Reviewer 1:

The paper Mathematical foundation of Capon's method for planetary magnetic field analysis provides the underlying formalism for applying Capon's method to planetary magnetic fields and illustrates it with simulated data relevant for BepiColombo mission. While this is a valuable contribution to the field, a number of points need to be addressed before publication.

1. Extension of Capon's method to planetary magnetic fields. As indicated in the first and last paras of Section 3, L73–75 and L202–206, the paper generalizes the Capon method, previously used for the analysis of wave data. This is a major result that could be emphasized better, perhaps in a separate discussion section. This section could include a closer analysis of the case in the paper as compared to the wave case, by referring to Motschmann et al. (1996).
The discussion section could also detail the key principle(s) underlying the method, like maximum likelihood / minimum variance, along the line of Narita (2019). The divergence free feature of the magnetic field could be discussed on top, as in Motschmann et al. (1996).

Reply: Agreed. We added a section dedicated to the connection of the Capon method to other inversion methods (e.g., maximum likelihood, least square fit) and highlighted the difference from the wave analysis method (Motschmann et al., 1996) on p. 16. We also added a discussion about the divergence-free nature of the magnetic field (p. 2, ll. 45--52).

Related:
• L166-168: I am not sure I understand the text here, even though I could essentially follow Eqs. 30 to 41. Eventually, the underlying model makes the main contribution to the data (e.g., in the test application of Section 5). I guess this 'minimal contribution' has rather the meaning of Motschmann et al. (1996), where the filter w absorbs all the energy not associated with k (here not associated with the parametrized field) and leaves the part related to k undistorted (here the part related to the parametrized field). Same issue at L256.

Reply: Surely it is possible that the underlying model makes the main contribution to the data. But the distribution of the parameterized and the non-parameterized parts is unknown. Therefore, we stay conservative and assume safely that a large part of data is influenced by the noise and non-parametrized signals. (Explanation added on p. 7, l. 182 f)

• L200-201: Is this expression derived by Narita (2019)? Or could be derived by further processing of the maximum likelihood estimator? (e.g., in the suggested discussion section?)

Reply: Yes, Narita (2019) derived the expression for Capon's estimator by regarding the likehood function as nearly Gaussian (particularly around the peak of the likelhood function).

2. Illustration & Validation. In view of upcoming BepiColombo data, the authors chose to illustrate the method with simulated observations of Mercury magnetic field. While this is certainly helpful to prepare BepiColombo, I wonder if it is also the best test bed for the method. Earth magnetic field is known much better, at various altitudes – such that the weight of the external field and its influence on the results could be analyzed too. Including an example at the Earth, or at least a brief discussion of this validation possibility, would be more than welcomed.

Reply: The application of Capon‘s method to the analysis of Mercury‘s magnetic field has been emphasized as an example. Surely, the test against the Earth' magnetic field would be a good alternative. But such an analysis would have the size of a separate paper and it could not be pressed in a paragraph. The simulated data have the advantage that the ideal solution is known and thus a doubtless evaluation is possible. (Discussion added on p. 15, ll. 336--339)

Regarding the test exercise of Section 5, Table 1 shows that the largest errors are associated with g21 and, to some extent, with h21 . Is this by chance, or related to some systematics?

Reply: The deviation of these coefficients is related to the underlying model and therefore it is systematic noise. (Discusson added on p. 15, ll. 352--354).

3. Technicalities. Considering the target audience of the journal, different to a good extent from the signal processing community, the mathematical language of the paper may prevent the optimal transmission of the message. Additional explanations may help, inserted in the text or collected in an Appendix – when detailing the math would perturb the flow too much:

• *L69-71: Please clarify this sentence, possibly including an example.*
Reply: For example, when the magnetic field data are known on a dense grid in the vicinity of the planet, the Gauss coefficients can be estimated via integration. But in the case of a limited data set those integrals cannot be evaluated. (Sentence added on p. 3, ll. 78--80)

• *L93-97: The intuitive introduction of the filter matrix w via Eq. 8 s a bit confusing, since eventually the non-parametrized (external) part of B does not show up in the g C formula, Eq. 41.*
Reply: Eq. 8 has been written down, since it is the first intuitive idea to solve the inverse problem. An explanation that the non-parameterized parts <v> are unknown and have to be truncated by w has been added on p. 4, l. 104.

• *Eqs. 9 and 10 fall pretty much out of the blue. The use of M and P becomes clear later, but some clarification would be good already at this point.*
Reply: Agreed. We added motivation on page 4, l. 109.

• L106-110: Please detail why the determinant vanishes (even though it may look straight), how does statistical average prevent this, how is statistical average achieved.
Reply: It is a mathematical nature. The calculation of vanishing determinant of the outer product is shown in Appendix A on page 18. The statistical average is achieved by averaging over several numbers of measurements (added on p. 5, l. 119)

• L127: Please indicate also the second order moments.
Reply: Maybe the structure of the sentence was confusing, but <g o g> are the second order moments. We changed the order of the sentence and the equation accordingly (p. 6, l. 140).

• Eq. 21: Please explain why 2<Hg> o <v> and not <Hg> o <v> + <v> o <Hg> (given that, in general, the external product does not commute).
Reply: The outer product commutates in this special case because <Hg> and <v> have the same dimension (added on p. 6, l. 156).

• Eq. 27 and L154: Please explain why this is not enough to uniquely determine w.
Reply: It is because the parts that have to be truncated are unknown. (added on p. 7, l. 174)

• L155-158: Feels confuse. As long as the filter matrix truncates the non-parametrized part, it is not clear why its contribution to the data matters, neither how 'this yields the following procedure'.
Reply: The contribution matters because the parts that have to be truncated are unknown. The sentence 'this yields the following procedure' has been deleted (p. 7, l. 173).

• L191: Why is tr P a convex function?
Reply: trP is a convex function since it is the sum of the quadratic averaged expansion coefficients and thus it follows the same structure as for example the function f(x)=x^2 (explanation added on p. 8, Eq. 41)

•Eqs. 42 and 43: Please detail what is meant by 'input' and 'output'. Regarding Eq. 43, is there an equation analogous to Eq. 23, to clarify the meaning of 'signal' and 'noise' also for output?
Reply: By input we mean the measured data. By output we mean the filtered data (added on p. 9, l. 227 and l. 230);
Yes, one can construct an equation analogous to Eq. 23, which results, when w^T(...)w is applied to this equation. This can be seen within SNR_o.

•Eq. 44: Please explain why this ratio is dominated by 1/trace.
Reply: The array gain is dominated by 1/trace, since P, H and v are given by the model and the data and do not depend on the method (w does). (explanation added on page 9, l. 235)

•L266-267: Please provide a brief demonstration.
Reply: The eigenvalues of M are calculated in the Appendix B (p. 18/19).

•L267-269: Please explain briefly what is this about.
Reply: Done. (p. 12, l. 288)

•L278: How is the 'compromise' quantified?
Reply: The compromise can be understood in the sense that trace reaches its minimal value under the condition that P_d is maximal. (added on p. 12, l.297/298)

•L278-282: This is quite opaque for those not familiar with signal processing and in particular with these techniques.
Reply: The basic idea of the Tikhonov regularisation is briefly explained on page 12, ll. 299--304.

4. Others
• L17-18: What is non-ideal orbits?
Reply: The sentence was adversely formulated. We changed it to „data sampled on single orbits" (p. 1, l. 18).

• L18: simulated Mercury magnetic field data
Reply: Formulation has been changed (p. 1, l. 18).

• L54: 'closing the void' => 'covering the range' ?
Reply: Done. (p. 3, l. 60)

• Eq. 54 is identical to Eq. 41.
Reply: Reference has been added. (p. 14, l. 312)

• L325: in => at
Reply: Done. (p. 15, l. 355)

• L352: In principle, one could analyze also the external field, if some model is adopted.
Reply: That's very true! Sentence has been added. (p. 15, l. 356)

<u>Reviewer 2:</u>

The manuscript presents well founded support for mathematical analysis of planetary magnetic field basing on experimental data. 1. Line 33: "N data points xi, i = 1, ...,N" and Line 102 "Q indicates the number of measurements". Are N and Q the same numbers? 2. Line 94: "fulfills in principle the resolution of Eq. (7) with respect to g." Did you mean solution of Eq.(7)? 3. As a usual practice, for validation of the model, experimental data are divided into two parts. The first one is used for selection/tuning of model parameters with the help of the various optimization algorithms. The second part provides verification of the model by means comparison of experimental data with the data predicted by the built model. It could be helpful to demonstrate such an approach here.

Reply:
1.) Q and N are not the same numbers. N is the number of spatial data points, whereas Q indicates the number of measurements at each of these data points (for example the number of flybys at each point). A comment has been added on p. 5, l. 113.

2.) Agreed, we modified the word „resolution" to „solution" (p. 4, l. 103).

3.) For the application of several inversion methods (e.g. machine learning) it is useful/necessary to divide the data into two parts. Capon's method does not require this segmentation. For example, each data set corresponds with an optimal diagonal loading parameter. Since this parameter depends on the measurements and on the underlying model, it has to be calculated for each data set individually. When the data and the model are known, for each data set the diagonal loading parameter is calculated with the measurements and the data points themselves and then Capon's estimator can be calculated directly.

<u>General changes in the manuscript:</u>

- Changes in the manuscript are marked with „latexdiff", i.e., added text is marked in blue and the old version of the formulation is crossed out and marked in red
- The position of changes that are related to Reviewer comments are directly stated at the reply.
- We added a section about the discussion of Capon's method on p. 16.
- Appendix A and B have been added on p. 18/19
- The literature list has been extended by:
  Tikhonov, A. N., Goncharsky, A., Stepanov, V. V., Yagola, A. G.: Numerical Methods for the Solution of Ill-Posed Problems, Springer Netherlands, 1995. ISBN 079233583X

Mathematical foundation of Capon's method for planetary magnetic field analysis

Simon Toepfer[1], Yasuhito Narita[2,3], Daniel Heyner[3], Patrick Kolhey[3], and Uwe Motschmann[1,4]

[1]Institut für Theoretische Physik, Technische Universität Braunschweig, Braunschweig, Germany
[2]Space Research Institute, Austrian Academy of Sciences, Graz, Austria
[3]Institut für Geophysik und extraterrestrische Physik, Technische Universität Braunschweig, Braunschweig, Germany
[4]DLR Institute of Planetary Research, Berlin, Germany

Correspondence: Simon Toepfer (s.toepfer@tu-braunschweig.de)

Abstract.

Minimum variance distortionless projection, the so-called Capon method, serves as a powerful and robust data analysis tool when working on various kinds of ill-posed inverse problems. The method has not only successfully been applied to multi-point wave and turbulence studies in the context of space plasma physics, but also is currently being considered as a technique to perform the multipole expansion of planetary magnetic fields from a limited data set, such as Mercury's magnetic field analysis. The practical application and limits of the Capon method are discussed in a rigorous fashion by formulating its linear-algebraic derivation in view of planetary magnetic field studies. Furthermore, the optimization of Capon's method by making use of diagonal loading is considered.

1 Introduction

Nonlinear and adaptive filter technique has a wider range of applications in geophysical and space science studies to find the most likely parameter set describing the measurement data or to decompose the data into a set of signals and noise. Above all, the minimum variance distortionless projection introduced by Capon (1969) (hereafter, Capon's method) has successfully been applied to multi-point data analyses for the waves, turbulence fields, and current sheets (Motschmann et al., 1996; Glassmeier et al., 2001; Narita et al., 2003, 2013; Contantinescu et al., 2006; Plaschke et al., 2008). The strength of Capon's method lies in the fact that the method performs a robust data fitting even when the spatial sampling or data amount is limited in the measurement (e.g., successfully applied to four-spacecraft data (Motschmann et al., 1996)). Capon's method is currently being considered for planetary magnetic field studies in which the data (i.e., magnetic field samples) are more limited (e.g., ~~non-ideal~~ data sampled on single orbits), and has recently been applied to ~~Mercury's~~ simulated Mercury magnetic field data in view of the BepiColombo mission (Toepfer et al., 2020).

From a theoretical point of view there exist several origins for the derivation of the method. The first derivation of Capon's method (Capon, 1969), constructed for the analysis of seismic waves, is based on the estimation of frequency-wavenumber spectra. Later on, this approach has been reformulated in terms of matrix algebra (Motschmann et al., 1996). In the light of mathematical statistics Capon's estimator can be regarded as a special case of the maximum likelihood estimator (Narita, 2019). In this work the linear-algebraic formulation of the method (Motschmann et al., 1996) with specific attention to the magnetic field analysis is extended and the application of diagonal loading is discussed to improve the quality of data analysis with more justified applications and limits.

2 Motivation of Capon's method

The analysis of planetary magnetic fields is of great interest and one of the main tasks in space science. Here we pay special attention to the analysis of Mercury's internal magnetic field which is one of the primary goals of the BepiColombo mission (Benkhoff et al., 2010). The magnetometer on board the Mercury Planetary Orbiter (MPO) (Glassmeier et al., 2010) measures the magnetic field vectors $\boldsymbol{b}^i = \left(b_x^i, b_y^i, b_z^i\right)^T \in \mathbb{R}^3$ at N data points $\boldsymbol{x}^i$, $i = 1, \ldots, N$ along the orbit in the vicinity of Mercury. The magnetic fields around Mercury are considered as a composition or superposition of internal fields generated by the dynamo process, crustal and induced fields, which are mainly dominated by dipole and quadrupole fields and external fields generated by the currents flowing in the magnetosphere. For Mercury the external fields contribute a significant amount to the total magnetic field within the magnetosphere (Anderson et al., 2011) and therefore, a robust method is required for seperating the internal fields from the total measured field. Yet, each component has to be properly modeled and parameterized when decomposing the field.

For example, when only data in current-free regions are analyzed, the planetary magnetic field is irrotational and can be parameterized via the Gauss representation (Gauss, 1839). Whithin these current-free regions, the internal magnetic field can be expressed as the gradient of a scalar potential Φ, which can be expanded into a set of basis functions. ~~Considering the expansion into spherical harmonics, the potential for the planetary dipole and quadrupole fields results in~~

$$\Phi(\boldsymbol{x}^i) = R_\mathrm{M} \sum_{l=1}^{2} \left(\frac{R_\mathrm{M}}{r^i}\right)^{l+1} \sum_{m=0}^{l} \left[g_l^m \cos(m\lambda^i) + h_l^m \sin(m\lambda^i)\right] P_l^m\left(\cos(\theta^i)\right),$$

Because of

$$\partial_{\boldsymbol{x}} \cdot \boldsymbol{B} = 0, \tag{1}$$

the scalar potential Φ has to satisfy

$$\partial_{\boldsymbol{x}}^2 \Phi = 0, \tag{2}$$

where $\partial_{\boldsymbol{x}}^2$ is the Laplacian. Choosing planetary centered coordinates with radius $r^i \in [R_\mathrm{M}, \infty)$, azimuth angle $\lambda^i \in [0, 2\pi]$ and polar angle $\theta^i \in [0, \pi]$ ~~are chosen.~~ , the solution of Eq. (2) is given by spherical harmonics, so that the potential results in

$$\Phi(\boldsymbol{x}^i) = R_\mathrm{M} \sum_{l=1}^{2} \left(\frac{R_\mathrm{M}}{r^i}\right)^{l+1} \sum_{m=0}^{l} \left[g_l^m \cos(m\lambda^i) + h_l^m \sin(m\lambda^i)\right] P_l^m\left(\cos(\theta^i)\right) \tag{3}$$

for the planetary dipole and quadrupole fields. Here, R_{M} indicates the radius of Mercury and P_l^m are the Schmidt-normalized associated Legendre polynomials of degree l and order m. Within this series expansion there occurs a set of expansion coefficients g_l^m and h_l^m, named internal Gauss coefficients. By constructing the Gauss coefficients in a vectorial fashion and defining the true coefficient vector as $\boldsymbol{g} = \left(g_1^0, g_1^1, h_1^1, g_2^0, g_2^1, h_2^1, g_2^2, h_2^2\right)^T$, the magnetic field vectors at every data point $\boldsymbol{x}^i$ and the expansion coefficients are related via

$$\boldsymbol{b}^i = -\partial_{\boldsymbol{x}^i}\Phi(\boldsymbol{x}^i) + \boldsymbol{v}^i + \boldsymbol{n}^i \tag{4}$$

$$= \mathbf{H}^i\boldsymbol{g} + \boldsymbol{v}^i + \boldsymbol{n}^i, \tag{5}$$

where the terms of the series expansion are arranged in the matrix $\mathbf{H}^i\left(r^i, \theta^i, \lambda^i\right)$, called shape matrix. The vector $\boldsymbol{v}^i$ describes the parts that are not parameterized by the underlying model, e.g. the external parts, ~~closing the void~~ covering the range between the parameterized field and the measurement noise of the sensors, which is symbolized by the vector $\boldsymbol{n}^i$. Especially, the measurement noise is neither correlated with the parameterized part $\mathbf{H}^i\boldsymbol{g}$ nor with the non-parameterized part $\boldsymbol{v}^i$.

Summarizing the magnetic field measurements for all N data points into a vector $\boldsymbol{B} = \left(\left(\boldsymbol{b}^1\right)^T \ldots \left(\boldsymbol{b}^N\right)^T\right)^T \in \mathbb{R}^{3N}$, the field can be written as

$$\boldsymbol{B} = \mathbf{H}\,\boldsymbol{g} + \boldsymbol{v} + \boldsymbol{n}, \tag{6}$$

where $\boldsymbol{n} = \left(\boldsymbol{n}^1 \ldots \boldsymbol{n}^N\right)^T \in \mathbb{R}^{3N}$, $\boldsymbol{v} = \left(\boldsymbol{v}^1 \ldots \boldsymbol{v}^N\right)^T \in \mathbb{R}^{3N}$ and $\mathbf{H} = \left[\mathbf{H}^1 \ldots \mathbf{H}^N\right]^T \in \mathbb{R}^{3N\times G}$. G indicates the number of expansion coefficients.

Within Eq. (6) the magnetic field vector $\boldsymbol{B}$ and the shape matrix $\mathbf{H}$, given by the underlying model, are known. The coefficient vector $\boldsymbol{g}$ is to be determined by data fitting. Since in most applications the number of known magnetic data points is much larger than the number of wanted expansion coefficients ($G \ll 3N$), $\mathbf{H}$ is a rectangular matrix in general. Furthermore, the non-parameterized parts of the field and the noise are unknown. Therefore, the direct inversion of Eq. (6) is impossible and $\boldsymbol{g}$ has to be estimated. In this case, Capon's method establishes a robust and useful tool to find the estimated solution for the expansion coefficients in Eq. (6).

Since the method does not require the orthogonality of the basis functions, it has a wider range of applications when decomposing the measured data into a set of superposed signals, especially when the number of data points is limited. For example, when the magnetic field data are measured on a dense grid in the vicinity of the planet, the Gauss coefficients can be estimated via integration of the data. But in the case of a limited data set those integrals cannot be evaluated.

3 Derivation of Capon's method

The following derivation of Capon's method is based on the linear-algebraic formulation (Motschmann et al., 1996) which was formerly applied to the analysis of plasma waves in the terrestrial magnetosphere. Now we are focussing on the analysis of planetary magnetic fields.

As illustrated in the previous section, the magnetic field $\boldsymbol{b}^i$ measured at data point $\boldsymbol{x}^i$ in the vicinity of Mercury and the wanted expansion coefficients $\boldsymbol{g}$ are related via

$$\boldsymbol{b}^i = \mathbf{H}^i \boldsymbol{g} + \boldsymbol{v}^i + \boldsymbol{n}^i \tag{7}$$

or in a compact form for all N data points

$$\boldsymbol{B} = \mathbf{H}\, \boldsymbol{g} + \boldsymbol{v} + \boldsymbol{n}, \tag{8}$$

where the shape matrix $\mathbf{H}$ describes the underlying model.

For every data point $\boldsymbol{x}^i$ the noise vector $\boldsymbol{n}^i$ is assumed to be Gaussian with variance σ_n and zero mean, so that $\langle \boldsymbol{n}^i \rangle = 0$ and $\langle \boldsymbol{n}^i \circ \boldsymbol{n}^i \rangle = \sigma_n^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The angular brackets indicate averaging over an ensemble, e.g., different samples, realizations or measurements. Therefore, $\langle \boldsymbol{b}^i \rangle = \mathbf{H}^i\, \boldsymbol{g} + \langle \boldsymbol{v}^i \rangle$ holds and equivalently

$$\langle \boldsymbol{B} \rangle = \mathbf{H}\, \boldsymbol{g} + \langle \boldsymbol{v} \rangle, \tag{9}$$

since the model $\mathbf{H}$ and the true coefficient vector $\boldsymbol{g}$ are not affected by the averaging.

Because $\mathbf{H}$ is not always a square matrix but is in general a rectangular matrix with different sizes between rows and columns, the direct inversion of Eq. (9) is not guaranteed. Let us ignore for the moment the non-existence of $\mathbf{H}^{-1}$ and write down the equation

$$\boldsymbol{g} = \mathbf{H}^{-1} \left(\langle \boldsymbol{B} \rangle - \langle \boldsymbol{v} \rangle \right). \tag{10}$$

Despite its simplicity it is obviously incorrect. As $\mathbf{H}^{-1}$ does not exist, let us look for another matrix $\mathbf{w}$ called filter matrix, which follows the structure of this equation and fulfills in principle the ~~resolution~~ solution of Eq. (9) with respect to $\boldsymbol{g}$. Furthermore, the non-parameterized parts $\langle \boldsymbol{v} \rangle$ are unknown and therefore, it is desirable to truncate these parts by the filter matrix. Capon's method is just the procedure to construct the filter matrix $\mathbf{w}$ and to calculate or better to say estimate $\boldsymbol{g}$. To do so, some helpful quantities are introduced. In order to distinguish between the true coefficient vector $\boldsymbol{g}$ and the estimated solution, in the following Capon's estimator will be symbolized by $\boldsymbol{g}_C$.

For the inversion of Eq. (6) ~~the~~ it is useful to rewrite the vectors $\boldsymbol{B}$ and $\boldsymbol{g}_C$ in terms of a matrix representation. Thus, the data covariance matrix $\mathbf{M}$ and the coefficient matrix $\mathbf{P}$ are introduced as follows:

$$\mathbf{M} = \langle \boldsymbol{B} \circ \boldsymbol{B} \rangle = \frac{1}{Q} \sum_{\alpha=1}^{Q} \boldsymbol{B}^\alpha \circ \boldsymbol{B}^\alpha \in \mathbb{R}^{3N \times 3N} \tag{11}$$

$$\mathbf{P} = \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle \in \mathbb{R}^{G \times G} \tag{12}$$

Here Q indicates the number of measurements at each data point x^i, for example the number of flybys at each data point and the circle '$\circ$' symbolizes the outer product, which is defined by

$$\boldsymbol{x} \circ \boldsymbol{y} = \mathbf{x} \cdot \mathbf{y}^\dagger \in \mathbb{R}^{n \times m} \tag{13}$$

for any pair of vectors $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$. The dagger $\dagger$ indicates the Hermitian conjugate and the dot stands for the multiplication of the matrices $\mathbf{x} \in \mathbb{R}^{n \times 1}$ and $\mathbf{y}^\dagger \in \mathbb{R}^{1 \times m}$. Therefore, the diagonal of the matrix $\mathbf{P}$ contains the quadratic averaged components of the wanted estimator. It is important to note here that the covariance matrix $\mathbf{M}$ must be statistically averaged ~~,~~ ~~otherwise~~ over several numbers of measurements. Otherwise the matrix is singular with a vanishing determinant (cf. Appendix A) and the further analysis cannot be achieved.

If the non-invertibility of the matrix $\boldsymbol{B}^\alpha \circ \boldsymbol{B}^\alpha$ is neglected, for every measurement $\alpha = 1, \ldots, Q$ an estimator $\boldsymbol{g}_C^\alpha$ for the true coefficient vector $\boldsymbol{g}$ can be determined, so that

$$\boldsymbol{B}^\alpha = \mathbf{H}\, \boldsymbol{g}_C^\alpha + \boldsymbol{v}^\alpha + \boldsymbol{n}^\alpha \tag{14}$$

is valid. Thereby, each estimator deviates from the true coefficient vector by an error vector $\boldsymbol{\varepsilon}^\alpha = \boldsymbol{g} - \boldsymbol{g}_C^\alpha$. Note that because of the non-invertibility of the matrix $\boldsymbol{B}^\alpha \circ \boldsymbol{B}^\alpha$ the single estimator $\boldsymbol{g}_C^\alpha$ cannot be calculated. Since the invertibility is solely given by the averaging over Q measurements, only the averaged estimator

$$\boldsymbol{g}_C = \frac{1}{Q} \sum_{\alpha=1}^{Q} \boldsymbol{g}_C^\alpha, \tag{15}$$

with its related error

$$\langle \boldsymbol{\varepsilon} \rangle = \frac{1}{Q} \sum_{\alpha=1}^{Q} \boldsymbol{\varepsilon}^\alpha \tag{16}$$

is available.

In contrast to the estimator, the true coefficient vector is a theoretical given vector, that is not affected by the averaging ($\boldsymbol{g} \equiv \langle \boldsymbol{g} \rangle$, Eq. (9)). This property directly links the estimator to the true coefficient vector, which can be rewritten as

$$\boldsymbol{g} = \boldsymbol{g}_C^\alpha + \boldsymbol{\varepsilon}^\alpha. \tag{17}$$

Averaging over Q measurements and using $\boldsymbol{g} \equiv \langle \boldsymbol{g} \rangle$ results in

$$\boldsymbol{g} = \boldsymbol{g}_C + \langle \boldsymbol{\varepsilon} \rangle \tag{18}$$

and

$$\boldsymbol{g} \circ \boldsymbol{g} = \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle + \langle \boldsymbol{g}_C \circ \boldsymbol{\varepsilon} \rangle + \langle \boldsymbol{\varepsilon} \circ \boldsymbol{g}_C \rangle + \langle \boldsymbol{\varepsilon} \circ \boldsymbol{\varepsilon} \rangle$$

analogously for the second order moments

$$\boldsymbol{g} \circ \boldsymbol{g} = \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle + \langle \boldsymbol{g}_C \circ \boldsymbol{\varepsilon} \rangle + \langle \boldsymbol{\varepsilon} \circ \boldsymbol{g}_C \rangle + \langle \boldsymbol{\varepsilon} \circ \boldsymbol{\varepsilon} \rangle. \tag{19}$$

In the limit of vanishing errors $\langle \boldsymbol{\varepsilon} \rangle \to 0$, $\langle \boldsymbol{\varepsilon} \circ \boldsymbol{\varepsilon} \rangle \to 0$ respectively, Capon's estimator converges to the true coefficient vector

$$\boldsymbol{g}_C \to \boldsymbol{g} \tag{20}$$

and therefore

$$\langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle \to \boldsymbol{g} \circ \boldsymbol{g}. \tag{21}$$

For the further evaluation of Eq. (6) the definition of the outer product (Eq. 13) is utilized. Matrix multiplication of Eq. (6) with its Hermitian adjoint and averaging yields

$$\langle \boldsymbol{B} \cdot \boldsymbol{B}^\dagger \rangle = \langle (\mathbf{H}\,\boldsymbol{g} + \boldsymbol{v} + \boldsymbol{n}) \cdot (\mathbf{H}\,\boldsymbol{g} + \boldsymbol{v} + \boldsymbol{n})^\dagger \rangle \tag{22}$$

and therefore

$$\langle \boldsymbol{B} \circ \boldsymbol{B} \rangle = \mathbf{H} \cdot \boldsymbol{g} \circ \boldsymbol{g} \cdot \mathbf{H}^\dagger + 2\,(\mathbf{H}\boldsymbol{g}) \circ \langle \boldsymbol{v} \rangle + \langle \boldsymbol{v} \circ \boldsymbol{v} \rangle + \langle \boldsymbol{n} \circ \boldsymbol{n} \rangle \qquad , \tag{23}$$

assuming, that $\boldsymbol{n}$ is Gaussian with variance σ_n and zero mean ($\langle \boldsymbol{n} \rangle = 0$). By means of the limit $\langle \boldsymbol{\varepsilon} \rangle \to 0$ in Eq. (21) the unknown matrix $\boldsymbol{g} \circ \boldsymbol{g}$ and the true coefficient vector $\boldsymbol{g}$ can be approximated by Capon's estimator, resulting in

$$\langle \boldsymbol{B} \circ \boldsymbol{B} \rangle = \mathbf{H} \cdot \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle \cdot \mathbf{H}^\dagger + 2\,(\mathbf{H}\,\boldsymbol{g}_C) \circ \langle \boldsymbol{v} \rangle + \langle \boldsymbol{v} \circ \boldsymbol{v} \rangle + \langle \boldsymbol{n} \circ \boldsymbol{n} \rangle. \tag{24}$$

Taking into account that $\langle \boldsymbol{n} \circ \boldsymbol{n} \rangle = \sigma_n^2 \mathbf{I}$ and using the above defined abbreviations, Eq. (24) can be rewritten as

$$\mathbf{M} = \mathbf{H}\,\mathbf{P}\,\mathbf{H}^\dagger + 2\,(\mathbf{H}\,\boldsymbol{g}_C) \circ \langle \boldsymbol{v} \rangle + \langle \boldsymbol{v} \circ \boldsymbol{v} \rangle + \sigma_n^2 \mathbf{I}, \tag{25}$$

because $\mathbf{H}\,\boldsymbol{g}_C$ and $\langle \boldsymbol{v} \rangle$ have the same dimension, so that the outer product commutates.

Since Eq. (25) cannot be directly solved for $\mathbf{P}$, the goal is to find the best estimator $\boldsymbol{g}_C$ for $\boldsymbol{g}$, so that $\mathbf{P} = \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle$ is obtained as an approximate solution of Eq. (25). Therefore, a filter matrix $\mathbf{w}$ is constructed, that separates the parameterized field from the noise by projecting the measured data onto the parameter space

$$\mathbf{w}^\dagger \langle \boldsymbol{B} \rangle = \boldsymbol{g}_C \tag{26}$$

and simultanously truncates the non-parameterized parts, i.e.

$$\mathbf{w}^\dagger \langle \boldsymbol{v} \rangle = 0. \tag{27}$$

Applying the filter matrix to the average of the non-parameterized parts of the field in Eq. (6)

$$0 = \mathbf{w}^\dagger \langle \boldsymbol{v} \rangle = \mathbf{w}^\dagger \left(\langle \boldsymbol{B} \rangle - \mathbf{H} \boldsymbol{g}_C \right) = \mathbf{w}^\dagger \langle \boldsymbol{B} \rangle - \mathbf{w}^\dagger \mathbf{H} \boldsymbol{g}_C = \boldsymbol{g}_C - \mathbf{w}^\dagger \mathbf{H} \boldsymbol{g}_C, \tag{28}$$

where the true coefficient vector has been replaced by Capon's estimator (Eq. 18) and taking into account that $\langle \boldsymbol{n} \rangle = 0$ leads to the distortionless constraint

$$\mathbf{w}^\dagger \mathbf{H} = \mathbf{I}. \tag{29}$$

This equation is one of the important constraints for the construction of the wanted filter matrix $\mathbf{w}$ but it is not enough. Let us look for another criterion. The basic idea is that in Eq. (6) there may be contributions $\langle \boldsymbol{v} \rangle$ in the data $\boldsymbol{B}$ which are not caused by the internal magnetic field and thus, these contributions are not modeled by $\mathbf{H}\boldsymbol{g}$. Although the filter matrix $\mathbf{w}$ is already designed to truncate these parts, i.e. $\mathbf{w}^\dagger \langle \boldsymbol{v} \rangle = 0$, their contributions to the data are unknown ~~. This yields the following procedure.~~and therefore, the parts that have to be truncated by $\mathbf{w}$ are unknown.

Conferring to Eq. (26) the average output power, which is defined as the sum of the quadratic averaged components of the estimator, can be rewritten as (Pillai , 1989)

$$\operatorname{tr} \mathbf{P} = \operatorname{tr} \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle = \operatorname{tr} \left(\mathbf{w}^\dagger \langle \boldsymbol{B} \circ \boldsymbol{B} \rangle \mathbf{w} \right), \tag{30}$$

where $\operatorname{tr} \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle$ is the trace of the matrix $\langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle$.

Using Eq. (30), the coefficient matrix $\mathbf{P}$ can be expressed by the weight $\mathbf{w}$ and the data covariance $\mathbf{M}$ as

$$\mathbf{P} = \langle \boldsymbol{g}_C \circ \boldsymbol{g}_C \rangle = \mathbf{w}^\dagger \langle \boldsymbol{B} \circ \boldsymbol{B} \rangle \mathbf{w} = \mathbf{w}^\dagger \mathbf{M} \mathbf{w}. \tag{31}$$

~~Since the~~ The amount of the data's noise and the amount of the non-parameterized parts are unknown. Thus, it is conservative and save to assume, ~~we assumehere,~~ that a large part of the data ~~are~~ is influenced by the noise and the non-parameterized parts, that have to be truncated by the matrix $\mathbf{w}$ and the underlying model keeps the minimal contribution to the data. This minimal contribution has to be extracted.

Therefore, the output power $P = \operatorname{tr} \mathbf{P}$ has to be minimized with respect to $\mathbf{w}^\dagger$, subject to the distortionless constraint $\mathbf{w}^\dagger \mathbf{H} = \mathbf{I}$ or equivalently $\mathbf{H}^\dagger \mathbf{w} = \mathbf{I}$. Using the Lagrange multiplier method this minimization problem can be formulated as

$$\text{minimize} \quad \operatorname{tr} \left[\mathbf{w}^\dagger \mathbf{M} \mathbf{w} + \boldsymbol{\Lambda} \left(\mathbf{I} - \mathbf{H}^\dagger \mathbf{w} \right) \right], \tag{32}$$

where $\mathbf{\Lambda}$ are the related Lagrange multipliers and the minimum is taken with respect to $\mathbf{w}$. Since the components w_{ij} and $w^\dagger_{ij}$ of the matrix $\mathbf{w}$ and $\mathbf{w}^\dagger$ respectively, are independent of each other, Eq. (32) can be expanded as

$$\text{minimize} \qquad P = \text{tr}\left[\mathbf{w}^\dagger\,\mathbf{M}\,\mathbf{w} + \mathbf{\Lambda}\left(\mathbf{I} - \mathbf{H}^\dagger\mathbf{w}\right) + \left(\mathbf{I} - \mathbf{w}^\dagger\,\mathbf{H}\right)\mathbf{\Gamma}\right] \tag{33}$$

or equivalently

$$\text{minimize} \qquad P = w^\dagger_{ij} M_{jk} w_{ki} + \Lambda_{ii} - \Lambda_{ij} H^\dagger_{jk} w_{ki} + \Gamma_{ii} - w^\dagger_{ij} H_{jk} \Gamma_{ki} \tag{34}$$

with related additional Lagrange multipliers $\mathbf{\Gamma}$. Taking the derivatives with respect to w_{ki} and $w^\dagger_{ij}$ results in

$$0 = \partial_{w_{ki}} P = w^\dagger_{ij} M_{jk} - \Lambda_{ij} H^\dagger_{jk} \tag{35}$$

yielding

$$\mathbf{w}^\dagger\,\mathbf{M} = \mathbf{\Lambda}\,\mathbf{H}^\dagger \tag{36}$$

and

$$0 = \partial_{w^\dagger_{ij}} P = M_{jk} w_{ki} - H_{jk} \Gamma_{ki} \tag{37}$$

resulting in

$$\mathbf{M}\,\mathbf{w} = \mathbf{H}\,\mathbf{\Gamma} \quad . \tag{38}$$

Multiplication of Eq. (36) with $\mathbf{w}$ from the right and multiplication of Eq. (38) with $\mathbf{w}^\dagger$ from the left considering the distortionless constraint delivers

$$\mathbf{P} = \mathbf{w}^\dagger\,\mathbf{M}\,\mathbf{w} = \mathbf{\Gamma} = \mathbf{\Lambda} \tag{39}$$

and therefore

$$P = \text{tr}\,\mathbf{P} = \text{tr}\,\mathbf{\Gamma} = \text{tr}\,\mathbf{\Lambda}. \tag{40}$$

Because $\cancel{P = \text{tr}\,\mathbf{P}}$

$$P = \text{tr}\,\mathbf{P} = \left|\langle \boldsymbol{g}_C \rangle\right|^2 \tag{41}$$

is a convex function, $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ are realizing the minimal output power.

Due to the ensemble averaging the matrix $\mathbf{M}$ is invertible and $\mathbf{M}^{-1}$ exists. Multiplying Eq. (38) with $\mathbf{H}^\dagger\,\mathbf{M}^{-1}$ and again considering the distortionless constraint yields

$$\mathbf{P} = \mathbf{\Lambda} = \mathbf{\Gamma} = \left[\mathbf{H}^\dagger\,\mathbf{M}^{-1}\,\mathbf{H}\right]^{-1}. \tag{42}$$

By means of Eq. (36) the filter matrix results in

$$\mathbf{w}^\dagger = \mathbf{P}\,\mathbf{H}^\dagger\,\mathbf{M}^{-1} \tag{43}$$

and therefore Capon's estimator is given by

$$\boldsymbol{g}_C = \left[\mathbf{H}^\dagger\,\mathbf{M}^{-1}\,\mathbf{H}\right]^{-1}\mathbf{H}^\dagger\,\mathbf{M}^{-1}\langle\boldsymbol{B}\rangle. \tag{44}$$

Regarding the expensive derivation, this compact formular for Capon's estimator is surprising. ~~The same expression can be derived by treating Capon's method as a special case of the maximum likelihood estimator (Narita, 2019).~~

~~Since Capon's method has formerly been applied to the analysis of waves, the existing derivations treat the non-parameterized parts of the field as Gaussian noise. Considering the analysis of planetary magnetic fields this assumption is indefensible. For example, the external parts of the field are systematic noise and cannot be modeled by a Gaussian distribution. Therefore, the above presented derivation generalizes the previous derivations of Capon's method.~~

4 Diagonal Loading

The filter matrix $\mathbf{w}$ is the key parameter to distinguish between the parameterized and the non-parameterized parts of the field. Conferring to Eq. (25) the ratio of these parts defines the input signal-noise-ratio (given by the data)

$$SNR_i = \frac{\mathrm{tr}\left(\mathbf{H}\,\mathbf{P}\,\mathbf{H}^\dagger\right)}{\mathrm{tr}\left(2\left(\mathbf{H}\,\boldsymbol{g}_C\right)\circ\langle\boldsymbol{v}\rangle + \langle\boldsymbol{v}\circ\boldsymbol{v}\rangle + \sigma_n^2\mathbf{I}\right)}. \tag{45}$$

The filter matrix is applied to the disturbed data for estimating the output power, that is related to Capon's estimator. Thus, the output signal-noise-ratio (resulting from the filtering) can be expressed as (Haykin, 2014; Van Trees, 2002)

$$SNR_o = \frac{\mathrm{tr}\left(\mathbf{w}^\dagger\,\mathbf{H}\,\mathbf{P}\,\mathbf{H}^\dagger\,\mathbf{w}\right)}{\mathrm{tr}\left(2\,\mathbf{w}^\dagger\left(\mathbf{H}\,\boldsymbol{g}_C\right)\circ\langle\boldsymbol{v}\rangle\,\mathbf{w} + \mathbf{w}^\dagger\langle\boldsymbol{v}\circ\boldsymbol{v}\rangle\,\mathbf{w}\right) + \sigma_n^2\,\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)} = \frac{\mathrm{tr}\left(\mathbf{P}\right)}{\sigma_n^2\,\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)}, \tag{46}$$

since the filter fulfills the distortionless constraint $\mathbf{w}^\dagger\mathbf{H} = \mathbf{I}$ and truncates the non-parameterized parts of the field, i.e. $\mathbf{w}^\dagger\langle\boldsymbol{v}\rangle = 0$. The ratio of the output and the input signal-noise-ratio is the so-called array gain (Van Trees, 2002)

$$\frac{SNR_o}{SNR_i} = \frac{1}{\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)}\,\frac{\mathrm{tr}\left(\mathbf{P}\right)}{\sigma_n^2\,\mathrm{tr}\left(\mathbf{H}\,\mathbf{P}\,\mathbf{H}^\dagger\right)}\,\mathrm{tr}\left(2\left(\mathbf{H}\,\boldsymbol{g}_C\right)\circ\langle\boldsymbol{v}\rangle + \langle\boldsymbol{v}\circ\boldsymbol{v}\rangle + \sigma_n^2\mathbf{I}\right) \sim \frac{1}{\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)}, \tag{47}$$

which is controlled by $\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)$, since $\mathbf{H}$, $\mathbf{P}$ and $\boldsymbol{v}$ are given by the data and the underlying model.

The input signal-noise-ratio is given by the data and the underlying model and therefore, $SNR_i = \mathrm{const.}$ in Eq. (47). In contrast to the input signal-noise-ratio, the output signal-noise-ratio depends on the filtering. When $\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)$ is large, the output signal-noise-ratio can decrease ($SNR_o \to 0$), resulting in signal elimination and thus, the performance of Capon's estimator degrades. To prevent the signal elimination and to improve the robustness of Capon's estimator it is desirable to restrict tr $\left(\mathbf{w}^\dagger\mathbf{w}\right)$ with an upper boundary $T_0 = \mathrm{const.}$ (Van Trees, 2002), which can be expressed by the additional quadratic constraint

$$\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right) = T_0. \tag{48}$$

For reasons of mathematical aesthetics, the constant T_0 is expressed as the trace of a matrix $\mathbf{T}$. For example, one can choose

$$\mathbf{T} = \frac{T_0}{G}\mathbf{I} \in \mathbb{R}^{G\times G}, \tag{49}$$

where G again indicates the number of wanted expansion coefficients and $\mathbf{I} \in \mathbb{R}^{G\times G}$ is the identity matrix, so that

$$\mathrm{tr}\left(\mathbf{T}\right) = \frac{T_0}{G}\,\mathrm{tr}\left(\mathbf{I}\right) = T_0 \tag{50}$$

holds. Thus, Eq. (48) can be rewritten as

$$\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w} - \mathbf{T}\right) = 0. \tag{51}$$

It should be noted that the matrix $\mathbf{T}$ can be chosen arbitrarily as long as it is independent of $\mathbf{w}$ and $\mathbf{w}^\dagger$.

Conferring to the previous section and considering the additional quadratic constraint, the filter matrix can be calculated by solving

$$\text{minimize} \qquad \mathrm{tr}\left[\mathbf{w}^\dagger\mathbf{M}\mathbf{w} + \sigma_d^2\left(\mathbf{w}^\dagger\mathbf{w} - \mathbf{T}\right) + \mathbf{\Lambda}\left(\mathbf{I} - \mathbf{H}^\dagger\mathbf{w}\right) + \left(\mathbf{I} - \mathbf{w}^\dagger\mathbf{H}\right)\mathbf{\Gamma}\right] \tag{52}$$

with respect to $\mathbf{w}$, where σ_d^2 is the related additional Lagrange multiplier. Carrying out the same procedure as described in section 3, the constrained minimizer is given by

$$\mathbf{w} = \left(\mathbf{M} + \sigma_d^2\mathbf{I}\right)^{-1}\mathbf{H}\left[\mathbf{H}^\dagger\left(\mathbf{M} + \sigma_d^2\mathbf{I}\right)^{-1}\mathbf{H}\right]^{-1}. \tag{53}$$

The comparison of Eq. (43) with Eq. (53) shows that the quadratic constraint results in the addition of the constant value σ_d^2 to the diagonal of the data covariance matrix $\mathbf{M}$, which is known as diagonal loading (Van Trees, 2002). Consequently, the filter matrix is designed for a higher Gaussian background noise than is actually present (Van Trees, 2002).

In Figure 1 tr $\left(\mathbf{w}^\dagger\mathbf{w}\right)$ is displayed with respect to $\sigma = \sqrt{\sigma_d^2 + \sigma_n^2}$. For $\sigma_d \to 0$, tr $\left(\mathbf{w}^\dagger\mathbf{w}\right)$ is large resulting in a small output signal-noise ratio, which can cause signal elimination. For increasing values of σ_d, tr $\left(\mathbf{w}^\dagger\mathbf{w}\right)$ decreases and for $\sigma_d \to \infty$ Capon's filter converges to the least square fit filter

$$\mathbf{w}^\dagger \overset{\sigma_d\to\infty}{\longrightarrow} \left[\mathbf{H}^\dagger\mathbf{H}\right]^{-1}\mathbf{H}^\dagger \tag{54}$$

or equivalently

$$\mathbf{w}^\dagger\mathbf{w} \overset{\sigma_d\to\infty}{\longrightarrow} \left[\mathbf{H}^\dagger\mathbf{H}\right]^{-1}, \tag{55}$$

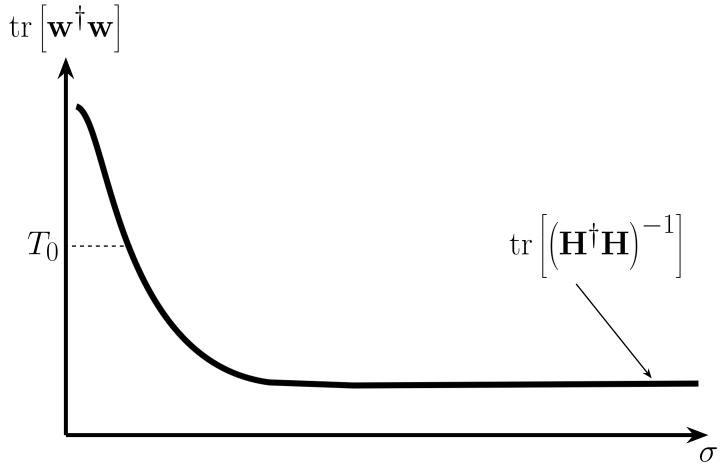

Figure 1. Sketch of the trace of the filter matrix with respect to $\sigma = \sqrt{\sigma_d^2 + \sigma_n^2}$. For increasing values of the diagonal loading parameter σ_d the output signal-noise ratio increases and Capon's filter converges to the least square fit filter, if the diagonal loading parameter is large.

that treats all data equally.

Since $\mathrm{tr}\left(\mathbf{w}^\dagger\mathbf{w}\right)$ is a monotonically decreasing function, $\sigma_d \to \infty$ might be the best choice for the diagonal loading parameter. To check this expectation we have to take a look at the output power P_d for the synthetic increased noise, which can be estimated by replacing $\mathbf{M} \to \mathbf{M} + \sigma_d^2\mathbf{I}$ in Eq. (42), resulting in (Pajovic, 2019; Richmond et al., 2005)

$$P_d = \mathrm{tr}\left[\mathbf{w}^\dagger\left(\mathbf{M} + \sigma_d^2\mathbf{I}\right)\mathbf{w}\right] = \mathrm{tr}\left[\mathbf{H}^\dagger\left(\mathbf{M} + \sigma_d^2\mathbf{I}\right)^{-1}\mathbf{H}\right]^{-1} \overset{\sigma_d \to \infty}{\longrightarrow} \sigma_d^2\,\mathrm{tr}\left[\mathbf{H}^\dagger\mathbf{H}\right]^{-1}. \tag{56}$$

Thus, P_d is an increasing function of σ_d. Since the output power has to be minimized one would expect that $\sigma_d = 0$ is the best choice, which is in contradiction to the argumentation above. Therefore, the maximization of the array gain is not equivalent to the minimization of the output power (Van Trees, 2002). Since σ_d cannot be directly calculated within the minimization procedure (Eq. 52), it is not clear how to choose the optimal diagonal loading parameter $\sigma_{opt.}$, that lies somewhere between those extrema.

In the literature there exist several methods for determining the optimal diagonal loading parameter (Pajovic, 2019). In contrast to the analysis of waves, we favor measurements that are stationary up to the Gaussian noise for the analysis of planetary magnetic fields. Comparing the measurement times with planetary geology time scales this assumption is surely valid for the internal magnetic field. For the external parts of the field this can be realized by choosing data sets of preferred situations, e.g. calm solar wind conditions. By virtue of the stationarity, the data covariance matrix $\mathbf{M} = \langle\boldsymbol{B}\rangle \circ \langle\boldsymbol{B}\rangle + \sigma_n^2\,\mathbf{I}$ contains only one non-trivial eigenvalue $\lambda_1 = \left|\langle\boldsymbol{B}\rangle\right|^2 + \sigma_n^2$ and $\lambda_i = \sigma_n^2$, for $i = 2,\ldots,3N$ elsewhere ~~.~~ (cf. Appendix B). Therefore, estimators for the diagonal loading parameter, that are related to eigenvalues corresponding to interference and noise (Carlson, 1988) or estimators taking into account the standard deviation of the diagonal elements of the data covariance matrix, i.e. $\sigma_d^2 = \mathrm{std}\{\lambda_1, \ldots, \lambda_{3N}\}$ (Ma and Goh, 2003) cannot be applied. When simulated data are analyzed the deviation between Capon's estimator and the true coefficient vector, implemented in the simulation, is a useful metric for estimating the optimal diagonal loading parameter (Pajovic, 2019; Toepfer et al., 2020). But when the method is applied to real spacecraft data no true coefficient vector is available and therefore, another estimation method for the diagonal loading parameter, that solely depends on the data and the underlying model is required.

The additional quadratic constraint (Eq. 48), resulting in diagonal loading, bounds the trace of the filter matrix, that is the solution of the minimization procedure (Eq. 52). To prevent signal elimination, $\mathrm{tr}\left(\mathbf{w}^\dagger \mathbf{w}\right)$ and P_d have to be minimized simultanously. Since $\mathrm{tr}\left(\mathbf{w}^\dagger \mathbf{w}\right)$ is decreasing for higher values of σ_d (cf. Figure 1), P_d is an increasing function and thus, they act as competitors. At the optimal value $\sigma_{opt.}$ the two competitors compromise, which can be unterstood as $\mathrm{tr}\left(\mathbf{w}^\dagger \mathbf{w}\right)$ reaches its minimal value under the constraint that P_d is maximal. Therefore, the diagonal loading of the data covariance matrix is equivalent to the Tikhonov regularization for ill-posed problems~~and the~~. The Tikhonov regularization improves the robustness of the least square fit problem:

$$\text{minimize} \quad \left|\mathbf{H}\boldsymbol{g} - \boldsymbol{B}\right|^2 \tag{57}$$

with respect to $\boldsymbol{g}$ under the constraint of solutions $\boldsymbol{g}$ with minimal norm, i.e.

$$\text{minimize} \quad \left|\mathbf{H}\boldsymbol{g} - \boldsymbol{B}\right|^2 + \alpha\left|\boldsymbol{g}\right|^2, \tag{58}$$

where α is the corresponding Lagrange multiplier, which is also known as the regularization parameter (Tikhonov et al., 1995). In analogy to the Lagrange multiplier α, the optimal diagonal loading parameter can be estimated ~~analogously~~ by the method of the L-curve (Hiemstra et al. (2002)). The L-curve arises by plotting $\lg\left[\mathrm{tr}\left(\mathbf{w}^\dagger \mathbf{w}\right)\right]$ versus $\lg\left[\mathrm{tr}\left(\mathbf{w}^\dagger\left(\mathbf{M} + \sigma_d^2 \mathbf{I}\right)\mathbf{w}\right)\right]$ for different values of σ_d and is displayed in Figure 2. The optimal value $\sigma_{opt.}$ is located in the vicinity of the L-curve's knee, which is defined by the maximum curvature of the L-curve (Hiemstra et al., 2002).

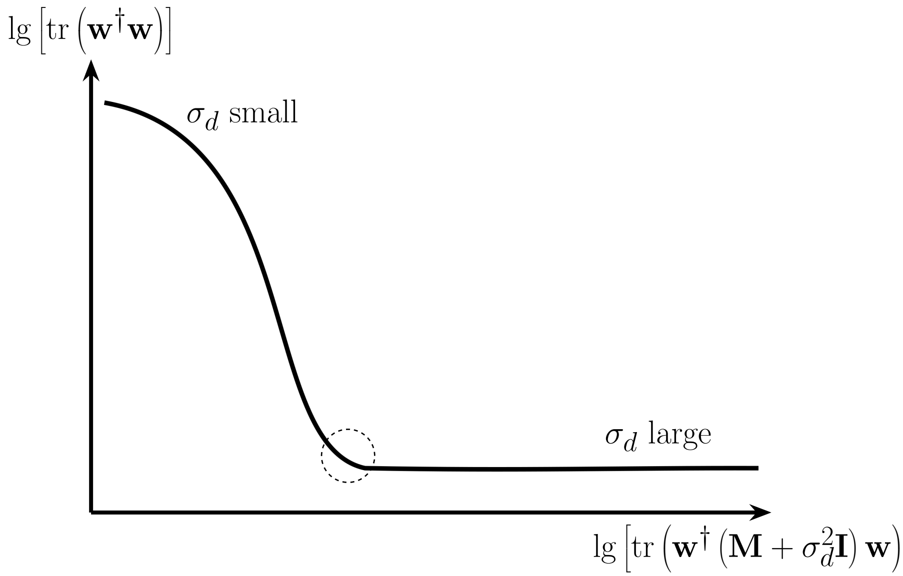

Figure 2. Sketch of the L-curve for estimating the optimal diagonal loading parameter σ_d. The optimal value is located in the vicinity of the L-curves knee (dashed circle).

5 Practical application of Capon's method

When Capon's method is applied to the reconstruction of Mercury's internal magnetic field (Toepfer et al., 2020), several computationally burdensome matrix inversions are necessary to calculate the estimator (cf. Eq. 44)

$$\boldsymbol{g}_C = \left[\mathbf{H}^\dagger \mathbf{M}^{-1} \mathbf{H}\right]^{-1} \mathbf{H}^\dagger \mathbf{M}^{-1} \langle \boldsymbol{B} \rangle. \tag{59}$$

Thus, for the practical application of Capon's method it is useful to rewrite the method in terms of the least square fit method.

Capon's estimator inherits the matrix operation structure of the least square fit estimator, which is given by (Haykin, 2014)

$$\boldsymbol{g}_L = \left[\mathbf{H}^\dagger \mathbf{H}\right]^{-1} \mathbf{H}^\dagger \langle \boldsymbol{B} \rangle. \tag{60}$$

Substituting $\mathbf{H} \rightarrow \mathbf{M}^{-1/2}\mathbf{H}$ and $\langle \boldsymbol{B} \rangle \rightarrow \mathbf{M}^{-1/2}\langle \boldsymbol{B} \rangle$ in Eq. (60), the least square fit estimator may be converted into Capon's estimator.

The least square fit method minimizes the deviaton between the disturbed measurements $\boldsymbol{B}$ and the underlying model $\mathbf{H}\boldsymbol{g}$, measured in the Euclidian norm $||.||_2$, so that

$$\boldsymbol{g}_L = \arg\min_{\boldsymbol{g}} \left|\left|\mathbf{H}\boldsymbol{g} - \boldsymbol{B}\right|\right|_2^2. \tag{61}$$

Conferring to the above mentioned substitutions, Capon's method can be interpreted as measuring the deviation in Eq. (61) in a different norm

$$\left|\left|\mathbf{H}\boldsymbol{g} - \boldsymbol{B}\right|\right|_{\mathbf{M}^{-1}}^2 = \left|\left|\mathbf{M}^{-1/2}\left(\mathbf{H}\boldsymbol{g} - \boldsymbol{B}\right)\right|\right|_2^2 \quad , \tag{62}$$

so that Capon's estimator is given by

$$\boldsymbol{g}_C = \arg\min_{\boldsymbol{g}} \left|\left|\mathbf{H}\boldsymbol{g} - \boldsymbol{B}\right|\right|_{\mathbf{M}^{-1}}^2. \tag{63}$$

Thus, Capon's method can be regarded as a special case of the least square fit method or more precisely of the weighted least square fit method, where the data and the model are measured and weighted with the inverse data covariance matrix. This property is useful for the practical application of Capon's method. In contrast to the computationally burdensome matrix inversions in Eq. (44), the usage of minimization algorithms, such as gradient descent or conjugate gradient method, for solving Eq. (63) is more stable and computationally inexpensive.

To illustrate the above presented mathematical foundations, Capon's method is applied to simulated magnetic field data in order to reconstruct Mercury's internal magnetic field. As a proof of concept, the underlying model is restricted to the internal dipole and quadrupole contributions to the magnetic field as discussed in section 2. Since simulated magnetic field data are analyzed, the special application to Mercury's magnetic field is of minor importance, because the ideal solution is known from the simulation. If the method would be tested against in-situ measurement data, the application to the analysis of the Earth's magnetic field is more suitable since the Earth's magnetic field is better known than Mercury's magnetic field.

For the reconstruction of Mercury's internal dipole and quadrupole field the internal Gauss coefficients $g_1^0 = -190\,\mathrm{nT}$ and $g_2^0 = -78\,\mathrm{nT}$ (Anderson et al., 2012; Wardinski et al., 2019), defining the non-vanishing components of the true coefficient vector $\boldsymbol{g}$ are implemented in the hybrid code AIKEF (Müller et al., 2011) and the magnetic field data resulting from the plasma interaction of Mercury with the solar wind are simulated. The data are evaluated along selected parts of the prospective trajectories of the BepiColombo mission on the night side of Mercury within a distance of $0.2\,R_\mathrm{M}$ up to $0.4\,R_\mathrm{M}$ from Mercury's surface. Since simulated data are analyzed, the deviation between the true coefficient vector $\boldsymbol{g}$ and Capon's estimator $\boldsymbol{g}_C$ can be used as a metric to verify the estimation of the optimal diagonal loading parameter by making use of the L-curve technique. The optimal diagonal loading parameter results in $\sigma_{opt.} \approx 800\,\mathrm{nT}$ which corresponds with the vicinity of the L-curve's knee. The reconstructed Gauss coefficients for the internal dipole and quadrupole field are presented in Table 1.

Table 1. Implemented and reconstructed Gauss coefficients for the internal dipole and quadrupole field.

Gauss coefficient	input in nT	output Capon in nT
g_1^0	-190.0	-189.2
g_1^1	0.0	1.9
h_1^1	0.0	0.2
g_2^0	-78.0	-68.4
g_2^1	0.0	26.1
h_2^1	0.0	11.4
g_2^2	0.0	-2.4
h_2^2	0.0	0.0

The deviation $\left|\boldsymbol{g}_C - \boldsymbol{g}\right|$ between Capon's estimator and the true coefficient vector results in $30.2\,\mathrm{nT}$ or $14.7\,\%$, respectively, for the optimal diagonal loading parameter. When the magnetic field data are evaluated at an ensemble of data points with a distance of $0.2\,R_\mathrm{M}$ from Mercury's surface this deviation is of the same order (Toepfer et al., 2020). Since the underlying model neglects the external parts and only the internal parts are considered, the coefficients g_2^1 and h_2^1 show large deviations to the implemented coefficients. Extending the underlying model by a parameterization of the external parts of the magnetic field improves Capon's estimator especially when the magnetic field data are evaluated ~~in~~ at some distance above Mercury's surface. In principle, one can analyze also the external field, if some model is adopted.

6 Discussion of Capon's method

Capon's method has formerly been applied to the analysis of waves. Thus, the existing derivations treat the non-parameterized parts of the field $\boldsymbol{v}$ as Gaussian noise, so that $\boldsymbol{v}$ and the measurement noise $\boldsymbol{n}$ are of the same character (Motschmann et al., 1996) . Considering the analysis of planetary magnetic fields this assumption is indefensible. For example, the external parts of the field are systematic noise and cannot be modeled by a Gaussian distribution. When the non-parameterized parts are Gaussian, i.e. $\langle \boldsymbol{v} \rangle = 0$, the truncation of these parts by the filter matrix $\mathbf{w}^\dagger \langle \boldsymbol{v} \rangle = 0$ (cf. Eq. 27) is fulfilled trivially, which reduces the terms within the derivation and the estimation of the diagonal loading parameter. Therefore, the above presented mathematical foundations generalize the previous derivations of Capon's method and transist into the derivation presented by Motschmann et al. (1996) for the special case of $\langle \boldsymbol{v} \rangle = 0$.

As already mentioned in the introduction (Sec. 1), Capon's method can be regarded from several mathematical perspectives. Within the derivation presented above, the output power P, which is defined as the trace of the coefficient (covariance) matrix is minimized with respect to the filter matrix $\mathbf{w}$, subject to the distortionless contraint $\mathbf{w}^\dagger \mathbf{H} = \mathbf{I}$. This procedure corresponds with the name Minimum Variance Distortionless Response Estimator (MVDR), since P contains the variance of the model coefficients. Narita (2019) showed, that Capon's estimator can also be derived by treating Capon's method as a special case of the maximum likelihood estimator by regarding the likehood function as nearly Gaussian (particularly around the peak of the likelhood function). As discussed in Sec. 5, Capon's method can also be interpreted as a special case of the weighted least square fit. This illustrates, that the several existing inversion methods for linear inverse problems are connected with each other and are not as different as they seem to be at first appearance.

7 Conclusions

Capon's method is a robust and useful tool for various kinds of ill-posed inverse problems, such as Mercury's planetary magnetic field analysis. The derivation of the method can be regarded from different mathematical perspectives. Here we revisited the linear-algebraic matrix formulation of the method and extended the derivation for Mercury's magnetic field analysis. Capon's method becomes even more robust by incorporating the diagonal loading technique. Thereby, the construction of a filter matrix is vital to the derivation of Capon's estimator.

Especially the trace of the filter matrix determines the array gain, which is defined as the ratio of the output and the input signal-noise-ratio. If the trace is large, the output signal-noise-ratio can decrease resulting in signal elimination and thus, the performance of Capon's estimator degrades. Bounding the trace of the filter matrix results in diagonal loading of the data covariance matrix, which improves the robustness of the method. The main problem of the diagonal loading technique is that in general it is not clear how to choose the optimal diagonal loading parameter. Since for the analysis of planetary magnetic fields we prefer measurements that are stationary up to the Gaussian noise, estimators for the diagonal loading parameter, that are related to eigenvalues of the data covariance matrix corresponding to interference and noise cannot be applied. Making use of the L-curve's technique enables a robust procedure for estimating the optimal diagonal loading parameter.

For the calculation of Capon's estimator several computationally burdensome matrix inversions are necessary. Interpretation of Capon's method as a special case of the least square fit method enables the usage of numerically more stable and less burden minimization algorithms, e.g. gradient descent or conjugate gradient method, for calculating Capon's estimator.

It should be noted that the parameterization of Mercury's internal magnetic field via the Gauss representation, as mentioned in section 2, is only one of several possibilities for modeling the magnetic field in the vicinity of Mercury. The underlying model can be extended by other parameterizations, for example the Mie representation (toroidal-poloidal decomposition) or magnetospheric models and Capon's method can be applied to estimate the related model coefficients. Besides the analysis of Mercury's internal magnetic field, the extention of the model also enables the reconstruction of current systems flowing in the magnetosphere. Concerning the BepiColombo mission this work establishes a mathematical basis for the application of Capon's method to analyze Mercury's internal magnetic field in a robust and manageable way.

Data availability. The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Appendix A: Determinant of the outer product

The influence of the averaging to the determinant of the data covariance marix $\mathbf{M}$ is exemplarily illustrated for the three-dimensional case. Thus, the magnetic field vector is given by

$$\boldsymbol{B} = \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} \tag{A1}$$

and therefore, the outer product results in

$$\boldsymbol{B} \circ \boldsymbol{B} = \begin{bmatrix} B_x^2 & B_x B_y & B_x B_z \\ B_x B_y & B_y^2 & B_y B_z \\ B_x B_z & B_y B_z & B_z^2 \end{bmatrix} \tag{A2}$$

with a vanishing determinant

$$\det(\boldsymbol{B} \circ \boldsymbol{B}) = 3B_x^2 B_y^2 B_z^2 - 3B_x^2 B_y^2 B_z^2 = 0. \tag{A3}$$

Throughout the averaging of the data, the data covariance matrix results in

$$\mathbf{M} = \langle \boldsymbol{B} \circ \boldsymbol{B} \rangle = \langle \boldsymbol{B} \rangle \circ \langle \boldsymbol{B} \rangle + \sigma_n^2 \underline{\underline{I}} = \begin{bmatrix} \langle B_x^2 \rangle + \sigma_n^2 & \langle B_x \rangle \langle B_y \rangle & \langle B_x \rangle \langle B_z \rangle \\ \langle B_x \rangle \langle B_y \rangle & \langle B_y^2 \rangle + \sigma_n^2 & \langle B_y \rangle \langle B_z \rangle \\ \langle B_x \rangle \langle B_z \rangle & \langle B_y \rangle \langle B_z \rangle & \langle B_z^2 \rangle + \sigma_n^2 \end{bmatrix} \tag{A4}$$

with a non-vanishing determintant

$$\det(\mathbf{M}) = (\langle B_x^2 \rangle + \sigma_n^2)(\langle B_y^2 \rangle + \sigma_n^2)(\langle B_z^2 \rangle + \sigma_n^2) - \langle B_x^2 \rangle \langle B_y^2 \rangle \langle B_z^2 \rangle - \sigma_n^2 \langle B_y^2 \rangle \langle B_z^2 \rangle - \sigma_n^2 \langle B_x^2 \rangle \langle B_z^2 \rangle - \sigma_n^2 \langle B_x^2 \rangle \langle B_y^2 \rangle \tag{A5}$$

$$= \sigma_n^4 (\langle B_x^2 \rangle + \langle B_y^2 \rangle + \langle B_z^2 \rangle) + \sigma_n^6 \tag{A6}$$

$$\neq 0. \tag{A7}$$

Thus, the inverse of $\mathbf{M}$ exists, whereas the outer product $\boldsymbol{B} \circ \boldsymbol{B}$ is singular.

Appendix B: Eigenvalues of the data covariance matrix

The data covariance matrix is defined as

$$\mathbf{M} = \langle \boldsymbol{B} \circ \boldsymbol{B} \rangle = \langle \boldsymbol{B} \rangle \circ \langle \boldsymbol{B} \rangle + \sigma_n^2 \mathbf{I}. \tag{B1}$$

This matrix is quadratic and especially diagonalisable. Thus, there exists a matrix $\mathbf{D_M}$ which is similar to the matrix $\mathbf{M}$, so that

$$\mathbf{D_M} = \mathbf{V}^\dagger \mathbf{M} \mathbf{V}, \tag{B2}$$

where $\mathbf{V}$ is an orthogonal transformation, i.e. $\mathbf{V}^{\dagger}\mathbf{V} = \mathbf{I}$ and $\mathbf{D_M}$ is a diagonal matrix which diagonal elements are given by the eigenvalues of $\mathbf{M}$. Inserting the definition of the matrix $\mathbf{M}$ delivers

$$\mathbf{D_M} = \mathbf{V}^{\dagger}\,\mathbf{M}\,\mathbf{V} = \mathbf{V}^{\dagger}\,\langle \boldsymbol{B} \circ \boldsymbol{B} \rangle\,\mathbf{V} = \mathbf{V}^{\dagger}\left(\langle \boldsymbol{B} \rangle \circ \langle \boldsymbol{B} \rangle\right)\mathbf{V} + \sigma_n^2 \mathbf{I}, \tag{B3}$$

since $\mathbf{V}^{\dagger}\mathbf{V} = \mathbf{I}$. To determine the diagonal form of the outer product, the two-dimensional case where

$$\langle \boldsymbol{B} \rangle = \begin{pmatrix} \langle B_x \rangle \\ \langle B_y \rangle \end{pmatrix} \tag{B4}$$

is considered exemplarily. Thus,

$$\langle \boldsymbol{B} \rangle \circ \langle \boldsymbol{B} \rangle = \begin{bmatrix} \langle B_x \rangle^2 & \langle B_x \rangle \langle B_y \rangle \\ \langle B_x \rangle \langle B_y \rangle & \langle B_y \rangle^2 \end{bmatrix} \tag{B5}$$

and the characteristic polynomial results in

$$\left(\langle B_x \rangle^2 - \beta\right)\left(\langle B_y \rangle^2 - \beta\right) - \langle B_x \rangle^2 \langle B_y \rangle^2 = 0 \tag{B6}$$

or equivalently

$$\beta^2 - \beta\left(\langle B_x \rangle^2 + \langle B_y \rangle^2\right) = \beta^2 - \beta \left|\langle \boldsymbol{B} \rangle\right|^2 = 0, \tag{B7}$$

where β denotes the eigenvalue which is given by $\beta = 0$ and $\beta = \left|\langle \boldsymbol{B} \rangle\right|^2$. Therefore, in general the diagonal matrix of the outer product is given by

$$\mathbf{D}_{\langle \boldsymbol{B} \rangle \circ \langle \boldsymbol{B} \rangle} = \begin{bmatrix} \left|\langle \boldsymbol{B} \rangle\right|^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{B8}$$

The noise matrix $\sigma_n^2 \mathbf{I}$ is already a diagonal matrix, so that the diagonal form of $\mathbf{M}$ is given by

$$\mathbf{D_M} = \begin{bmatrix} \left|\langle \boldsymbol{B} \rangle\right|^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} + \sigma_n^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \left|\langle \boldsymbol{B} \rangle\right|^2 + \sigma_n^2 & 0 & \cdots & 0 \\ 0 & \sigma_n^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n^2 \end{bmatrix}. \tag{B9}$$

Thus, the data covariance matrix contains only one non-trivial eigenvalue $\lambda_1 = \langle \boldsymbol{B} \rangle\big|^2 + \sigma_n^2$ and $\lambda_i = \sigma_n^2$ for $i = 1, \ldots, 3N$.

Author contributions. All authors contributed conception and design of the study; ST, YN and UM wrote the first draft of the manuscript; All authors contributed to manuscript revision, read and approved the submitted version.

Funding: We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Technische Universität Braunschweig.

The work by Y. Narita is supported by the Austrian Space Applications Programme at the Austrian Research Promotion Agency under contract 865967.

D. Heyner was supported by the German Ministerium für Wirtschaft und Energie and the German Zentrum für Luft- und Raumfahrt under contract 50 QW1501.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors are grateful for stimulating discussions and helpful suggestions by Karl-Heinz Glassmeier and Alexander Schwenke.

References

Anderson, B. J., Johnson, C. L., Korth, H., Purucker, M. E., Winslow, R. M., Slavin, J. A., Solomon, S. C., McNutt Jr., R. L., Raines, J. M., and Zurbuchen, T. H.: The global magnetic field of Mercury from MESSENGER orbital observations, Science, 333, 1859–1862, https://doi.org/10.1126/science.1211001, 2011.

Anderson, B. J., Johnson, C. L., Korth, H., Winslow, R. M., Borovsky, J. E., Purucker, M. E., et al.: Low-degree structure in Mercury's planetary magnetic field, J. Geophys. Res., 117, E00L12, https://doi.org/10.1029/2012JE004159, 2012.

Benkhoff, J., van Casteren, J., Hayakawa, H., Fujimoto, M., Laakso, H., Novara, M., Ferri, P., Middleton, H. R., and Ziethe, R.: BepiColombo–Comprehensive exploration of Mercury: Mission overview and science goals, Planet. Space Sci., 85, 1–2, 2–20, https://doi.org/10.1016/j.pss.2009.09.020, 2010.

Capon, J.: High resolution frequency-wavenumber spectrum analysis, Proc. IEEE, 57, 1408–1418, https://doi.org/10.1109/PROC.1969.7278, 1969.

Carlson, B. D.: Covariance matrix estimation errors and diagonal loading in adaptive arrays, IEEE Trans. Aerosp. Electron. Syst., 24, 397–401, https://doi.org/10.1109/7.7181, 1988.

Constantinescu, O. D., Glassmeier, K.-H., Motschmann, U., Treumann, R. A., Fornaçon, K. -H., and Fränz, M.: Plasma wave source location using CLUSTER as a spherical wave telescope, J. Geophys. Res., 111, A09221, https://doi.org/10.1029/2005JA011550, 2006.

Gauß, C. F.: Allgemeine Theorie des Erdmagnetismus: Resultate aus den Beobachtungen des magnetischen Vereins im Jahre 1838, edited by: Gauss, C. F. and Weber, W., 1–57, Weidmannsche Buchhandlung, Leipzig, 1839.

Glassmeier, K.-H., Motschmann, U., Dunlop, M., Balogh, A., Acuña, M. H., Carr, C., Musmann, G., Fornaçon, K.-H., Schweda, K., Vogt, J., Georgescu, E., and Buchert, S.: Cluster as a wave telescope – first results from the fluxgate magnetometer, Ann. Geophys., 19, 1439–1447, https://doi.org/10.5194/angeo-19-1439-2001, 2001.

Glassmeier, K.-H., Auster, H.-U., Heyner, D., Okrafka, K., Carr, C., Berghofer, G., Anderson, B. J., et al.: The fluxgate magnetometer of the BepiColombo Mercury Planetary Orbiter, Planet. Space Sci., 58, 287–299, https://doi.org/10.1016/j.pss.2008.06.018, 2010.

Haykin, S.: Adaptive Filter Theory (5th ed.), International edition, Pearson, 2014.

Hiemstra, J. H., Wippert, M. W., Goldstein, J. S., and Pratt, T.: Application of the L-curve technique to loading level determination in adaptive beamforming, Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2002, 2, 1261–1266, 2002.

Ma, N., and Goh, J. T.: Efficient method to determine diagonal loading value, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 5, 341–344, https://doi.org/10.1109/ICASSP.2003.1199948, 2003.

Motschmann, U., Woodward, T. I., Glassmeier, K.-H., Southwood, D. J., and Pinçon, J.-L.: Wavelength and direction filtering by magnetic measurements at satellite arrays: Generalized minimum variance analysis, J. Geophys. Res., 101, 4961–4966, https://doi.org/10.1029/95JA03471, 1996.

Müller, J., Simon, S., Motschmann, U., Schüle, J., Glassmeier, K.-H., and Pringle, G. J.: A.I.K.E.F.: Adaptive hybrid model for space plasma simulations, Comput. Phys. Commun., 182: 946–966, https://doi.org/10.1016/j.cpc.2010.12.033, 2011.

Narita, Y.: A note on Capon's minimum variance projection for multi-spacecraft data analysis, Front. Phys., 7, 8, https://doi.org/10.3398/fphy.2019.00008, 2019.

Narita, Y., Glassmeier, K.-H., Schäfer, S., Motschmann, U., Sauer, K., Dandouras, I., Fornaçon, K. H., Georgescu, E., and Rème, H.: Dispersion analysis of ULF waves in the foreshock using cluster data and the wave telescope technique, Geophys. Res. Lett., 30, https://doi.org/10.1029/2003GL017432, 2003.

Narita, Y., Nakamura, R., and Baumjohann, W.: Cluster as current sheet surveyor in the magnetotail, Ann. Geophys., 31, 1605–1610, https://doi.org/10.5194/angeo-31-1605-2013, 2013.

Pajovic, M., Preisig, J. C., and Baggeroer, A. B.: Analysis of optimal diagonal loading for MPDR-based spatial power estimators in the snapshot deficient regime, IEEE Journal of Oceanic Engineering, 44, 451-465, https://doi.org/10.1109/JOE.2018.2815480, 2018.

Pillai, S. U.: Array Signal Processing, Springer Verlag, New York, 1989.

Plaschke, F., Glassmeier, K.-H., Constantinescu, O. D., Mann, I. R., Milling, D. K., Motschmann, U., and Rae, I. J.: Statistical analysis of ground based magnetic field measurements with the field line resonance detector, Ann. Geophys., 26, 3477–3489, https://doi.org/10.5194/angeo-26-3477-2008, 2008.

Richmond, C. D., Nadakuditi, R. R., and Edelman, A.: Asymptotic mean squared error performance of diagonally loaded Capon-MVDR processor, Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers 2005, Pacific Grove, CA, 1711-1716, 2005.

Toepfer, S., Narita, Y., Heyner, D., and Motschmann, U.: The Capon method for Mercury's magnetic field analysis, Front. Phys., 8, 249, https://doi.org/10.3389/fphy.2020.00249, 2020.

Tikhonov, A. N., Goncharsky, A., Stepanov, V. V., Yagola, A. G.: Numerical Methods for the Solution of Ill-Posed Problems, Springer Netherlands, Netherlands, 1995. ISBN 079233583X

Van Trees, H. L.: Detection, Estimation, and Modulation Theory, Optimum Array Processing, Wiley, New York, 2002.

Wardinski, I., Langlais, B., and Thébault, E.: Correlated time-varying magnetic field and the core size of Mercury, J. Geophys. Res., 124, 2178–2197, https://doi.org/10.1029/2018JE005835, 2019.