dispersion surface in Fig. 4 is merely a part of the complete solution of the dispersion relation. In principle, our reconstruction scheme can handle any number of dispersion surfaces. However, in this study we concentrate on a frequency range where the solution of the dispersion relation is single-valued, and therefore, only the surface in Fig. 4 needs to be included. Only $k_z > 0$ is shown in the figure, but an identical surface exists for $k_z < 0$. Waves with $k_z > 0$ and $k_\perp = 0$ correspond to waves propagating parallel to the background magnetic field. In the northern hemisphere these waves propagate towards the Earth. A distinction between downgoing ($k_z > 0$) and upgoing ($k_z < 0$) waves requires measurements of both electric and magnetic wave fields [39]. Here we reconstruct the WDF from magnetic field data only, and in such cases we know beforehand that the reconstructed WDF is symmetric with respect to $\mathbf{k}$ and $-\mathbf{k}$ [29]. Therefore, to keep computations to a minimum, we exclude $k_z < 0$ in the following reconstructions.

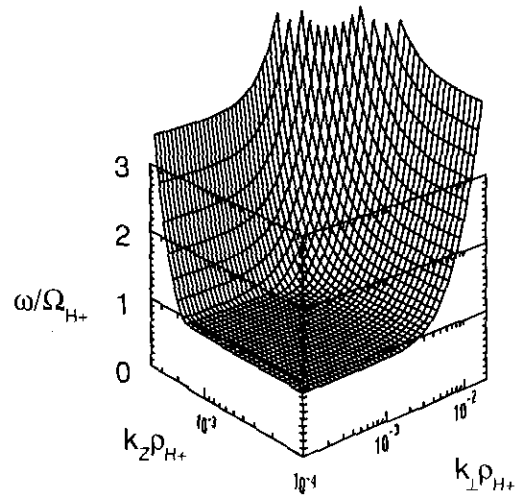

FIG. 4. The real part of the frequency plotted as a dispersion surface. The surface shows normalized frequency as a function of wave vector coordinates parallel and perpendicular to the background magnetic field. The wave vector coordinates are normalized to the inverse of the proton gyroradius ρ_{H^+}. To produce the surface we use a plasma model containing 15% protons, 5% He^+, and 80% O^+. The ions as well as the electrons are Maxwellian distributed in velocity with a temperature of 1 eV (11500 K). Together with a background magnetic field strength of 27.7 μT, this makes $\rho_{H^+} \approx 5$ m. The plasma density is 7190 electrons/cm^3.

4.2. *Synthetic Data*

We compute synthetic Aureol-3 magnetic field spectra $\tilde{\mathbf{C}}$ from a model WDF $\tilde{\psi}(\mathbf{k})$ and the relation

$$\tilde{\mathbf{C}} = \mathbf{S}(\tilde{\mathbf{x}}), \tag{45}$$

where $\tilde{x}_l = \tilde{\psi}(\mathbf{k}_l)\,\delta_l$. The model WDF is shown by contours of constant WDF in Figs. 5 and 6. Figure 5 shows the WDF in the plane perpendicular to the background magnetic field, where the model consists of a single peak at $\phi = 90°$. Figure 6 shows the model WDF in the $(k_\perp, k_z)$-plane projected onto the dispersion surface in Fig. 4.

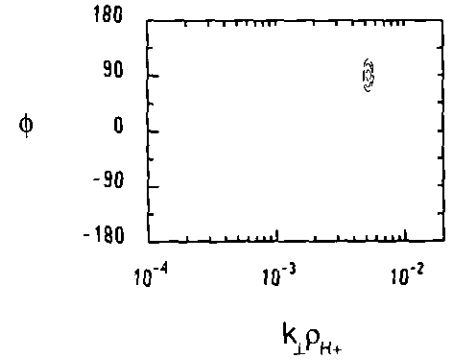

FIG. 5. Model WDF shown by contours of constant WDF in the plane perpendicular to the background magnetic field. The WDF varies by a factor of 1.41 between adjacent contour levels.

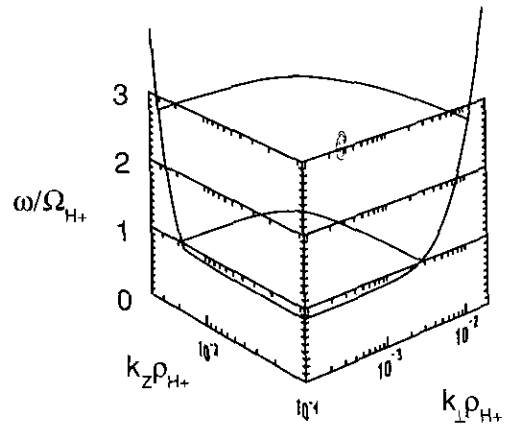

FIG. 6. Model WDF shown by constant WDF contours projected onto the dispersion surface in Fig. 4. The WDF changes by a factor of 1.41 between adjacent contour levels. The dispersion surface is indicated for $k_\perp = 0$ and $k_z = 0$, and the constant frequency contours running across the surface mark the frequency interval included in later reconstructions.

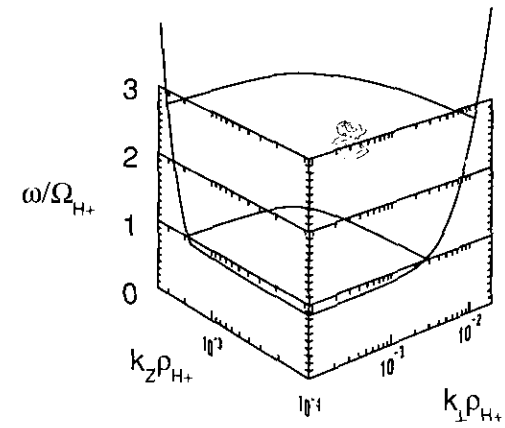

FIG. 8. Reconstructed WDF shown by constant WDF contours projected onto the dispersion surface in Fig. 4. There is a factor of 1.41 between adjacent contours. As in Fig. 6 the dispersion surface is indicated for $k_\perp = 0$ and $k_z = 0$, and the constant frequency contours running across the surface mark the frequency interval included in the reconstruction.

We use the plasma model described in the caption of Fig. 4 and compute the required kernels from the dispersion relation and from the geometry of the magnetic field antennas. A logarithmic resolution of 0.02 in the $(k_\perp, k_z)$-plane and an angular resolution of 45° makes $n = 23592$. The two constant frequency contours across the surface in Fig. 4 mark the included frequency range $[1.02, 2.80]\,\Omega_{H^+}$. We divide this frequency range into six intervals, each of width $0.3\Omega_{H^+}$. With three magnetic field components this amounts to $m = 54$ data points (three real power spectra and three independent complex cross spectra in each of the six frequency intervals). The spectral densities computed in the six frequency intervals are assumed to be afflicted with a 10% error, $\sigma_j = 0.1\tilde{C}_j$.

The WDF reconstructed from synthetic data is shown in Figs. 7 and 8. Besides some discretization effects visible in Fig. 8, the reconstructed WDF displays a broadening of the solution with respect to the model. This broadening comes from the M-E condition, which strives to maintain a flat solution while satisfying the constraint imposed by the data.

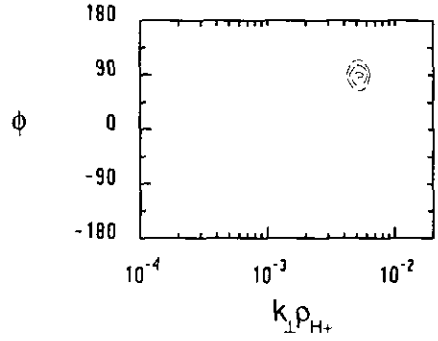

FIG. 7. WDF reconstructed from synthetic data generated by the model WDF shown in Figs. 5 and 6. The solution is shown by contours of constant WDF in the plane perpendicular to the background magnetic field. The WDF varies by a factor of 1.41 between the contour levels.

Except for the discretization effects and a slight spread of energy in **k**-space, the reconstruction reproduces the model accurately. Note that even though the dispersion relation is independent of the azimuthal angle ϕ, the integration kernels are not, and therefore, the reconstruction schemes places the reconstructed WDF in the correct angular interval.

To solve this particular problem our scheme tried three different μ-values and roughly 10 different λ-values for each value of μ. The number of μ-iterations required depends on the initial guess. In this case we had already experimented with different μ-values, hence the small number of iterations. The total cpu time required for this job on a 33 MHz HP/APOLLO workstation was nearly 9 h. The final value of β was roughly 10^{-6}.

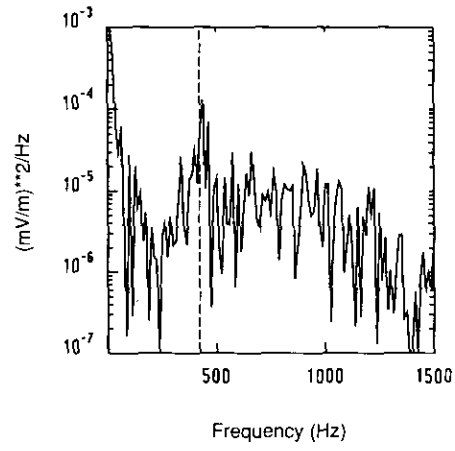

FIG. 9. Power spectral density for the BZ magnetic component measured with Aureol-3 at 08.47 (UT) on June 10, 1982. Coordinates: geomagnetic time 11:50, altitude 1612 km, geographic longitude 20.7°, and geographic latitude 65.83°. The spectrum is computed from 80 ms of data giving a frequency resolution of 12.5 Hz. The dashed vertical line at 422 Hz indicates $\Omega_{H^+}/2\pi$.

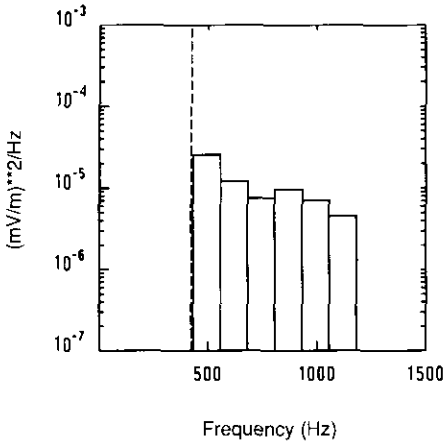

FIG. 10. Average power spectral density in six frequency intervals above the proton gyrofrequency. Each interval has a width of 125 Hz and the lowest interval starts at 431 Hz, just above $\Omega_{H^+}/2\pi$.

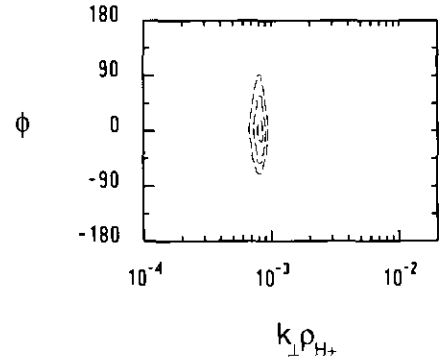

FIG. 11. WDF reconstructed from real Aureol-3 data. As before, the WDF varies by a factor of 1.41 between contour levels.

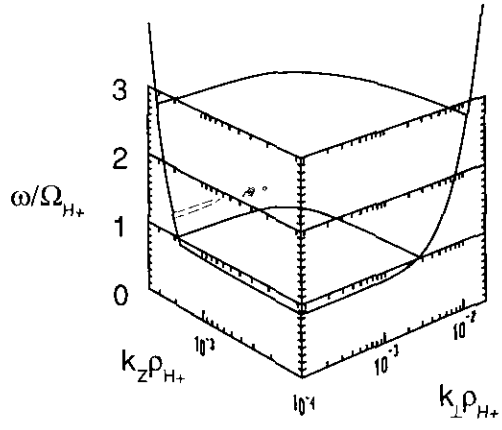

FIG. 12. Reconstructed WDF shown in the $(k_\perp, k_z)$-plane by constant WDF contours separated by a factor of 1.41. Real Aureol-3 data are used.

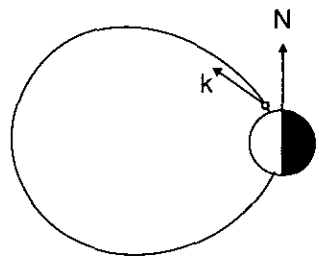

FIG. 13. Schematic picture of the Earth viewed in the geographic meridian plane. The figure shows the location of the Aureol-3 satellite on the dayside (small circle), and a magnetic field line passing through this point. The direction of the wave vector that we infer from the reconstructed WDF is indicated at the location of the satellite. Only the upgoing ($k_z < 0$) waves are represented here.

4.3. *Real Data*

Figure 9 shows a 120-point FFT spectrum of one of the magnetic field components observed on Aureol-3 on June 10, 1982. The observations were made on the day side near local noon at an altitude of 1612 km and a geographic latitude of 66°. The dashed vertical line at 422 Hz indicates $\Omega_{H^+}/2\pi$. The power spectrum in Fig. 9 is a highly structured function of frequency. However, we are not interested in resolving all these structures. Instead, we average the spectrum over frequency intervals of width 125 Hz. We concentrate on waves with angular frequencies above the proton gyrofrequency, and start from a frequency of 431 Hz. The resulting average spectral densities are shown in Fig. 10. When the frequency is normalized to $\Omega_{H^+}/2\pi$, these frequency intervals correspond to the six intervals used in the model problem earlier. In other respects we set up the reconstruction as in the model problem. Thus, we reconstruct the WDF over the dispersion surface in Fig. 4 with the same number of points in **k**-space, $n = 23592$, and with the same number of data points, $m = 54$. The standard deviations are assumed to be 10% of the spectral densities, and only magnetic field data are used.

The reconstructed WDF presented in Figs. 11 and 12 shows a distinct peak in **k**-space for $k_\perp \rho_{H^+} \approx 0.0008$, $k_z \rho_{H^+} \approx 0.002$, and $\phi \approx 0°$, with only a small spread of energy to lower and higher perpendicular wave numbers. Due to the $\pm\mathbf{k}$ symmetry we can say immediately that the reconstructed WDF contains an identical peak also at $k_\perp \rho_{H^+} \approx 0.0008$, $k_z \rho_{H^+} \approx -0.002$, and $\phi \approx 180°$. The wavelength, $2\pi/k$, is roughly 15 km, and the wave vector makes an angle of about 20° with the background magnetic field (see Fig. 13).

5. DISCUSSION

Before any reliable information can be obtained from a reconstruction, the solution must be validated in many ways. When new inversion techniques are introduced one

often sees a variety of tests performed on model problems. A successful reconstruction of the model distribution in such tests is assumed to justify the reconstruction scheme. This has been done also in the case of WDF reconstruction [3, 41]. However, the successful reconstruction in one case does not automatically guarantee reliable reconstructions in another. How accurately the reconstruction is depends strongly on the structure of the integration kernels. Uniform kernels give no resolution capability at all, while highly structured kernels may allow accurate reconstructions. In our case the kernels vary between different satellites and different dispersion surfaces in a way that we can hardly predict beforehand. Therefore, it is necessary to investigate relevant model problems to obtain a feeling for how the reconstruction scheme works in each given situation. For this reason reconstructions from synthetic data have a significant role to play in the analysis of real data. This should be true, not only in WDF reconstructions, but also in other cases. In our example in this paper, a sharply peaked model WDF is reproduced accurately by the reconstruction scheme. Even though more extensive tests are required before any far-reaching conclusions can be drawn, this gives credibility to the sharply peaked solution we obtain by reconstructing the WDF from real data.

The solution to an underdetermined problem must not depend strongly on small variations in the data, since the finer details in the data can be strongly affected by noise. When reconstructing the WDF, stability in this sense can be checked by investigating the effect on the WDF of small random variations in the data. To do this explicitly is rather laborious, and therefore, we normally test the solution by reconstructing the WDF with different choices of ξ_{bound} [3, 32]. Although this is not an equivalent procedure, it is much simpler and gives an indication of the stability of the solution with respect to noise in the data. Solutions obtained so far have proved to be stable to substantial variations in ξ_{bound}.

In addition, the solution should be stable to small perturbations in the plasma model. Satellite measurements of plasma parameters such as density, temperature, and ion composition are normally rather uncertain. For instance, the ion composition in the example in Section 4 is considered acceptable even though the raw data give 2% H^+, 1% He^+, and 97% O^+ [42]. To ensure stability, the WDF must be reconstructed with a representative selection of plasma models consistent with the observed plasma parameters. In this paper we present the results without performing either of the above tests since we do not draw any conclusions regarding the physics behind the data. Our purpose here is merely to present our reconstruction scheme.

Due to the fact that the inverse problem is underdetermined, we know by almost certainty that the WDF produced by our reconstruction scheme is not the "true" solution. It is of greatest importance that this fact is remembered, not only in WDF reconstruction, but when solving all types of ill-posed inversion problems. There are several ways of regularizing an inverse problem, and the solution we come up with may depend crucially on the particular regularization method that we use. When we reconstruct the WDF our aim is to keep the solution as flat and structureless as possible. The reason for this is that structures in the WDF are crucial for understanding the underlying physics. Therefore, we want the reconstructed WDF to reflect only such structures that are supported by the data, and not imposed by our particular reconstruction method. In practice, there is no way to guarantee this, but the least we should demand of our scheme is that when the data are consistent with a perfectly flat WDF, then this must also be the solution that we obtain.

One can devise several functionals that will produce a flat solution when such a solution is allowed by the data. As we have seen earlier, the entropy functional $H_2(f)$ in (19) has this property, and so has any functional of the form

$$H(f; p) = -\int dy \frac{f(y)}{F^p} \log\left[\frac{f(y)}{F\Delta(y)}\right], \qquad p \geq 0. \quad (46)$$

One of the simplest functions with the above property is the quadratic form

$$\sum_{i=1}^{n} \left(1 - \frac{x_i}{X\Delta_i}\right)^2. \quad (47)$$

One reason why we have not attempted to use this function when recontructing the WDF is that it does not automatically make the solution positive. However, it could be interesting to see how it compares to the entropy function. In this context one can mention the comparison made by Gull and Newton [40], which does not seem adequate since their quadratic function has its minimum at $x = 0$.

The entropy $H(\psi)$ in (17) seems to be one of the simplest functionals with the flat solution property that automatically produces a strictly positive solution. We think that these are two of the most important arguments for using entropy regularization in general, and our entropy functional in particular. For those interested in philosophical arguments, one can show that H is in reality the logarithm of the prior probability of a macrosignal defined by, for instance, Loredo [43], provided that the limit to a continuous distribution is taken properly. It is interesting to note that Loredo does not seem to realize that one of the two parameters in his prior probability is redundant and can be incorporated in the entropy as in $H(\psi)$.

Even if reconstruction with maximum-entropy seems to be one of the best and safest ways of resolving underdetermined inverse problems, we are still faced with the fact that

the solution we obtain is almost certainly not the true solution. This fact must always be taken into account when interpreting reconstructions of any kind. The value of a reconstructed WDF on its own lies in providing a test for consistency in **k**-space between possible physical generation mechanisms and the data. For instance, the reconstructed WDF presented in Section 4 is probably not consistent with any local generation mechanism. The fact that the M-E condition in some sense selects the smoothest solution consistent with the data indicates that also the true WDF *is inconsistent with any conceivable local generation* mechanism. This suggests that the waves have reached the satellite by propagating from the source region located somewhere else in the magnetosphere. The direction of the wave vector shown in Fig. 13 may then be taken as a starting point in a search for this source region. Without the reconstructed WDF such a search would be difficult.

Although the information that can be obtained from the reconstructed WDF on its own can be of great value, we feel that WDF reconstruction has its greatest merits when the solution can be combined successfully with other types of analysis techniques and observations that cannot easily be incorporated as prior information in the reconstruction. Oscarsson and Rönnmark [32] show one such example, where the reconstructed WDF can be combined consistently with other types of information, allowing a convincing interpretation of observations made by the Swedish Viking satellite.

The reconstruction method presented in this paper is developed for reconstructing the WDF. However, the only thing pertaining to the WDF case is the integration kernels. Since other inverse problems can be phrased in terms of integral equations similar to those we use, we believe that our method should be applicable in many other cases. The importance of dual principles is emphasized also by others [24], but the advantage with reducing the number of unknowns to the number of independent measurements should be obvious in cases with sparse data. The dual formulation employed in our scheme should be useful also when χ^2 constraints are used, together with the fixed normalization entropy $H_1(f)$. (In fact, setting $X=1$ in our scheme gives the traditional fixed normalization M-E solution.) However, we feel that the entropy function $H(\psi)$ that we use is the more appropriate one, not only in WDF reconstruction, but also in areas such as image reconstruction and computer tomography.

APPENDIX A

In order to exploit the possibility of reducing the dimension of the problem further, we form the matrix

$$\mathbf{B} = [\bar{\mathbf{A}}_1, \bar{\mathbf{A}}_2, ..., \bar{\mathbf{A}}_n], \tag{48}$$

where the kernels $\bar{\mathbf{A}}_l$ are treated as column vectors. Let n_r be the rank of **B**, and let the vectors $\mathbf{v}_i$, $i=1, ..., n_r$, form an orthonormal basis for the column space of **B**. The basis vectors can be obtained by diagonalizing $\mathbf{B}\mathbf{B}^T$. We can then write

$$\mathbf{r} = \mathbf{V}\mathbf{z}, \tag{49}$$

where

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{n_r}] \tag{50}$$

and

$$\mathbf{z} \in R^{n_r \times 1}. \tag{51}$$

Now we can re-write the dual problem as

$$\begin{aligned} \mathbf{z} &= \bar{\mathbf{C}}' - X \sum_{l=1}^{n} \frac{\delta_l}{V} \bar{\mathbf{A}}_l' \exp\left[\frac{1}{\mu} \mathbf{z}\bar{\mathbf{A}}_l'\right] \\ 1 &= \sum_{l=1}^{n} \frac{\delta_l}{V} \exp\left[\frac{1}{\mu} \mathbf{z}^T \bar{\mathbf{A}}_l'\right], \end{aligned} \tag{52}$$

where

$$\bar{\mathbf{C}}' = \mathbf{V}^T \bar{\mathbf{C}} \tag{53}$$

and

$$\bar{\mathbf{A}}_l' = \mathbf{V}^T \bar{\mathbf{A}}_l. \tag{54}$$

ACKNOWLEDGMENTS

We thank J. L. Rauch at the Laboratoire de Physique et Chimie de l'Environnement/CNRS in Orléans for his help with the Aureol-3 data. We have also benefited from discussions with Tommy Elfving and Mårten Gulliksson concerning the dual formulation of the reconstruction problem. Professor K. Rönnmark initiated and was deeply involved in the development of the original reconstruction scheme, and his continued support has been of great value. The cartoon of the magnetosphere was produced with the aid of computer programs provided by Kristof Stasiewicz.

REFERENCES

1. S. Ichimaru, *Basic Principles of Plasma Physics* (Benjamin, New York, 1973).
2. K. Rönnmark and J. Larsson, *J. Geophys. Res.* **93**, 1809 (1988).
3. T. Oscarsson and K. Rönnmark, *J. Geophys. Res.* **94**, 2417 (1989).
4. L. R. O. Storey and F. Lefeuvre, *Space Res.* **14**, 381 (1974).
5. C. R. Smith and W. T. Grandy, Jr. (Eds.), *Maximum-Entropy and Bayesian Methods in Inverse Problems* (Reidel, Dordrecht, 1985).
6. C. R. Smith and G. J. Erickson (Eds.), *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems* (Reidel, Dordrecht, 1987).

7. G. J. Erickson and C. R. Smith (Eds.), *Maximum-Entropy and Bayesian Methods in Science and Engineering. Volume 2. Applications* (Kluwer Academic, Dordrecht, 1988).
8. J. Skilling, *Maximum Entropy and Bayesian Methods*, edited by P. F. Fougère (Kluwer Academic, Dordrecht, 1990).
9. E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
10. E. T. Jaynes, *Phys. Rev.* **108**, 171 (1957).
11. E. T. Jaynes, *IEEE Trans. Systems Sci. Cybern.* **SSC-4**, 227 (1968).
12. N. S. Tzannes and P. A. Jonnard, *Opt. Eng.* **26**, 1077 (1987).
13. J. Skilling and S. F. Gull, in [5].
14. E. A. Robinson, in [5, p. 171].
15. N. J. Dusaussoy and I. E. Abdou, *IEEE Trans. Signal Process* **39**, 1164 (1991).
16. J. P. Burg, in [5].
17. A. Lippman, in [7].
18. T. Elfving, *Math. Comput. Modelling* **12**, 729 (1989).
19. J. E. Shore, in [6].
20. E. T. Jaynes, *Proc. IEEE* **70**, 939 (1982).
21. S. F. Brown, J.-F. Donati, D. E. Rees, and M. Semel, *Astron. Astrophys.* **250**, 463 (1991).
22. C. S. Fisk, *Trans. Res.-B* **22**, 69 (1988).
23. G. T. Herman, in [5].
24. G. T. Gullberg and D. N. Ghosh Roy, *J. Soc. Photo-Opt. Instrum Eng.* **671**, 25 (1986).
25. R. K. Bryan, in [6].
26. X. Zhuang, R. M. Haralick, and Y. Zhao, *IEEE Trans. Signal Process.* **39**, 1478 (1991).
27. J. J. Hu and F. H. Li, *Ultramicroscopy* **35**, 339 (1991).
28. A. Mohammad-Djafari and G. Demoment, in [7].
29. T. Oscarsson, Ph.D. thesis, University of Umeå, 1989.
30. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic Press, San Diego, CA, 1981).
31. T. Oscarsson and K. Rönnmark, *Proceedings, 1989 URSI International Symposium on Electromagnetic Theory* (The Royal Institute of Technology, Stockholm, 1989), p. 126.
32. T. Oscarsson and K. Rönnmark, *J. Geophys. Res.* **95**, 21187 (1990).
33. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, New York, 1986).
34. Yu. I. Galperin, H. Reme, C. Beghin, J. J. Berthelier, J. M. Bosqued, and B. E. Khmyrov, *Ann. Géophys.* **38**, 543 (1982).
35. J. J. Berthelier, F. Lefeuvre, M. M. Mogilevsky, O. A. Molchanov, Yu. I. Galperin, J. F. Karczewski, R. Ney, G. Gogly, C. Guerin, M. Leveque, J.-M. Moreau, and F. X. Sene, *Ann. Géophys.* **38**, 643 (1982).
36. K. Rönnmark, Rep. 179, Kiruna Geophys. Inst., Kiruna, Sweden, 1982.
37. K. Rönnmark, *Plasma Phys.* **25**, 699 (1983).
38. M. André, *J. Plasma Phys.* **33**, 1 (1985).
39. F. Lefeuvre, M. Parrot, and C. Delannoy, *J. Geophys. Res.* **86**, 2359 (1981).
40. S. F. Gull and T. J. Newton, *Appl. Opt.* **25**, 156 (1986).
41. F. Lefeuvre and C. Delannoy, *Ann. Télécommun.* **34**, 204 (1979).
42. F. Lefeuvre, J. L. Rauch, D. Lagoutte, J. J. Berthelier, and J. C. Cerisier, *J. Geophys. Res.* **97**, 10601 (1992).
43. T. J. Loredo, *Maximum Entropy and Bayesian Methods*, edited by P. F. Fougére (Kluwer Academic, Dordrecht, 1990), p. 134.