A Note on Capon's Minimum Variance Projection for Multi-Spacecraft Data Analysis

Yasuhito Narita [1,2]*

[1] Space Research Institute, Austrian Academy of Sciences, Graz, Austria, [2] Institut für Geophysik und Extraterrestrische Physik, Technische Universität Braunschweig, Braunschweig, Germany

Capon's minimum variance projection for the multi-point measurements is revisited using the method of likelihood function to derive the minimum variance projection and a simplified error estimate analytically. Theoretical construction of the minimum variance projection assumes a Gaussian form of the likelihood function and also regards the data covariance as a proxy of the noise covariance. The minimum variance projection is extended to the problem of two-spacecraft mode decomposition in the Mercury magnetosphere in which the magnetic field is a superposition of the constant field from the current sheet and the dipolar field from the planet. The extension of the Capon estimator (the data-variance projection) can identify the signal amplitudes of the different fields with a sufficient accuracy when the statistical averaging is properly done. The Capon estimator serves as a powerful analysis tool when the spatial resolution is limited to only a few points.

Keywords: adaptive filter theory, Capon estimator, multi-spacecraft data analysis, waves and turbulence, mode decomposition

1. CAPON'S MINIMUM VARIANCE PROJECTION

Minimum variance projection introduced by Capon [1] has a wide range of applications in the geophysical and space physical research fields whenever multi-point measurements are available.

Various formations are possible in the multi-point measurements (**Figure 1**): the THEMIS mission in a one-dimensional array aligned with a magnetic field line [2], the Swarm mission spanning a plane with three spacecraft [3, 4], and the Cluster mission [5] the MMS mission [6] forming a tetrahedron.

The projection works with various kinds of shape vectors (or models for the data) by minimizing the projection error without changing the amplitude of the signal amplitude. The minimum variance projection is, after Capon [1] or Haykin [7], obtained by imposing a constrained optimization:

$$\text{minimize} \quad \vec{w}\mathbf{R}\vec{w} \quad \text{subject to} \quad \vec{w}\vec{h} = 1, \qquad (1)$$

or, by formulating into a variational problem using the variation operator $\delta[\cdots]$ and the Lagrangian multiplier λ,

$$\delta\left[\vec{w}\mathbf{R}\vec{w} - \lambda\left(\vec{w}\vec{h} - 1\right)\right] = 0. \qquad (2)$$

Here $\vec{w}$ is the weight vector operating on the measurement covariance matrix $\mathbf{R} = \langle\vec{d}\vec{d}\rangle$, $\vec{d}$ the measurement data in a vectorial form, $\langle\cdots\rangle$ the operation of ensemble averaging, $\vec{h}$ the shape

vector. The problem (Equation 2) can analytically be solved. The solution is, blue by treating the shape vector as a complex-number vector,

$$\vec{w} = \frac{\mathbf{R}^{-1}\vec{h}}{\vec{h}^{\dagger}\mathbf{R}^{-1}\vec{h}}. \tag{3}$$

See, for example, Haykin [7] for the derivation. The projected power (squared signal amplitude) is

$$P = \vec{w}^{\dagger}\mathbf{R}\vec{w} = \left[\vec{h}^{\dagger}\mathbf{R}^{-1}\vec{h}\right]^{-1}. \tag{4}$$

A larger set of algorithms has so far been developed in the frame of direction-of-arrival estimation in the adaptive filter theory. Many algorithms are recently reviewed by Khmou et al. [8], including beamforming method, Bartlett method, Capon method, linear prediction method, maximum entropy method, Pisarenko harmonic decomposition, minimum norm, MUSIC algorithm, propagator method, and partial covariance matrix method. In the multi-spacecraft wave-field analysis, three methods are most relevant: beamforming method, Capon method, and MSR method [9].

- The beamforming method can easily be implemented to the data analysis, but on the other hand provides a lower resolution of the signal-to-noise ratio in the spectral analysis compared to the Capon method when only few data points are used in the analysis [10, 11]
- The Capon method is versatile in the multi-spacecraft data analysis because the method can be extended in various ways, e.g., to the vectorial data set (the wave telescope technique) [12, 13] to different field types (e.g., electric field and magnetic field used in the k-filtering technique) [14], and to the mode decomposition (this paper).
- The MSR method [11] uses both the Capon (or the wave telescope) method and the eigenvector-based method (the extended MUSIC algorithm), and provides an improved signal-to-noise ratio in the wavevector spectrum. The MSR method is optimized in the estimate of the total fluctuation energy (i.e., trace of the spectral density matrix) but not to the matrix elements. Also, the MSR method is constrained to the isotropic noise assumption.

So far, three shape vectors for the Capon method have successfully been applied to the multi-spacecraft data analysis blue in the field of space physics

1. Plane waves [1, 12, 14]:

$$h_i(k_x, k_y, k_z) = \exp\left[\mathrm{i}\left(k_x r_{xi} + k_y r_{yi} + k_z r_{zi}\right)\right], \tag{5}$$

where $\{k_x, k_y, k_z\}$ are the three components of the wavevectors, $\{r_{xi}, r_{yi}, r_{zi}\}$ the spatial coordinates of the i-th sensor, and i the imaginary number unit.

2. Spherical waves [15, 16]:

$$h_i(k, r_{xc}, r_{yc}, r_{zc}) = \left[\sum_{i=1}^{n}\frac{1}{|\vec{r}_i - \vec{r}_c|^2}\right]^{-1/2}\frac{\exp\left[\mathrm{i}k|\vec{r}_i - \vec{r}_c|\right]}{|\vec{r}_i - \vec{r}_c|}, \tag{6}$$

where k is the wavenumber, and $\vec{r}_c = [r_{xc}, r_{yc}, r_{zc}]$ the coordinate of the spherical wave center. The summation in the normalization constant runs over the number of sensors n.

3. Phase-shifted waves [17]:

$$h_i(\Phi, r_{xc}, \Delta x, k_y) = \exp\left[\mathrm{i}\left(\Phi\arctan\left(\frac{r_{xi} - r_{xc}}{\Delta x}\right) + k_y r_{yi}\right)\right], \tag{7}$$

where Φ is the amount of phase jump, r_{xc} the x-coordinate of the phase jump center, Δx the x-width (or range in x) of the phase jump around the center, and k_y the y-component of the wavenumber.

Capon's minimum variance projection uses the measurement data to guide the projection by minimizing the estimated power during the projection (in spirit of minimizing uncertainty) yet keeping the gain. We address the question here, "Why is the inversion of the data covariance matrix $\mathbf{R}^{-1}$ used in the Capon projection (Equation 3) and in the spectrum (Equation 4)?" In this paper, we offer an answer to this question, that is, the Capon method uses the measured data as a proxy of the noise property upon the optimization procedure. The method of the likelihood function is introduced to give an alternative and more instinctive derivation of Capon's method. Moreover, by revisiting Capon's minimum variance projection through the likelihood function method, it becomes clear that one may project the measurement data not only onto a single shape vector but also onto a multitude of shape vectors, which will enhance the capability of the multi-spacecraft data analysis.

It is worthwhile to note that the maximum likelihood and the minimum variance aspects of the Capon estimator are already covered in detail in the original paper by Capon et al. [18] for a seismic array problem. In this paper, the essence of the Capon estimator is reviewed (in section 2) and the estimator is extended to the problem of mode decomposition. As an application, a new method is constructed (in section 3) for two-spacecraft measurements in the magnetosphere to identify the magnetic field of the current sheet origin and the dipolar field from the planet, which is relevant to the BepiColombo mission.

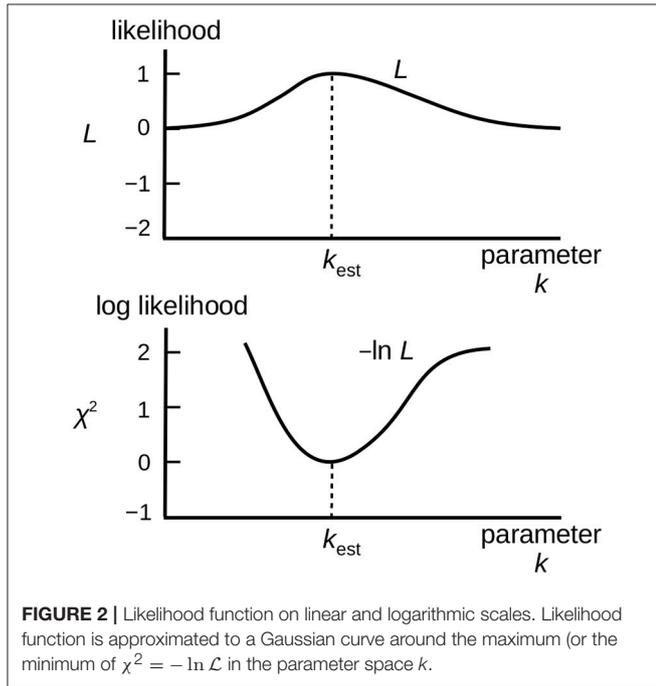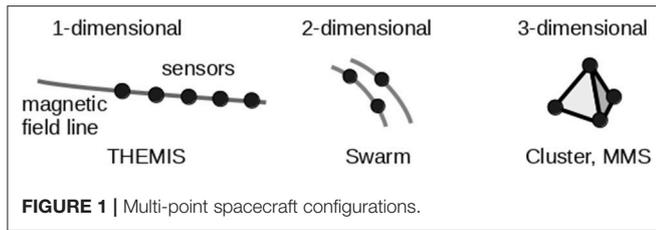## 2. A VIEW FROM LIKELIHOOD FUNCTION

2.1. Scalar Field

Minimum variance projection can be formulated using the likelihood function $\mathcal{L}$ as follows (see also [19]). Consider a model for the observational data as

$$d_i = h_i s + \eta_i, \tag{8}$$

where d_i is the data element at the i-th sensor ($i = \{1, 2, \cdots, n\}$, so n is the number of sensors), h_i the shape vector given a priori as a model, s the signal (in which we are interested), and η_i the noise at the i-th sensor. The noise is characterized by the $n \times n$ noise matrix $\mathbf{N}$

$$N_{ij} = \langle\eta_i\eta_j\rangle. \tag{9}$$

Again, the angular bracket $\langle\cdots\rangle$ denotes the ensemble averaging over different realizations. The averaging is important because

FIGURE 1 | Multi-point spacecraft configurations.

FIGURE 2 | Likelihood function on linear and logarithmic scales. Likelihood function is approximated to a Gaussian curve around the maximum (or the minimum of $\chi^2 = -\ln \mathcal{L}$ in the parameter space k.

otherwise the determinant of the matrix vanishes and the matrix cannot be inverted.

We minimize the squared deviation between the model and the data, i.e., we minimize χ^2 constructed (**Figure 2**) because of the steeper gradient toward the minimum (or the extremum) than that of the Gaussian curve as follows,

$$\chi^2 = \left\langle \sum_{j=1}^{n} \sum_{i=1}^{n} (d_i - h_i s) N_{ij}^{-1} (d_j - h_j s) \right\rangle \tag{10}$$

with respect to the signal s. We now assume a Gaussian shape for the likelihood function,

$$\mathcal{L} \propto \exp\left[-\frac{\chi^2}{2}\right]. \tag{11}$$

The likelihood function $\mathcal{L}$ represents the probability of finding the true signal value for s under the given data set d_i.

Our goal is to estimate the signal amplitude (denoted as $\tilde{s}$) for the shape vector for the given data set by finding a maximum of the likelihood function or equivalently by minimizing the deviation χ^2.

We differentiate χ^2 with respect to the signal s,

$$\frac{\partial \chi^2}{\partial s} = -2 \left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} \left(d_j - h_j s\right) \right\rangle. \tag{12}$$

Requiring that the derivative be zero (which corresponds to the extremum), $\partial(\chi^2)/\partial s$, we obtain

$$\left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} \left(d_j - h_j s\right) \right\rangle = 0. \tag{13}$$

Equation (13) can be arranged into the following form,

$$\left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} h_j s \right\rangle = \left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} d_j \right\rangle. \tag{14}$$

Note that the part in the angular bracket on the left hand side of Equation (14) is already statistically evaluated, so one may leave the angular bracket out in the following discussion. For a scalar quantity of s, one may obtain the estimator $\tilde{s}$ as

$$\tilde{s} = \left[\sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} h_j \right]^{-1} \left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} d_j \right\rangle \tag{15}$$

$$= C_{\mathrm{N}} \left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} d_j \right\rangle, \tag{16}$$

where C_{N} is the noise covariance,

$$C_{\mathrm{N}} = \left[\sum_{i=1}^{n} \sum_{j=1}^{n} h_i N_{ij}^{-1} h_j \right]^{-1}. \tag{17}$$

Or, in a matrix notation,

$$\tilde{s} = C_{\mathrm{N}} \left\langle \vec{h} \mathbf{N}^{-1} \vec{d} \right\rangle \tag{18}$$

and

$$C_{\mathrm{N}} = \left(\mathbf{h} \mathbf{N}^{-1} \mathbf{h}\right)^{-1} \tag{19}$$

The signal covariance (or the optimized power estimate) is (by noting that the ensemble averaging is taken after the covariance calculation)

$$C_{\mathrm{R}} = \langle \tilde{s}\tilde{s} \rangle \tag{20}$$

$$= \left\langle C_{\mathrm{N}} \vec{h} \mathbf{N}^{-1} \vec{d}\vec{d} \left(\mathbf{N}^{-1}\right) \vec{h} C_{\mathrm{N}} \right\rangle \tag{21}$$

$$= C_{\mathrm{N}} \vec{h} \mathbf{N}^{-1} \left\langle \vec{d}\vec{d} \right\rangle \left(\mathbf{N}^{-1}\right) \vec{h} C_{\mathrm{N}}. \tag{22}$$

In general, the data covariance matrix $\mathbf{R} = \langle \vec{d}\vec{d} \rangle$ and the noise covariance matrix $\mathbf{N}$ must be evaluated separately, or the noise covariance matrix needs to be known a priori to determine the signal covariance C_{S}. Capon's minimum variance projection is

obtained by imposing (or using) the data covariance matrix as the noise covariance matrix, $\mathbf{N} \rightarrow \mathbf{R}$. An explicit calculation yields

$$C_R = C_N \vec{h} \mathbf{R}^{-1} \vec{h} C_N \tag{23}$$

$$= \left(\vec{h} \mathbf{R}^{-1} \vec{h} \right)^{-1} \vec{h} \mathbf{R}^{-1} \vec{h} \left(\vec{h} \mathbf{R}^{-1} \vec{h} \right)^{-1} \tag{24}$$

$$= \left(\vec{h} \mathbf{R}^{-1} \vec{h} \right)^{-1} \tag{25}$$

2.2. Vector Field

The method with the likelihood function can be extended to the vector field treatment in a straightforward fashion. We construct the data model as follows.

$$\vec{d} = \mathbf{H} \vec{s} + \boldsymbol{\eta}. \tag{26}$$

The data are arranged into a long vector, $\vec{d} = [d_{x1}, d_{y1}, d_{z1}, d_{x2}, d_{y2}, d_{z2}, \cdots d_{zN}]$. The signal amplitude vector is $\vec{s} = [s_x, s_y, s_z]$. $\mathbf{H}$ is the shape-and-pointing matrix.

$$\mathbf{H} = \begin{bmatrix} h_{x1} & 0 & 0 \\ 0 & h_{y1} & 0 \\ 0 & 0 & h_{z1} \\ h_{x2} & 0 & 0 \\ 0 & h_{y2} & 0 \\ 0 & 0 & h_{z2} \\ & \vdots & \\ 0 & 0 & h_{zN} \end{bmatrix}. \tag{27}$$

$\boldsymbol{\eta}$ is the noise with the same construction as that of $\vec{d}$ (with $3 \times n$ elements). The essential difference from the scalar-field treatment is that the covariance becomes a matrix, e.g., $\mathbf{C}_S = \langle \vec{s} \vec{s} \rangle$ for the signal covariance matrix. Repeating the same procedure (by taking care of the vector and matrix operations), the signal estimator for the m-th component ($k = \{x, y, z\}$) is obtained as

$$\tilde{s}_k = \sum_{\ell}^{x,y,z} \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{C}_N)_{\ell k} H_{i\ell} N_{ij}^{-1} d_j, \tag{28}$$

where the summation on ℓ runs over the x, y, and z components, and that on i and j over the number of sensors. The signal estimator is given in a matrix notation as

$$\tilde{\tilde{s}} = \mathbf{C}_N \left\langle \mathbf{H} \mathbf{N}^{-1} \vec{d} \right\rangle \tag{29}$$

The noise covariance matrix $\mathbf{C}_N$ is given by

$$\left(\mathbf{C}_N^{-1} \right)_{\ell k} = \sum_{i=1}^{n} \sum_{j=1}^{n} H_{i\ell} N_{ij}^{-1} H_{jk} \tag{30}$$

or, in a matrix notation,

$$\mathbf{C}_N = \left[\mathbf{H} \mathbf{N}^{-1} \mathbf{H} \right]^{-1}. \tag{31}$$

The signal covariance matrix, when using the data covariance matrix again as the noise covariance matrix, is

$$\mathbf{C}_R^{(vec)} = \mathbf{C}_N \mathbf{H} \mathbf{N}^{-1} \mathbf{R} \left(\mathbf{N}^{-1} \right)^t \mathbf{H} \mathbf{C}_N \tag{32}$$

$$\rightarrow \left[\mathbf{H} \mathbf{R}^{-1} \mathbf{H} \right]^{-1}. \tag{33}$$

2.3. Mode Decomposition

The minimum variance projection can be extended to multiple shape vectors. By doing so, it is possible to decompose the measurement data set into a spectrum of m different modes or shapes. We construct a model for the measurement data (scalar field) with a multitude of modes and a noise.

$$d_i = \sum_{\alpha=1}^{m} h_i^{\alpha} s^{\alpha} + \eta_i, \tag{34}$$

where d_i is the measured field at the i-th sensor, m the number of the modes introduced into the model, $\vec{h}^{\alpha} = \{\vec{h}^{(1)}, \vec{h}^{(2)}, \cdots, \vec{h}^{(m)}\}$ is a set of shape vectors for modes $\alpha = \{1, 2, \cdots, m\}$, the symbol s^{α} is the signal amplitude at each mode, and n_i the noise at the i-th sensor.

Derivation of the minimum variance estimator is essentially the same as that for the single mode χ^2 minimization. The estimator for the signal amplitude at the α-th mode is obtained as

$$\tilde{s}^{\alpha} = \left\langle \sum_{\beta}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{C}_N)^{\alpha\beta} h_i^{\beta} N_{ij}^{-1} d_j \right\rangle. \tag{35}$$

The noise covariance matrix is a projection of the inverse noise matrix $\mathbf{N}^{-1}$ onto different modes (h^{α} and h^{β}),

$$\left(\mathbf{C}_N^{-1} \right)^{\alpha\beta} = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i^{(\alpha)} N_{ij}^{-1} h_j^{\beta}. \tag{36}$$

The signal covariance matrix for the Capon-type projection is obtained as

$$(\mathbf{C})^{\alpha\beta} = \mathbf{C}_N \vec{h}^{\alpha} \mathbf{N}^{-1} \mathbf{R} \mathbf{N}^{-1} \vec{h}^{\beta} \mathbf{C}_N \tag{37}$$

$$\rightarrow \left[\vec{h}^{\alpha} \mathbf{R}^{-1} \vec{h}^{\beta} \right]^{-1}. \tag{38}$$

The diagonal elements in the signal covariance matrix are the power (squared signal amplitude) of each mode.

2.4. Error Estimate

A useful form for the one-sigma error (68 % confidence) for the minimum variance estimator can be evaluated from the likelihood function. Ideally, the inverse of the noise matrix $\mathbf{N}^{-1}$ must be known, but the error estimate requires the knowledge on the noise property. Still, we obtain an insight by considering a special case, that is, we model the noise matrix $\mathbf{N}$ as diagonal with a value of each diagonal element $C_S + C_N$ (which is a scalar). The likelihood function for a simplified error estimate is again modeled as Gaussian (see, e.g., Equation 11.21 in Dodelson [19]),

$$\mathcal{L} \propto \frac{1}{(C_S + C_N)^{n/2}} \exp \left[-\frac{1}{2} \frac{\left\langle \sum_{i=1}^{n} (d_i - h_i s)^2 \right\rangle}{C_S + C_N} \right] \tag{39}$$

One-sigma error is obtained as the second-order derivative of the logarithm of the likelihood function (which is essentially χ^2 in the Gaussian likelihood model):

$$\sigma_S = \left[-\frac{\partial^2}{\partial C_S^2} \left(\ln \mathcal{L} \right) \right]^{-1/2} \tag{40}$$

$$= \sqrt{\frac{2}{n}} (C_S + C_N) \tag{41}$$

$$\rightarrow \sqrt{\frac{8 C_S^2}{n}}. \tag{42}$$

Here Equation (40) is evaluated at the peak of the likelihood function to obtain (Equation 41), and the signal covariance C_S is used as a proxy of the noise covariance C_N for Capon's minimum variance projection in deriving (Equation 42). Thus, the simplified error of the Capon-estimated signal power is $\sigma^2 = 8 C_S^2 / n$. See **Appendix** for derivation of Equation (40).

3. APPLICATION

Mode decomposition using the minimum variance projection serves as a powerful analysis tool when the measurements are limited to only few spatial points. A test using a synthetic data set is presented with two-spacecraft measurements in the Mercury magnetosphere in view of the BepiColombo mission [20].

3.1. Setup

Magnetic fields are modeled as a superposition of two different fields (or modes), $B^{(a)}$ and $B^{(b)}$ and noise η in the magnetosphere at two sensor locations, $r_1 = 480$ km (planetary orbiter) and $r_2 = 590$ km (magnetospheric orbiter) above surface (at a radius of $R_s = 2,440$ km for the planetary center). The measurement data are thus

$$B(r_1) = B^{(a)}(r_1) + B^{(b)}(r_1) + \eta_1 \tag{43}$$

$$B(r_2) = B^{(a)}(r_2) + B^{(b)}(r_2) + \eta_2, \tag{44}$$

where the first mode is the magnetic field from the current sheet

$$B^{(a)} = s^{(a)} = \text{const.} \tag{45}$$

and the second mode is the dipolar field from the planet

$$B^{(b)} = s^{(b)} \left(\frac{r}{R_s} \right)^{-3}. \tag{46}$$

Measurements are assumed to be on the magnetic equatorial plane such that the magnetic fields have only one non-vanishing component (say, the z component). The geometrical configuration is illustrated in **Figure 3**. The true signal amplitudes are $s^{(a)} = 20$ nT and $s^{(b)} = 200$ nT, respectively. Noise is blue assumed to be Gaussian distributed with a standard deviation of $\sigma_n = 1$ nT. The goal of the numerical test using the synthetic data is to estimate the signal amplitudes for the two modes, $\tilde{s}^{(a)}$ and $\tilde{s}^{(b)}$ using the noise-variance minimization and the data-variance minimization.
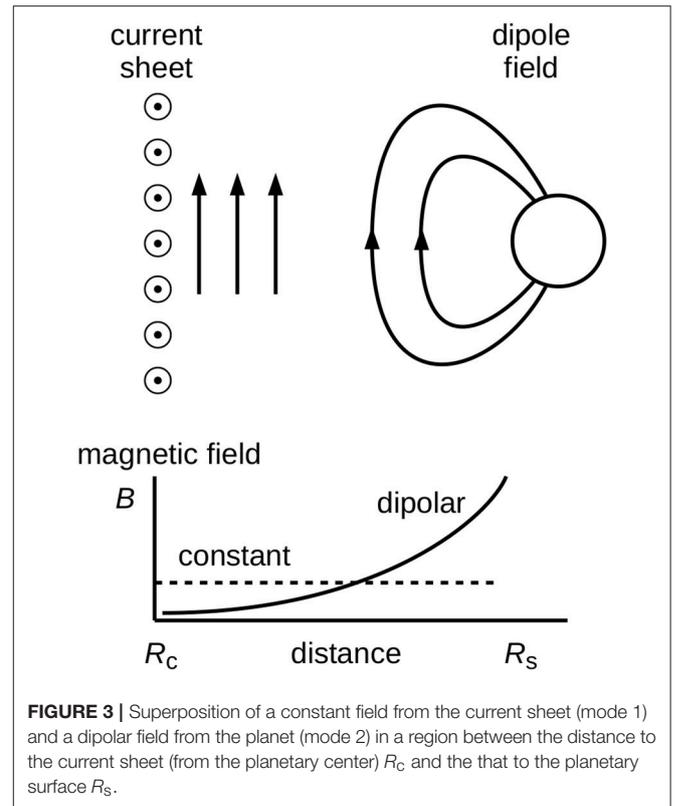

FIGURE 3 | Superposition of a constant field from the current sheet (mode 1) and a dipolar field from the planet (mode 2) in a region between the distance to the current sheet (from the planetary center) R_c and the that to the planetary surface R_s.

3.2. Analysis – Preparation

The shape vector for the first mode (a constant field) is

$$\vec{h}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{47}$$

and that for the second mode (decaying field) is

$$\vec{h}_2 = \begin{bmatrix} \left(\frac{r_1}{R_s} \right)^{-3} \\ \left(\frac{r_2}{R_s} \right)^{-3} \end{bmatrix}. \tag{48}$$

The shape matrix is constructed as $\left[\vec{h}_1 \vec{h}_2 \right]$,

$$\mathbf{h} = \begin{bmatrix} 1 & \left(\frac{r_1}{R_s} \right)^{-3} \\ 1 & \left(\frac{r_2}{R_s} \right)^{-3} \end{bmatrix}. \tag{49}$$

The data matrices are averaged over different realizations or samples,

$$\mathbf{R} = \langle \vec{d}\vec{d} \rangle, \tag{50}$$

using an averaging size of N_s. The measurement data vector is averaged when using the minimum variance estimator as $\langle \vec{d} \rangle$. We study the minimum variance estimators for different sampling

sizes N_s. The measurement noise matrix follows uncorrelated Gaussian statistics,

$$\mathbf{N} = \langle \vec{\eta}\vec{\eta} \rangle = \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_n^2 \end{bmatrix}. \tag{51}$$

3.3. Analysis – Projection

The noise minimum variance estimator for each of the modes $\alpha = \{a, b\}$ is

$$\left(\tilde{\vec{s}}^{(\mathrm{nv})} \right)_\alpha = (\mathbf{C}_N)_{\alpha\beta} \, (\mathbf{h})_{\beta i} \left(\mathbf{N}^{-1} \right)_{ij} \langle \vec{d}_j \rangle. \tag{52}$$

The data minimum variance estimator for each of the modes $\alpha = \{a, b\}$ is obtained by replacing the measurement noise covariance N by the measurement data covariance $\mathbf{R} = \langle \vec{d}\vec{d} \rangle$ and also replacing the noise projection $\mathbf{C}_N$ by the data projection $\mathbf{C}_R$ as

$$\left(\tilde{\vec{s}}^{(\mathrm{dv})} \right)_\alpha = (\mathbf{C}_R)_{\alpha\beta} \, (\mathbf{h})_{\beta i} \left(\mathbf{R}^{-1} \right)_{ij} \langle \vec{d}_j \rangle. \tag{53}$$

The noise covariance matrix and the data covariance matrix are essentially the Capon projection, namely,

$$\mathbf{C}_N = \left[\mathbf{h}\mathbf{N}^{-1}\mathbf{h} \right]^{-1} \tag{54}$$

$$\mathbf{C}_R = \left[\mathbf{h}\mathbf{R}^{-1}\mathbf{h} \right]^{-1}. \tag{55}$$

The noise minimum variance estimator needs the knowledge on the noise property and does not require the measurement data themselves. Therefore, the noise minimum variance estimator can be applied without the presence of the data and may be useful when planning a measurement of an experiment. For an uncorrelated Gaussian noise statistics, the noise minimum variance estimator computes the mode amplitude as a linear combination. The data minimum variance estimator, in contrast, requires the data but not the knowledge on the noise property. The mode amplitude is computed in a non-linear fashion, i.e., the measurement weight for each sensor is influenced by the data.

3.4. Results

Signal amplitudes for the two modes are obtained using the noise variance projection (Equation 52) and the data variance projection (Equation 52), and are graphically displayed as a function of the averaging size (or the number of realizations or samples) in **Figures 4**, **5** together with the one-sigma errors. The both estimators can find the true signal amplitudes (20 nT for the mode 1 and 200 nT for the mode 2) within the error bar, and the analysis using a larger statistical sampling size is beneficial in reducing the error. Yet, the accuracy is by far improved in the data-variance projection. The error bar is about 10 nT for an averaging size of 10–100 and becomes only a few nT or better (cf. the noise amplitude is 1 nT) for an even larger averaging size in the data-variance projection. The estimated amplitudes sufficiently converge to the true values for averaging sizes above 10, too. In contrast, the noise-variance projection still exhibits random deviations of the signal amplitudes from the true values for a larger averaging size (for example, size of 1,000).
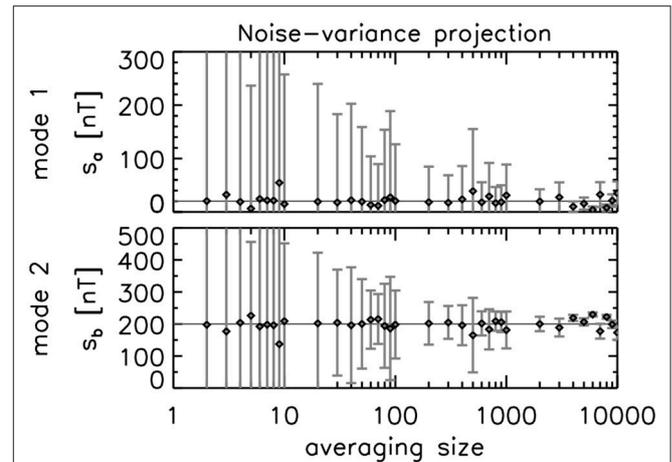

FIGURE 4 | Amplitudes for the mode 1 (constant field from the current sheet) and the mode 2 (dipolar field at the planetary surface) estimated by the noise-variance projection as a function of the averaging size.
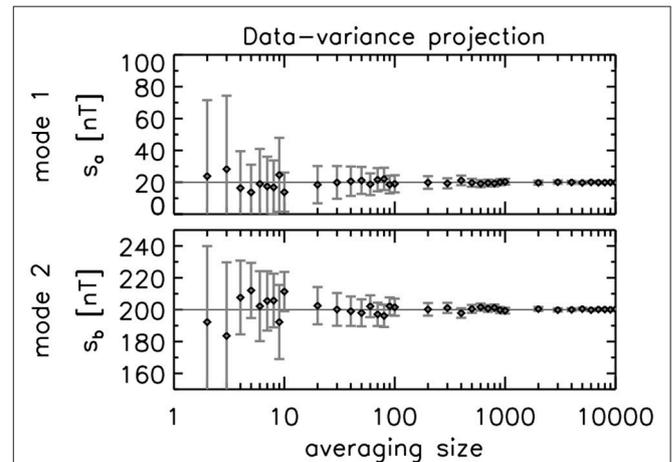

FIGURE 5 | Amplitudes for the mode 1 and the mode 2 (the same format as **Figure 4**) estimated by the data-variance projection.

4. OUTLOOK

Capon's projection is a useful tool when the noise property is unknown, and has a higher flexibility for various applications and extensions compared to the beamforming or the MSR methods. The extension of the minimum variance projection onto a multitude of shape vectors in the data opens the door to a decomposition method for the multi-point data. An application is presented for a decomposition of the multi-point data into a constant magnetic field and a dipolar field in view of the Mercury magnetosphere. It is comforting that even two-point measurements are capable of identifying the signal amplitudes when a sufficient amount of data is obtained for the proper averaging operation.

Another application is a decomposition into a set of orthogonal function basis. Capon estimator and its extension

to the mode decomposition can be applied to various wave fields in the solar wind, the foreshock, the magnetosheath, and the magnetotail regions as well as various static fields as far as the spatial structure can be properly modeled (such as the dayside magnetosphere as presented in this paper). For example, the source locator uses the lowest-order spherical Bessel function $j_0(x) = \sin(x)/x$ and the lowest-order Neumann function $n_0(x) = -j_{-1}(x) = -\cos(x)/x$. The expansion into a series of spherical Bessel functions or cylindrical functions (presumably with a cutoff) is a possible application to the multi-point measurements, e.g., identification or reconstruction of the spherical propagation or the vortical shape or motion.

Capon's minimum variance projection is not limited to the search for wave propagations or mode decomposition into different sources of the magnetic fields, but the method can be applied to solitary, spatially-localized structures. A useful example may be the KdV (Korteweg-de Vries) soliton (or ion-acoustic solitons in the case of plasmas) characterized by the following shape vector, $h_i(A, D, c, x_0) = A\,\mathrm{sech}^2\left[D(x_i - ct + x_0)\right]$ where c is the phase speed of the propagation, A the amplitude, and D the width of the soliton structure around the peak. The amplitude and the width are determined by the phase speed. For ion-acoustic solitons, the amplitude is given as $A = 3c$ and the width $D = \sqrt{c/2}$ [21].

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

REFERENCES

1. Capon J. High resolution frequency-wavenumber spectrum analysis. *Proc IEEE* (1969) **57**:1408–18. doi: 10.1109/PROC.1969.7278

2. Angelopoulos, V. The THEMIS mission. *Space Sci Rev.* (2008) **141**:5–34. doi: 10.1007/s11214-008-9336-1

3. Friis-Christensen E, Lühr H, Hulot G. Swarm: a constellation to study the Earth's magnetic field. *Earth Planets Space* (2006) **58**:351–8. doi: 10.1186/BF03351933

4. Friis-Christensen E, Lühr H, Knudsen D, Haagmans R. Swarm–an earth observation mission Investigating geospace. *Adv Space Res.* (2008) **41**:210–6. doi: 10.1016/j.asr.2006.10.008

5. Escoubet CP, Fehringer M, Goldstein, M. Introduction to the cluster mission, *Ann Geophys.* (2001) **19**:1197–200. doi: 10.5194/angeo-19-1197-2001

6. Burch JL, Moore TE, Torbert RB, Giles BL. Magnetospheric multiscale overview and science objectives. *Space Sci Rev.* (2016) **199**:5–21. doi: 10.1007/s11214-015-0164-9

7. Haykin S. *Adaptive Filter Theory*, 2nd. ed., *Prentice Hall Information and System Science Series*. Upper Saddle River, NJ: Prentice-Hall Inc. (1991).

8. Khmou Y, Safi S, Frikel M. Comparative study between several direction of arrival estimation methods. *J Telecomm Inform Tech.* (2014) **2014**:41–8.

9. Narita Y, Glassmeier KH, Motschmann U. Wave vector analysis methods using multi-point measurements. *Nonlin Processes Geophys.* (2010) **17**:383–94. doi: 10.5194/npg-17-383-2010

10. Motschmann U, Woodward TI, Glassmeier KH, Dunlop MW. Array signal processing techniques. In: Glassmeier KH, Motschmann U, Schmidt R. editors. *Proceedings of the Cluster Workshop on Data Analysis Tools*, Braunschweig (1995). ESA SP-371, p. 79–86. p. 28–3.

11. Narita Y, Glassmeier KH, Motschmann U. High-resolution wave number spectrum using multi-point measurements in space–the multi-point signal resonator (MSR) technique. *Ann Geophys.* (2011) **29**:351–60. doi: 10.5194-angeo-29-351-2011

12. Motschmann U, Woodward TI, Glassmeier KH, Southwood DJ, Pinçon J. Wavelength and direction filtering by magnetic measurements at satellite arrays: generalized minimum variance analysis. *J Geophys Res.* (1996) **101**:4961–6. doi: 10.1029/95JA03471

13. Glassmeier KH, Motschmann U, Dunlop M, Balogh A, Acuña MH, Carr C, et al. Cluster as a wave telescope–first results from the fluxgate magnetometer. *Ann Geophys.* (2001) **19**:1439–47. doi: 10.5194/angeo-19-1439-2001

14. Pinçon JL, Lefeuvre F. Local characterization of homogeneous turbulence in a space plasma from simultaneous measurements of field components at several points in space. *J Geophys Res.* (1991) **96**:1789–802. doi: 10.1029/90JA02183

15. Constantinescu OD, Glassmeier KH, Motschmann U, Treumann RA, Fornaçon KH, Fränz, M. Plasma wave source location using CLUSTER as a spherical wave telescope. *J Geophys Res Space Phys.* (2006) **111**:A09221. doi: 10.1029/2005JA011550

16. Constantinescu OD, Glassmeier KH, Décréau PME, Fränz M, Fornaçon KH. Low frequency wave sources in the outer magnetosphere, magnetosheath, and near Earth solar wind. *Ann Geophys.* (2007) **25**:2217–28. doi: 10.5194/angeo-25-2217-2007

17. Plaschke F, Glassmeier KH, Constantinescu OD, Mann IR, Milling DK, Motschmann U, et al. Statistical analysis of ground based magnetic field measurements with the field line resonance detector. *Ann Geophys.* (2008) **26**:3477–89. doi: 10.5194/angeo-26-3477-2008

18. Capon J, Greenfield RJ, Kolker RJ. Multidimensional maximul-likelihood processing of a large aperture seismic array. *Proc IEEE* (1967) **55**:192–211. doi: 10.1109/PROC.1967.5439

19. Dodelson S. *Modern Cosmology*. London: Academic Press (2003).

20. Benkhoff J, van Casteren J, Hayakawa H, Fujimoto M, Laakso H, Novara M, et al. BepiColombo–comprehensive exploration of mercury: mission overview and Science goals. *Planet Space Sci.* (2010) **58**:2–20. doi: 10.1016/j.pss.2009.09.020

21. Ichikawa YH, Watanabe S. Solitons, envelope solitons in collisionless plasmas. *J Phys Colloques* (1977) **38**:15–26. doi: 10.1051/jphyscol:1977603

APPENDIX: DERIVATIVE CALCULATION

Calculation for Equation 40 is as follows. The first-order derivative of the Gaussian error likelihood function (Equation 39) with respect to the variance C_S is obtained:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial C_R} &= \frac{n/2}{(C_R + C_N)^{n/2-1}} \exp\left[-\frac{1}{2} \frac{\left\langle \sum_{i=1}^{n}(d_i - h_i s)^2 \right\rangle}{C_R + C_N}\right] + \\
&\quad \frac{1}{2} \frac{1}{(C_R + C_N)^{n/2}} \frac{\left\langle \sum_{i=1}^{n}(d_i - h_i s)^2 \right\rangle}{(C_R + C_N)^2} \times \\
&\quad \exp\left[-\frac{1}{2} \frac{\left\langle \sum_{i=1}^{n}(d_i - h_i s)^2 \right\rangle}{C_R + C_N}\right] \qquad (A1) \\
&= \mathcal{L}\left[-\frac{n}{2(C_R + C_N)} + \frac{\left\langle \sum_{i=1}^{n}(d_i - h_i s)^2 \right\rangle}{2(C_R + C_N)^2}\right]. \qquad (A2)
\end{aligned}
$$

The second-order derivative of the logarithmic of the likelihood function is:

$$
\begin{aligned}
\frac{\partial^2 (\ln \mathcal{L})}{\partial C_R^2} &= \frac{\partial}{\partial C_R}\left[\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial C_R}\right] \qquad (A3) \\
&= \frac{\partial}{\partial C_R}\left[\frac{-n/2}{C_R + C_N} + \frac{1}{2} \frac{\left\langle \sum_{i=1}^{n}(d_i - h_i s)^2 \right\rangle}{(C_R + C_N)^2}\right] \qquad (A4) \\
&= \frac{n}{2} \frac{1}{(C_R + C_N)^2} - \frac{n}{(C_R + C_N)^2} \qquad (A5) \\
&= \frac{n}{2} \frac{1}{(C_R + C_N)^2}. \qquad (A6)
\end{aligned}
$$

Here the first-order derivative evaluated at the peak of the likelihood function, $\partial \mathcal{L}/\partial C_R = 0$, i.e.,

$$
\left\langle \sum_{i=1}^{n}(d_i - h_i s)^2 \right\rangle = n(C_R + C_N) \qquad (A7)
$$

is used in deriving Equation (A5). The one-sigma error (Equation 40) is evaluated using Equation (A6).