Multifrequency array calibration in presence of radio frequency interferences

Yassine Mhiri, Mohammed Nabil El Korso, Arnaud Breloy, Pascal Larzabal

Multifrequency Array Calibration in Presence of Radio Frequency Interferences

Yassine Mhiri[a,*], Mohammed Nabil El Korso[b], Arnaud Breloy[b], Pascal Larzabal[a]

[a]*SATIE, CNRS, ENS Paris-Saclay, Université Paris-Saclay ,4 Avenue des Sciences 91190 Gif-sur-Yvette, France*
[b]*LEME, Université Paris Nanterre, 50 rue de Sèvres 92410 Ville d'Avray, France*

Abstract

Radio interferometers are phased arrays producing high-resolution images from the covariance matrix of measurements. Calibration of such instruments is necessary and is a critical task. This is how the estimation of instrumental errors is usually done thanks to the knowledge of referenced celestial sources. However, the use of high sensitive antennas in modern radio interferometers (LOFAR, SKA) brings a new challenge in radio astronomy because they are more sensitive to Radio Frequency Interferences (RFI). The presence of RFI during the calibration process generally induces biases in state-of-the-art solutions. The purpose of this paper is to propose an alternative to alleviate the effects of RFI. For that, we first propose a model to take into account the presence of RFI in the data across multiple frequency channels thanks to a low-rank structured noise. We then achieve maximum likelihood estimation of the calibration parameters with a Space Alternating Generalized Expectation-Maximization (SAGE) algorithm for which we derive originally two sets of complete data allowing closed-form expressions for the updates. Numerical simulations show a significant gain in performance for RFI corrupted data in comparison with some more classical methods.

Keywords: Radio astronomy, calibration, SAGE, RFI mitigation, array signal processing

1. Introduction

Radio astronomy uses radio telescopes to detect weak emissions from celestial sources in the radio spectrum. Since the first radio observation of the sky in 1931 by Karl Jansky [1], we have been able to monitor the sky in unprecedented detail with growing resolution, bringing new insights in various scientific domains such as solar monitoring, planetology, and astrophysics. Moreover, there has been major improvements in the technology used by radio interferometers, making possible the first imaging of a black hole [2]. A radio interferometer is an array of antennas that are placed far apart from each other. Thanks to the correlations (called visibilities) between signals at each sensor it produces a high-resolution image of the sky [3]. A new generation of radio interferometers, such as the LOFAR or the SKA, are designed to further improve the sensitivity of radio observations over a large bandwidth using multiple sensors placed on a large area. These technological advances bring several array processing challenges to exploit the full scientific potential of such instruments. In this paper, we consider the problem of array calibration, which is a necessary step to produce reliable observations of the sky with high sensitivity.

Array calibration consists in finding an estimate of the systematic instrumental errors affecting the measurements. Such perturbations range from antenna electronic gains to atmospheric perturbations and can be modeled using the Jones framework [4][5]. Array calibration methods can usually be classified into two types of approaches depending on the presence or not of calibrator sources. Radio interferometer calibration

*Corresponding author
Email addresses: yassine.mhiri@universite-paris-saclay.fr (Yassine Mhiri), m.elkorso@parisnanterre.fr (Mohammed Nabil El Korso), a.breloy@parisnanterre.fr (Arnaud Breloy), pascal.larzabal@universite-paris-saclay.fr (Pascal Larzabal)

lies in the first category where we rely on the position and flux of strong celestial sources as calibrators to estimate the parameters of the model. The field of radio astronomy has seen several successive generations of calibration algorithms. The latest tendency brings the focus on statistical modeling by the formulation of robust calibration algorithms as the solution of a maximum likelihood problem. In the literature, we can find numerous non-linear least square approaches [6][7][8][9] that assume an additive Gaussian noise on the visibilities. Some alternate modeling of the noise has been proposed, notably, a Student's t-distribution [10][11][12][13][14] to take into account the eventual presence of outliers.

One of the main limitations to the sensitivity of radio telescopes is Radio Frequency Interferences (RFI) which are mostly due to man-made radio waves. RFI can come from various sources, from telecommunication signals to high voltage transmission lines. Moreover, the generalization of large scale telecommunication networks and the growing use of satellites makes the modeling and mitigation of RFI a significant challenge in array signal processing. As a matter of fact, despite the efforts that are made to build such instruments in radio-quiet zones (in which radio transmissions are rare or restricted), man-made radio waves still have a significant impact in radio astronomy observation [15]. Likewise, the new generation of large-scale radio interferometers can detect weaker signals thanks to sensors being more and more sensitive, increasing the chances of intercepting RFI signals. The presence of a strong RFI in a radio interferometer signal corrupts the data measured by the affected sensors. Such outliers are usually flagged in the raw data using a generalized likelihood ratio test [16] or by a manual inspection of the data to check for its consistency. Various methods to perform spatial filtering on the correlation matrix have been proposed [17]. They are particularly efficient when the power of the RFI is sufficiently large so that the subspace of the interfering signal is identifiable. However, they lose in performance when the power of the RFI decreases. Yet, the presence of RFI still significantly impact the calibration solutions making the data unusable for scientific purposes [18].

In this paper, we propose a unified model from which we derive a maximum likelihood estimate of the calibration solutions as well as the RFI subspace, allowing robust calibration in the presence of RFI. By modeling accurately the RFI contribution to the visibilities, we improve the robustness of the calibration process for weak and medium RFI. Considering RFI as man-made high-frequency electromagnetic waves [19][17], we prove that the subspace of the perturbing RFI yields a low-rank structured perturbation in the visibilities. Consequently, we build a model for multi-frequency observations using the formalism of mixed effect model [20]. We consider a model composed of fixed effects representing the observed celestial sources and random effects which account for the possible presence of RFI exclusively on a few frequency channels. We achieve maximum likelihood estimation for the introduced model by using the Space Alternating Generalized EM (SAGE) [21]. The proposed model is described in section 2. In section 3, we introduce a SAGE-based algorithm used to perform maximum likelihood estimation for the proposed model. Finally, numerical results are presented in section 4.

Notation : Matrices are represented in bold uppercase and vectors in bold lowercase. The expectation operator is denoted $\mathbb{E}_{.}[.]$. The transpose operator is noted ^{T}, the complex conjugate * and the complex conjugate transpose (Hermitian) ^{H}. $\otimes$ denotes the Kronecker operator. The vectorization of a matrix, vec(.), is done by stacking its columns in a vector and its inverse operator is denoted unvec(.). $\Re\{.\}$ and $\Im\{.\}$ respectively refer to the real and imaginary part of a complex. The cardinality of a set is denoted Card(.). diag(.) creates a block-diagonal matrix from its matrix argument, rank(.) computes the rank of a matrix, Tr(.) refers to the trace of a matrix, and $|.|$ to its determinant. The Frobenius norm is denoted $\|.\|_2^2$. $\sim$ "means distributed as", $\overset{d}{=}$ stands for "shares the same distribution" and $\xrightarrow{d}$ denotes convergence in distribution. $\mathcal{CN}(.,.)$ refers to the complex normal distribution, $\mathcal{GCN}(.,.,.)$ to the generalized complex normal distribution, $\mathcal{W}(.,.)$ to the Wishart distribution and $\mathcal{U}(.,.)$ to the uniform distribution.

2. Data model

We consider a radio interferometer composed of P receivers observing the sky, each receiver being dual-polarized such that the signal received by the p^{th} sensor at time t can be expressed as

$$\boldsymbol{x}_p(t) = \begin{bmatrix} x_{p,H}(t) \\ x_{p,V}(t) \end{bmatrix} \tag{1}$$

where $x_{p,H}(t)$ and $x_{p,V}(t)$ represent the horizontal and vertical polarizations.
In the context of radio interferometer calibration, the sky is usually modeled as the sum of D independant strong calibrator sources, each source decomposed into two orthogonal polarizations [22][23][4][12]. The known sources, $\boldsymbol{s}(t)$, are assumed to be sampled from a circular complex gaussian law with a brightness (or coherency) matrix $\mathbf{C}$, ie,

$$\boldsymbol{s}(t) = \begin{bmatrix} s_H(t) \\ s_V(t) \end{bmatrix} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}), \text{ in which } \mathbf{C} = \begin{bmatrix} I+Q & U+jV \\ U-jV & I-Q \end{bmatrix}, \tag{2}$$

where I, Q, U, V are the Stokes parameters of the source caracterizing its power and polarization [3].

The perturbations encountered by the signal from its emission to its reception by the telescope sensors are modeled by a matrix product of the emitted signal with a perturbation matrix, namely the Jones matrix. Jones matrices can accurately model all kinds of distorsions of the signal (more details on the Jones framework can be found in [4][5][3]). A Jones matrix is specified for each receiver p and source i such that the signal received at the p^{th} sensor can be ideally written as

$$\boldsymbol{x}_p(t) = \sum_{i=1}^{D} \mathbf{J}_{i,p}(t)\boldsymbol{s}_i(t). \tag{3}$$

2.1. Radio interferometric data model

Radio interferometry uses estimates of the correlations between sensors to recover the flux of celestial sources [3][22]. By concatenating the signals from all sensors in a vector $\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{x}_1^T(t) & \dots & \boldsymbol{x}_P^T(t) \end{bmatrix}^T$ The correlation matrix of the array can be defined as a P^2 2×2 block matrix, $\mathbf{V} = \mathbb{E}\left[\boldsymbol{x}(t)\boldsymbol{x}^H(t)\right]$ where each block is the correlation between the p^{th} and the q^{th} receivers,

$$\mathbf{V}_{pq} = \mathbb{E}\left[\boldsymbol{x}_p(t)\boldsymbol{x}_q^H(t)\right] = \sum_{i=1}^{D} \mathbf{J}_{i,p}\mathbf{C}_i\mathbf{J}_{i,q}^H, \tag{4}$$

where $\mathbf{V}_{pq}$ is referred to as the visibility for the baseline formed by the p^{th} and q^{th} sensors and equation (4) is referred to as the Radio Interferometric Measurement Equation (RIME) [4].

In order to reduce the memory cost and complexity of calibration algorithms we use a vectorized version of the correlation matrix that contains only its lower diagonal entries. Let us note $\mathcal{B} = \{(p,q) \in [1,P]^2, p < q\}$ the set of unique pairs in the sensor array and $N_{\mathcal{B}} = \text{Card}(\mathcal{B}) = \frac{P(P-1)}{2}$.
For $(p,q) \in \mathcal{B}$, the visibility vector for the baseline formed by the p^{th} and q^{th} sensors is noted,

$$\boldsymbol{v}_{pq} = \text{vec}(\mathbf{V}_{pq}) = \sum_{i=1}^{D} \left(\mathbf{J}_{i,q}^* \otimes \mathbf{J}_{i,p}\right) \text{vec}(\mathbf{C}_i), \tag{5}$$

where the subscript pq refers to the pair of antennas formed by the p^{th} and the q^{th} antennas. All the visibility vectors $\{\boldsymbol{v}_{pq}\}_{(p,q)\in\mathcal{B}}$ are then concatenated in a vector

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_{12} \\ \vdots \\ \boldsymbol{v}_{(P-1)P} \end{bmatrix}. \tag{6}$$

In practice the visibilities are estimated by the sample covariance matrix of the measurement vectors for a short time interval,

$$\mathbf{R}_{pq} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_p(n)\boldsymbol{x}_q^H(n), \tag{7}$$

where N is the number of bins chosen such that the sources signals are stationnary and the Jones matrices constant. For N sufficiently large it is commonly accepted to model $\text{vec}(\mathbf{R}_{pq})$ as the sum of a deterministic signal and a white gaussian noise [3][24][25][22][12] as

$$\mathbf{r}_{pq} = \text{vec}(\mathbf{R}_{pq}) \overset{d}{=} \boldsymbol{v}_{pq} + \boldsymbol{n}_{pq}, \quad \boldsymbol{n}_{pq} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_4). \tag{8}$$

The additive noise $\boldsymbol{n}_{pq}$ is often referred to as the thermal noise, incorporating the noise induced by the antenna and cables that carry the signal to the correlator [3]. Moreover, $\boldsymbol{r}_{pq}$ being a realization of the sample covariance matrix associated to $\boldsymbol{v}_{pq}$ in an asymptotic regime, it is convenient to model it with a complex circular gaussian distribution [26].

The general problem of radio interferometric calibration aims at estimating the Jones matrices and the noise parameter given a set of measurements of the visibilities from known stellar sources. Taking into account all the baselines, equation (8) becomes

$$\boldsymbol{r} = \boldsymbol{v} + \boldsymbol{n}, \quad \mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{N_\mathcal{B}}), \tag{9}$$

where $\boldsymbol{r} = \begin{bmatrix} \boldsymbol{r}_{12} \\ \vdots \\ \boldsymbol{r}_{(P-1)P} \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_{12} \\ \vdots \\ \boldsymbol{v}_{(P-1)P} \end{bmatrix}$.

2.2. Radio Frequency Interferences

RFI are jamming signals of various forms (telecommunication, high voltage transmission lines) that can be modeled as additional independent unwanted sources with their specific Jones matrices [19]. The model expressed in (9) is not suited for the presence of RFI. In response, various works have proposed statistical models to take into account the possible presence of RFI [12][11]. Spatial filtering and statistical tests have been used to tackle the possible presence of strong RFI [17][27][15][16][25]. Such methods become less effective when the power of the RFI decreases. As an alternative we propose to perform calibration in presence of RFI, taking into account the RFI contribution to the visibilities in the calibration step. To this end, we propose in this section a statistical model for the measured visibilities corrupted by RFI from which a Maximum Likelihood Estimate (MLE) is derived in section 3.

Let us consider a vector of measured visibilities corrupted by L RFI sources, we propose to model the presence of RFI in the measured visibilities by a stochastic process with a low-rank structure, $\boldsymbol{r}^{RFI} = \mathbf{W}\boldsymbol{y}$. The RFI, the celestial sources, and the Gaussian noise are considered independent so that the presence of RFI is modeled independently and added to equation (9). The model on the measured visibilities thus becomes

$$\boldsymbol{r} = \boldsymbol{v} + \mathbf{W}\boldsymbol{y} + \boldsymbol{n}, \ \boldsymbol{y} \sim \mathcal{CN}(\text{vec}(\mathbf{I}), \mathbf{I}), \ \boldsymbol{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}), \ \rho \leq 4L^2, \tag{10}$$

where $\mathbf{W} \in \mathbb{C}^{4N_\mathcal{B} \times 4L^2}$, $\rho = \text{rank}(\mathbf{W})$. Specifically, the rational behind the modelling of the RFI in (10) is motivated by the physical modelling of such man-made signal [19]. Indeed the RFI component of the received signal at the p^{th} sensor can be expressed as follows,

$$\boldsymbol{s}_l^{RFI}(t) = \begin{bmatrix} s_{l,H}^{RFI}(t) \\ s_{l,V}^{RFI}(t) \end{bmatrix} \sim \mathcal{CN}(0, \mathbf{C}_l^{RFI}), \quad \boldsymbol{x}_p^{RFI}(t) = \sum_{l=1}^{L} \mathbf{J}_{l,p}^{RFI}(t) \boldsymbol{s}_l^{RFI}(t). \tag{11}$$

The form of the correlation matrix of the RFI sources for the P-sensors telescope can be expressed using the RIME as defined in the previous section,

$$\mathbf{V}_{pq}^{RFI} = \mathbb{E}\left[\boldsymbol{x}_p^{RFI}(t) \boldsymbol{x}_q^{RFI^H}(t)\right] = \sum_{i=1}^{L} \mathbf{J}_{i,p}^{RFI} \mathbf{C}_i^{RFI} \mathbf{J}_{i,q}^{RFI^H}. \tag{12}$$

These expressions can be arranged as matrix products,

$$\mathbf{V}_{pq}^{RFI} = \mathbf{J}_p^{RFI}\mathbf{C}^{RFI}\mathbf{J}_q^{RFI^H}, \tag{13}$$

$$\mathbf{V}^{RFI} = \mathbf{J}^{RFI}\mathbf{C}^{RFI}\mathbf{J}^{RFI^H}, \quad \mathbf{J}^{RFI} = \begin{bmatrix} \mathbf{J}_1^{RFI} \\ \vdots \\ \mathbf{J}_P^{RFI} \end{bmatrix}, \tag{14}$$

with $\mathbf{J}_p^{RFI} = \begin{bmatrix} \mathbf{J}_{1,p}^{RFI} & \dots & \mathbf{J}_{D,p}^{RFI} \end{bmatrix}$ and $\mathbf{C}^{RFI} = \text{diag}(\mathbf{C}_1^{RFI}, \dots, \mathbf{C}_D^{RFI})$

Let us note $\mathbf{R}^{RFI}$ the sample covariance estimate associated to $\mathbf{V}^{RFI}$. By definition $\mathbf{R}^{RFI}$ follows a Wishart distribution of degree N and mean $\mathbf{V}^{RFI} = \mathbf{A}\mathbf{A}^H$, $\mathbf{A} = \mathbf{J}^{RFI}\mathbf{C}^{RFI^{1/2}}$, and can be written

$$\mathbf{R}^{RFI} = \mathbf{A}\mathbf{M}\mathbf{A}^H, \quad \mathbf{M} \sim \mathcal{W}(N, \mathbf{I}). \tag{15}$$

For $(p, q) \in \mathcal{B}$, the visibility matrix for the baseline formed by the p^{th} and q^{th} sensors can be expressed as,

$$\mathbf{R}_{pq}^{RFI} = \mathbf{A}_p\mathbf{M}\mathbf{A}_q^H, \quad \mathbf{A}_p = \mathbf{J}_p^{RFI}\mathbf{C}^{RFI^{1/2}}. \tag{16}$$

Applying the vec operation,

$$\boldsymbol{r}_{pq}^{RFI} = \text{vec}(\mathbf{R}_{pq}^{RFI}) = (\mathbf{A}_q^* \otimes \mathbf{A}_p)\text{vec}(\mathbf{M}). \tag{17}$$

Finally, all the visibilities induced by the RFI are concatenated in a vector to obtain the low-rank structure

$$\boldsymbol{r}^{RFI} = \mathbf{W}\text{vec}(\mathbf{M}), \text{ with } \mathbf{W} = \begin{bmatrix} \mathbf{A}_2^* \otimes \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_N^* \otimes \mathbf{A}_{N-1} \end{bmatrix} \text{ and } \mathbf{M} \sim \mathcal{W}(N, \mathbf{I}). \tag{18}$$

As the number of samples increases, $\mathbf{M}$ appears to be closer to the asymptotic distribution of Wishart which is a Generalized Complex Normal distribution [26]

$$\sqrt{N}\text{vec}(\mathbf{M} - \mathbf{I}) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{I}, \mathbf{I}). \tag{19}$$

Consequently, when considering sufficient samples so that the central limit theorem can be applied, equation (9) can be modified in (10) to accurately model the presence of RFI. The maximum value for the rank of $\mathbf{W}$ can be identified from equation (18), $\rho_{max} = 4L^2$.

2.3. Multi-frequency model

In the previous sections, we only considered one frequency channel. Let us consider measurements across F frequency channels. Following the work presented in [24][28], we model the frequency dependance of the Jones matrices by a polynom. Considering the f^{th} frequency channel, the p^{th} receiver and the i^{th} source, the Jones matrix $\mathbf{J}_{i,p}(f)$ is expressed as

$$\mathbf{J}_{i,p}(f) = \sum_{k=1}^{K} b_k(f)\mathbf{Z}_{i,p,k}, \quad b_k(f) = \left(\frac{f - f_0}{f_0}\right)^{k-1}, \tag{20}$$

where K is the polynomial order and f_0 the central frequency. The expression is simplified using $\mathbf{B}_f = [b_1(f), \dots, b_K(f)] \otimes \mathbf{I}_2$,

$$\mathbf{J}_{i,p}(f) = \mathbf{B}_f\mathbf{Z}_{i,p}, \quad \mathbf{Z}_{i,p} = \begin{bmatrix} \mathbf{Z}_{i,p,1} \\ \vdots \\ \mathbf{Z}_{i,p,K} \end{bmatrix}. \tag{21}$$

The block matrix containing all the polynomial coefficient is noted $\mathbf{Z}$. The ideal visibility vector for the f^{th} frequency channel and the baseline formed by the p^{th} and q^{th} antenna becomes

$$\boldsymbol{v}_{f,pq}(\mathbf{Z}) = \sum_{i=1}^{D}((\mathbf{B}_f^* \mathbf{Z}_{i,q}^*) \otimes (\mathbf{B}_f \mathbf{Z}_{i,p}))\text{vec}(\mathbf{C}_i) \quad \text{and } \boldsymbol{v}_f(\mathbf{Z}) = \begin{bmatrix} \boldsymbol{v}_{f,12}(\mathbf{Z}) \\ \vdots \\ \boldsymbol{v}_{f,(P-1)P}(\mathbf{Z}) \end{bmatrix}. \tag{22}$$

The model expressed in (10) can be written using the formalism of mixed effect model to take into account the diversity accross frequency channels. The presence of RFI in a given frequency channel as well as the thermal noise associated to the frequency channel are considered as random effects and modeled by $\mathbf{W}_f \boldsymbol{y}_f + \boldsymbol{n}_f$. On the other hand, the contribution of the calibrator sources to the visibilities, which frequency behavior is assumed known, is considered as fixed effect. The resulting model can be written as,

$$\boldsymbol{r}_f = \boldsymbol{v}_f(\mathbf{Z}) + \mathbf{W}_f \boldsymbol{y}_f + \boldsymbol{n}_f, \quad \boldsymbol{y}_f \sim \mathcal{CN}(\text{vec}(\mathbf{I}), \mathbf{I}) \text{ and } \boldsymbol{n}_f \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}) \tag{23}$$

Ideally, since RFI can be found on multiple frequency channels, a low-rank matrix $\mathbf{W}_f$ would be assigned to each frequency channel affected by RFI. To ease computation, we reduce the model dimension by considering a common low-rank matrix $\mathbf{W}$ that englobes the RFI structures for all the frequency channels. Thus, the model is factorized in a simpler mixed effect model as $\mathbf{W}_f = \sigma_f \mathbf{W}$, where $\mathbf{W}$ is shared across all frequencies (absorbing the RFI structure) and σ_f is a soft weight.

By concatenating the measured visibilities for all the frequency channels, we deduce the proposed model for RFI corrupted radio interferometric data.

$$\boldsymbol{r} = \boldsymbol{v}(\mathbf{Z}) + ((\sigma_1, \ldots, \sigma_F) \otimes \mathbf{W})\, \boldsymbol{y} + \boldsymbol{n}, \tag{24}$$

with $\mathbf{W} \in \mathbb{C}^{4N_\mathcal{B} \times M}$, $\boldsymbol{y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_F]$, $\forall f \in [1, F], \boldsymbol{y}_f \sim \mathcal{CN}(\text{vec}(\mathbf{I}), \mathbf{I})$ and $\boldsymbol{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. The number of columns, M, of the tall matrix, $\mathbf{W}$, corresponds to its maximal rank (and also the maximal rank of the matrix $\mathbf{W}\mathbf{W}^H$), so is set in order to ensure $\text{rank}(\mathbf{W}) \leq M \leq 4N_B$.

3. Maximum likelihood estimation using a SAGE algorithm

In this section, we present an algorithm that aims at estimating all the unknown parameters of the model specified by (24). The set of parameters is

$$\boldsymbol{\Theta} = \{\mathbf{Z}, \mathbf{W}, \sigma^2, \sigma_1, \ldots, \sigma_F\} \tag{25}$$

The parameter of interest for radio interferometer calibration is the set of polynomial coefficients $\mathbf{Z}$, used to calibrate the measurements. Given a data vector $\boldsymbol{r} \in \mathcal{C}^{1 \times 4N_\mathcal{B} F}$ of measured visibilities, the log-likelihood for the model specified by (24) is

$$\mathcal{L}(\boldsymbol{r}, \boldsymbol{\theta}) = -\log\left(|\sigma^2 \mathbf{I} + \boldsymbol{\Psi}\boldsymbol{\Psi}^H|\right) - (\boldsymbol{r} - \boldsymbol{v}(\mathbf{Z}))(\sigma^2 \mathbf{I} + \boldsymbol{\Psi}\boldsymbol{\Psi}^H)^{-1}(\boldsymbol{r} - \boldsymbol{v}(\mathbf{Z}))^H \tag{26}$$

where $\boldsymbol{\Psi} = (\sigma_1, ..., \sigma_F) \otimes \mathbf{W}$ and $\boldsymbol{\theta} = \left[\sigma^2, \text{vec}(\mathbf{Z}), \text{vec}\mathbf{W}, \sigma_1, ..., \sigma_F\right]$.

Direct maximization of the likelihood is not tractable, thus we propose to use a SAGE based algorithm to find the Maximum Likelihood Estimate (MLE) for the proposed model [21] [29]. The SAGE is a variant of the Expectation-Maximization (EM) algorithm. In its simplest form, the EM algorithm is an iterative approach to the computation of the MLE when observations can be viewed as incomplete data. There has been sufficient work on the convergence properties of the EM algorithm, though it is known that an EM procedure increases the likelihood at each iteration converging to at least a local maxima [30][31]. Yet, the EM algorithm is known for its slow convergence and possible difficult maximization step when the space of parameters is of high dimensions [32]. The SAGE algorithm proposed in [21] tackles those problems by using alternating hidden data spaces to sequentially update the parameter vector. In [23], the SAGE algorithm is used to separate signals along different directions while assuming a gaussian noise on the visibilities,

whereas, in this paper, the SAGE algorithm is used to separate celestial and RFI signals with a model that specifically considers the presence of RFI. In practice, the SAGE consists in defining multiple spaces of parameters and their corresponding complete data spaces in which EM procedures are performed.

Thereby, we define two sets of complete data from which we derive a SAGE procedure. For each hidden data space we compute the expectation of the log-likelihood of the complete data given the observed incomplete data in the *E-step* as a surrogate function that is maximized in the *M-step*. Thus the set of complete data is defined so that the likelihood of its associated parameters leads to closed form maximizations. The purpose of the proposed algorithm is to estimate the RFI subspace ($\mathbf{W}$ and $(\sigma_f)_{1\leq f\leq F}$) with the first hidden data space and to estimate the calibration parameters ($\mathbf{Z}$ and σ^2) with the second hidden data space.

The first set of complete data contains the measured visibilities as well as the RFI subspace,

$$\mathcal{X}_1 = \{\boldsymbol{r}, \boldsymbol{y}\}. \tag{27}$$

Its associated parameter vector defines the RFI subspace, $\boldsymbol{\theta}_1 = [\text{vec}(\mathbf{W}), \sigma_1, ..., \sigma_F]$. We choose the second set of complete data to be the source contributions to the measured visibilities and their respective thermal noise,

$$\mathcal{X}_2 = \{(\boldsymbol{u}_i)_{1\leq i\leq D} : \boldsymbol{u} = \boldsymbol{r} - \boldsymbol{\Psi}\boldsymbol{y} = \sum_{i=1}^{D} \boldsymbol{u}_i\}, \tag{28}$$

where $\boldsymbol{u}_i = \boldsymbol{v}_i(\mathbf{Z}) + \boldsymbol{n}_i$ with $\boldsymbol{v}_i(\mathbf{Z})$ being the i^{th} source visibilities and $\boldsymbol{n}_i \sim \mathcal{CN}(0, \beta_i\sigma^2\mathbf{I})$ st $\sum_{i=1}^{D}\beta_i = 1$ and $\beta_i > 0$. The calibration parameters, $\boldsymbol{\theta}_2 = \left[\text{vec}(\mathbf{Z}), \sigma^2\right]$, are updated using this second set of complete data. The corresponding surrogate functions can be expressed as,

$$\begin{aligned} Q_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1^{(m)}) &= \mathbb{E}_{\mathcal{X}_1|\boldsymbol{r};\boldsymbol{\theta}_1^{(m)}}\left[\log(p(\boldsymbol{r}, \boldsymbol{y}; \boldsymbol{\theta}_1))\right], \\ Q_2(\boldsymbol{\theta}_2, \boldsymbol{\theta}_2^{(m)}) &= \sum_{i=1}^{D} \mathbb{E}_{\mathcal{X}_2|\boldsymbol{r};\boldsymbol{\theta}_1^{(m)}}\left[\log(p(\boldsymbol{u}_i; \boldsymbol{\theta}_2))\right]. \end{aligned} \tag{29}$$

Proposition 1 *The surrogate function Q_1 reduces to*

$$\begin{aligned} Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = A - \log \mid \sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \mid \\ - \sum_{f=1}^{F}\Bigg(\frac{1}{\sigma^2}\|\boldsymbol{r}_f - \boldsymbol{v}_f - \sigma_f\mathbf{W}\hat{\boldsymbol{y}}_f\|_2^2 + \frac{\sigma_f^2}{\sigma^2}\,Tr(\mathbf{W}^H\mathbf{W}\hat{\boldsymbol{\Sigma}}_f) + Tr(\hat{\boldsymbol{\Sigma}}_f) \\ + \|\hat{\boldsymbol{y}}_f - vec(I)\|_2^2\Bigg) \end{aligned} \tag{30}$$

where A is a constant independant of the model parameters and,

$$\begin{aligned} \hat{\boldsymbol{y}}_f &= vec(\mathbf{I}) + \sigma_f^{(m)}\mathbf{W}^{(m)^H}(\sigma^{(m)^2}\mathbf{I} + \sigma_f^{(m)^2}\mathbf{W}^{(m)}\mathbf{W}^{(m)^H})^{-1}(\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}^{(m)})), \\ \hat{\boldsymbol{\Sigma}}_f &= \mathbf{I} - \sigma_f^{(m)}\mathbf{W}^{(m)^H}(\sigma^{(m)^2}\mathbf{I} + \sigma_f^{(m)^2}\mathbf{W}^{(m)}\mathbf{W}^{(m)^H})^{-1}\sigma_f^{(m)}\mathbf{W}^{(m)^H}. \end{aligned} \tag{31}$$

Proof : see Appendix A.

Maximization of Q_1 is done using Wirtinger derivatives [33] and leads to the following closed-form expressions (see Appendix B)

$$\sigma_f^{(m+1)} = \frac{\Re\Big((\boldsymbol{r}_f - \boldsymbol{v}_f)^H\mathbf{W}\hat{\boldsymbol{y}}_f\Big)}{\text{Tr}\big(\mathbf{W}^H\mathbf{W}(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)\big)}, \tag{32}$$

$$\mathbf{W}^{(m+1)} = \left(\sum_{f=1}^{F} \sigma_f \left(\hat{\boldsymbol{y}}_f (\boldsymbol{r}_f - \boldsymbol{v}_f)^H \right)^H \right) \left(\sum_{f=1}^{F} \sigma_f^2 \left(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f \hat{\boldsymbol{y}}_f^H \right)^H \right)^{-1}. \tag{33}$$

It is worth mentioning that there is an ambiguity between the updates of $\mathbf{W}$ and $\sigma_1, \ldots, \sigma_F$. In order to avoid any ambiguity in the estimation of $\mathbf{W}$ and $(\sigma_1, \ldots, \sigma_f)$, we choose to impose a unit norm to $\mathbf{W}$, $||\mathbf{W}||_2 = 1$,

$$\sigma_f^{(m+1)} = \|\mathbf{W}^{(m+1)}\|_2 \times \sigma_f^{(m+1)} \tag{34}$$

$$\mathbf{W}^{(m+1)} = \frac{\mathbf{W}^{(m+1)}}{\|\mathbf{W}^{(m+1)}\|_2}. \tag{35}$$

Proposition 2 *The surrogate function Q_2 reduces to*

$$Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{D} C_{u_i} - \log \mid \beta_i \sigma^2 \mathbf{I}_{4FN_{\mathcal{B}}} \mid - \frac{1}{\beta_i \sigma^2} \|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 - \frac{1}{\beta_i \sigma^2} Tr(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i}), \tag{36}$$

where C_{u_i} is a constant independant of the model parameters and,

$$\begin{aligned} \hat{\boldsymbol{u}}_i &= \boldsymbol{v}_i(\mathbf{Z}^{(m)}) + \beta_i \sigma^2 (\sigma^2 \mathbf{I} + \boldsymbol{\Psi}^{(m)} \boldsymbol{\Psi}^{(m)^H})^{-1} (\boldsymbol{r} - \boldsymbol{v}(\mathbf{Z}^{(m)})) \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i} &= \beta_i \sigma^2 \mathbf{I}_{4FN_{\mathcal{B}}} - \beta_i^2 \sigma^4 (\sigma^2 \mathbf{I} + \boldsymbol{\Psi}^{(m)} \boldsymbol{\Psi}^{(m)^H})^{-1}. \end{aligned} \tag{37}$$

Proof : see Appendix C.

Maximization of (36) using Wirtinger derivatives leads to the following expression for σ^2 (see Appendix D),

$$\sigma^{2^{(m+1)}} = \frac{1}{4FN_{\mathcal{B}}D} \sum_{i=1}^{D} \frac{1}{\beta_i} \left(\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 + \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i}) \right). \tag{38}$$

Q_2 is maximized over $\mathbf{Z}$ with a block coordinate descent [34], leading to the following update (see Appendix D),

$$\mathrm{vec}(\mathbf{Z}_{i,p}^{(m+1)^*}) = (\mathbf{M}_{i,p} + \tilde{\mathbf{M}}_{i,p})^{-1} \mathrm{vec}(\mathbf{T}_{i,p}), \tag{39}$$

with $\mathbf{M}_{i,p} = \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{F}_{i,f,q}^T \mathbf{F}_{i,f,q}^* \otimes \mathbf{B}_f^T \mathbf{B}_f^*$, $\tilde{\mathbf{M}}_{i,p} = \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{G}_{i,f,q}^T \mathbf{G}_{i,f,q}^* \otimes \mathbf{B}_f^T \mathbf{B}_f^*$, and
$\mathbf{T}_{i,p} = \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{B}_f^T \hat{\mathbf{U}}_{i,f,pq}^* \mathbf{F}_{i,f,q} + \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{B}_f^T \hat{\mathbf{U}}_{i,f,qp}^T \mathbf{G}_{i,f,q}$, where $\mathbf{F}_{i,f,q} = \mathbf{B}_f^* \mathbf{Z}_{i,q}^{(m+1)^*} \mathbf{C}_{i,f}^T$ and
$\mathbf{G}_{i,f,q} = \mathbf{B}_f^* \mathbf{Z}_{i,q}^{(m+1)^*} \mathbf{C}_{i,f}^*$.

The SAGE algorithm consists in applying the two SAGE steps iteratively until convergence. We consider that convergence is achieved when the variation of the log-likelihood between two iterations is less than a chosen threshold $\epsilon > 0$.

Algorithm 1 SAGE algorithm for multi-frequency calibration in presence of RFI

Input: ϵ, $N_{sources}$, $N_{sensors}$, $N_{frequencies}$, $\boldsymbol{v}$, $\boldsymbol{r}$, $\sigma^{2^{(0)}}$, $\mathbf{Z}^{(0)}$, $\mathbf{W}^{(0)}$, $\sigma_1^{(0)}, ..., \sigma_F^{(0)}$.
Output: σ^2, $\mathbf{Z}$, $\mathbf{W}$, σ_1, ..., σ_f

while $(\mathcal{L}(\boldsymbol{r}, \boldsymbol{\theta}^{(m+1)}) - \mathcal{L}(\boldsymbol{r}, \boldsymbol{\theta}^{(m)})) \geq \epsilon$ **do**
  Update $\hat{\boldsymbol{y}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}$ using (31)
  for f in $[1 : N_{frequencies}]$ **do**
    Update σ_f using (32)
  end for
  Update $\mathbf{W}$ using (33)
  Update $\sigma_1, \ldots, \sigma_F$ using (34)
  Update $\mathbf{W}$ using (35)
  for i in $[1 : N_{sources}]$ **do**
    Update $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i}$ using (37)
    for p in $[1 : N_{sensors}]$ **do**
      Update $\mathbf{Z}[i, p]$ using (39)
    end for
  end for
  Update σ^2 using (38)
end while
Return: (σ^2, $\mathbf{Z}$, $\mathbf{W}$, σ_1, ..., σ_f)

4. Numerical results

In this section, several simulations are presented to investigate the robustness of the proposed model and to study the performance of the proposed algorithm. To that end, we consider a radio interferometer of $P = 8$ receivers observing a sky composed of $D = 2$ calibrator sources along $F = 32$ frequency channels. The frequency channels and the central frequency are chosen such that $\frac{f - f_0}{f_0} \in [-1, 1]$. The sky is composed of two unpolarized calibrator sources of flux 100Jy and 50Jy. The matrix $\mathbf{Z}$ is generated using a uniform distribution, for $i \in [1, D]$, $p \in [1, P]$, $\Re(\mathbf{Z}_{i,p}) \sim \mathcal{U}(0, 1)$ and $\Im(\mathbf{Z}_{i,p}) \sim \mathcal{U}(0, 1)$. The maximum number of iterations for the various iterative algorithms used is set to 15 and 100 Monte Carlo repetitions are performed for a thermal noise SNR of 15dB. We compute for each run the Normalized Mean Square Error (NMSE) for the parameter of interest, NMSE $= \frac{1}{100} \sum_{j=1}^{100} \frac{\|\mathbf{Z} - \hat{\mathbf{Z}}_j\|_2^2}{\|\mathbf{Z}\|_2^2}$. We first show that our method outperforms standard projections methods [17] in the context of strong RFI. We then compare our algorithm to a method considering a robust noise model, namely the Student-t distribution [14]. Finally, we study the convergence rate and the influence of the rank of $\mathbf{W}$ for the proposed algorithm.

In the presence of strong RFI on specific frequency channels, it is possible to project out the RFI subspace by using the highest eigenvalues of the correlation matrix [17], reconstructing the filtered visibilities free from RFI. Calibration can then be performed on the filtered visibilities. We compare our proposed calibration algorithm that does not need to filter out the RFI to a classical calibration algorithm performed on filtered visibilities [23]. A sky corrupted by 2 RFI, with Stokes parameter [100 10 50 30] and [50 0 0 0], is simulated.

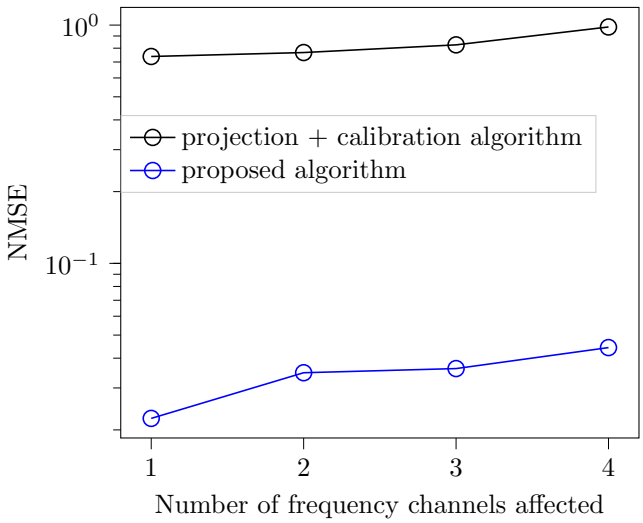

Figure 1: Evolution of the NMSE of $\mathbf{Z}$ versus the number of neighboring frequency band affected by RFI

In figure 1 we plot the evolution of the NMSE for the parameter of interest versus the number of frequency channels affected by RFI. It appears that the proposed calibration algorithm leads to better performance. While filtering out the RFI, a part of the sources might be attenuated, leading to possible biased calibration solutions, whereas the proposed algorithm calibrates on the original sources signals while considering the RFI subspace.

The Student-t distribution has proven itself to be a distribution of choice to model the possible presence of outliers in the visibilities [14][11][12]. Thus we compare our proposed model to the model presented in [14], considering a Student-t distribution as additive noise on the visibilities. Figure 2 shows the evolution of the NMSE versus the power of the RFI when only a few frequency channels are corrupted by interference. RFI with a power that ranges from -10dB to 10dB are added on 10% and 30% of the frequency channels. Figures 2.b and 2.d present the visibility amplitude across all the frequency bands for one baseline when respectively 10% and 30% of the frequency channels are affected by RFI. The proposed method displays, in figure 2.a and 2.c, performances that are similar to the state of the art methods, considering a Student-t distribution as additive noise, despite that such methods are prone to good behavior in this specific context.

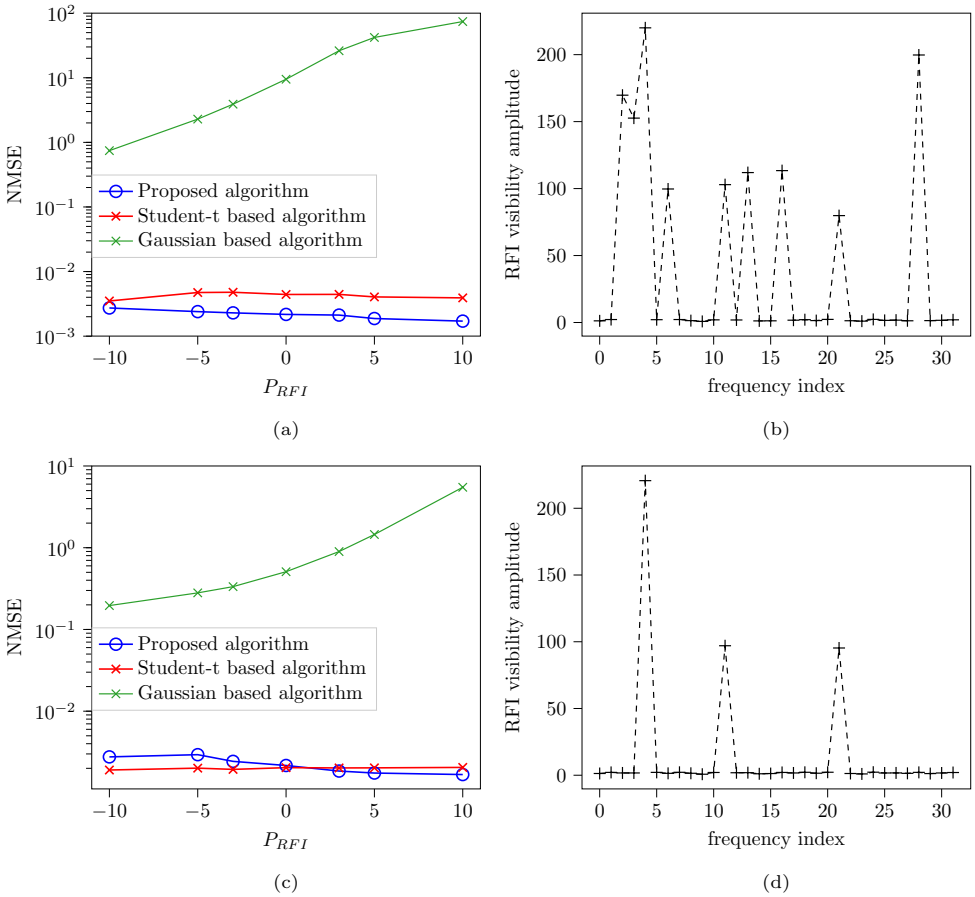

Figure 2: Evolution of the NMSE of $\mathbf{Z}$ versus the power of the strong RFI

In reality, low RFI are predominant across multiple frequency channels [27]. Subsequently, we simulated a sky perturbated by low RFI (-15dB) on each frequency channel. Additionally, strong RFI are added to 10% of the frequency bands with a power that ranges from -10dB to 10dB as illustrated in figure 3.b for one baseline.

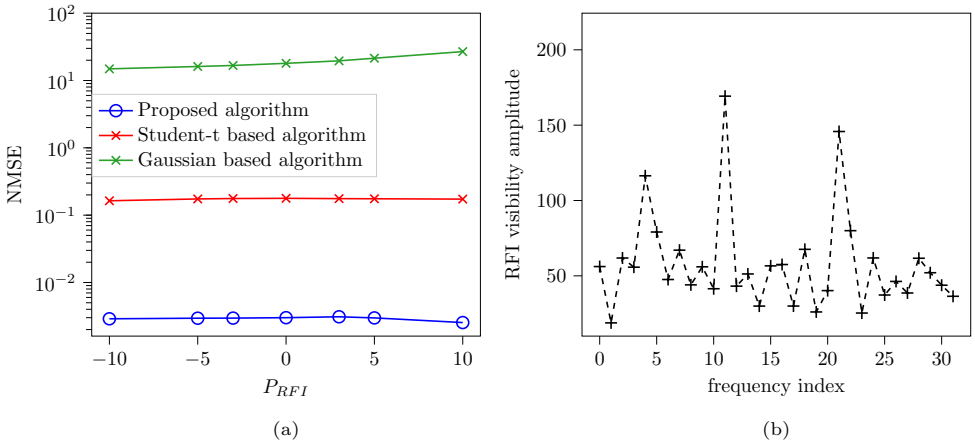

Figure 3: Evolution of the NMSE of $\mathbf{Z}$ versus the power of the strong RFI with additional low RFI on each frequency band

We plot in figure 3.a the evolution of the NMSE versus the power of the strong RFI. A significant gain in performance can be noticed for the proposed algorithm compared to the state-of-the-art. The contribution of our proposed model lies in taking into account the structure of the RFI, adding robustness to the presence of RFI of various forms. In figure 4, the evolution of the NMSE for $\mathbf{Z}$ versus the chosen rank is displayed.

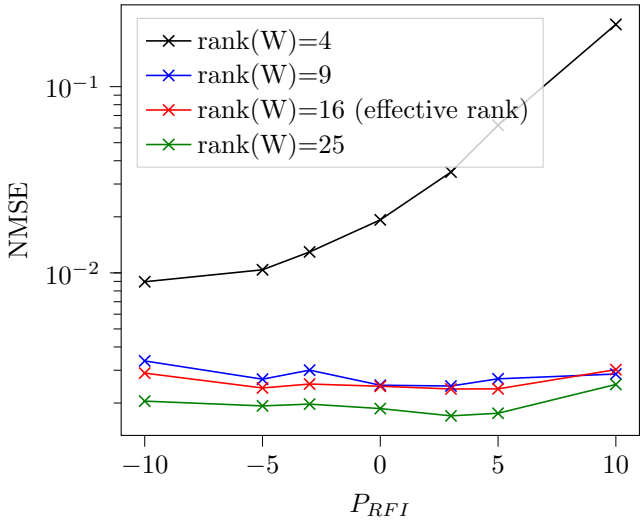

Figure 4: Evolution of the NMSE of $\mathbf{Z}$ versus the power of the RFI for multiple choice of the rank

The effective rank used to simulate the visibilities is $\rho_{\text{eff}} = 16$. It can be noticed that a slight misjudgment of the rank does not change the performance of the algorithm. These results illustrate the fact that the rank of $\mathbf{W}$ constrains the rank of the clutter noise covariance matrix, $\Phi = \mathbf{W}\mathbf{W}^H$. If the rank of $\mathbf{W}$ is highly underestimated, the clutter noise covariance matrix will also have a rank highly underestimated, leading to poor reconstruction. Conversely, since there is no constraint in the rank of the estimated RFI structure matrix, $\mathbf{W}$, the clutter noise covariance matrix can still be correctly estimated when the rank of $\mathbf{W}$ is overestimated. This particular case might lead to overfitting in the estimate of $\mathbf{W}$. However, in practice, we observe that our parameter of interest, $\mathbf{Z}$, is correctly estimated.

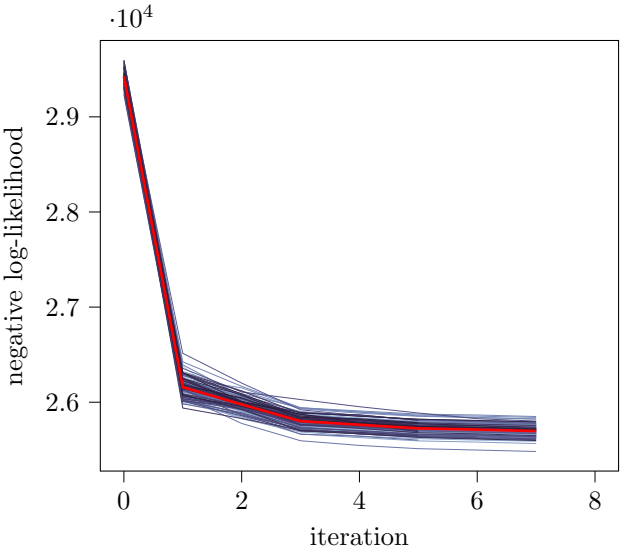

Figure 5: Evolution of the negative log likelihood versus the numnber of iteration for 100 realizations

Figure 5 presents the evolution of the negative log likelihood versus the number of iterations for 100 realizations. It appears that only a few iterations are needed for the proposed algorithm to converge, leading to a reasonable computational cost.

5. Conclusion

In this paper, we have first proposed a model for radio interferometer calibration that takes into account the presence of radiofrequency interferences, using a low-rank structured noise. The calibration parameters are obtained by a maximum likelihood estimation formulated thanks to a Space Alternating Generalized Expectation-Maximization algorithm. This latter leads to closed form updates for the model parameters thanks to the choice of two judicious complete data sets. Numerical simulations show that the proposed estimator is robust to the presence of radiofrequency interferences and demonstrate a gain in performance compared to the state-of-the-art with reasonable computational time. Moreover, simulations show that a misjudgment of the rank of the design matrix $\mathbf{W}$ does not impact the performance of the algorithm.

Appendix A. Proof of Proposition 1

We seek to compute $Q_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) = \mathbb{E}_{\mathcal{X}_{c,1}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\log(p(\boldsymbol{r}, \boldsymbol{y}; \boldsymbol{\theta}))\right]$. Since $\boldsymbol{r}$ and $\boldsymbol{y}$ can be decomposed into mutually independant frequency vectors, $\boldsymbol{r} = [\boldsymbol{r}_1^T, ..., \boldsymbol{r}_F^T]$ and $\boldsymbol{y} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_F^T]$ where $\boldsymbol{r}_f = \boldsymbol{v}_f + \sigma_f \mathbf{W}\boldsymbol{y}_f + \boldsymbol{n}_f$, Q_1 can be expressed as

$$Q_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) = \sum_{f=1}^{F} \mathbb{E}_{\mathcal{X}_{c,1}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\log(p(\boldsymbol{r}_f, \boldsymbol{y}_f; \boldsymbol{\theta}))\right] \tag{A.1}$$

It is straightforward from the model definition that $\boldsymbol{r}_f \mid \boldsymbol{y}_f \sim \mathcal{CN}(\boldsymbol{v}_f(\mathbf{Z}) + \sigma_f \mathbf{W}\boldsymbol{y},\ \sigma^2\mathbf{I}_{4B})$ and $\boldsymbol{y}_f \sim \mathcal{CN}(\mathrm{vec}(\mathbf{I}),\ \mathbf{I})$. Thus the log-likelihoods of $\boldsymbol{r}_f \mid \boldsymbol{y}_f$ and $\boldsymbol{y}_f$ can respectively be written as

$$\begin{aligned}
\log(p(\boldsymbol{r}_f \mid \boldsymbol{y}_f; \boldsymbol{\theta})) &= -4FN_{\mathcal{B}}\log(\pi) - \log \mid \sigma^2\mathbf{I} \mid -\frac{1}{\sigma^2}(\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}) - \sigma_f\mathbf{W}\boldsymbol{y})^H(\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}) - \sigma_f\mathbf{W}\boldsymbol{y}_f), \\
\log(p(\boldsymbol{y}_f; \boldsymbol{\theta})) &= -\rho\log(\pi) - \log \mid \mathbf{I}_\rho \mid \ - \sum_{f=1}^{F}(\boldsymbol{y}_f - \mathrm{vec}(\mathbf{I}))^H(\boldsymbol{y}_f - \mathrm{vec}(\mathbf{I})).
\end{aligned} \tag{A.2}$$

Thus, $Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})$ can be expressed as,

$$\begin{aligned}
Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) &= \sum_{f=1}^{F} \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\log(p(\boldsymbol{r}_f \mid \boldsymbol{y}_f; \boldsymbol{\theta})) + \log(p(\boldsymbol{y}_f; \boldsymbol{\theta}))\right] \\
&= A + \sum_{f=1}^{F} - \log \mid \sigma^2\mathbf{I}_{4B} \mid \\
&\quad - \frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[(\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}) - \sigma_f\mathbf{W}\boldsymbol{y}_f)^H(\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}) - \sigma_f\mathbf{W}\boldsymbol{y}_f)\right] \\
&\quad - \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[(\boldsymbol{y}_f - \mathrm{vec}(\mathbf{I}))^H(\boldsymbol{y}_f - \mathrm{vec}(\mathbf{I}))\right] \\
&= A - F\log \mid \sigma^2\mathbf{I}_{4B} \mid + \sum_{f=1}^{F} \frac{2\sigma_f}{\sigma^2}\Re((\boldsymbol{r}_f - \boldsymbol{v}_f)^H\mathbf{W}\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f\right]) \\
&\quad - \frac{1}{\sigma^2}(\boldsymbol{r}_f - \boldsymbol{v}_f)^H(\boldsymbol{r}_f - \boldsymbol{v}_f) - \frac{\sigma_f^2}{\sigma^2}\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f^H\mathbf{W}^H\mathbf{W}\boldsymbol{y}_f\right] \\
&\quad - \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f^H\boldsymbol{y}_f\right] + 2\mathrm{vec}(\mathbf{I})^H\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f\right] - \mathrm{vec}(\mathbf{I})^H\mathrm{vec}(\mathbf{I})
\end{aligned} \tag{A.3}$$

where A is a constant independant of the model parameters First the expression of the mean and covariance of $\boldsymbol{y}_f$ are derived, respectively $\hat{\boldsymbol{y}}_f = \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f\right]$ and $\hat{\boldsymbol{\Sigma}}_f = \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[(\boldsymbol{y}_f - \hat{\boldsymbol{y}}_f)(\boldsymbol{y}_f - \hat{\boldsymbol{y}}_f)^H\right]$. We consider $\boldsymbol{z} = [\boldsymbol{y}_f, \boldsymbol{r}_f]^T$, $\boldsymbol{z}$ being a multivariate Gaussian in which the mean of $\boldsymbol{y}_f$ and $\boldsymbol{r}_f$ are respectively

$$\boldsymbol{\mu}_{\boldsymbol{y}_f} = \mathrm{vec}(\mathbf{I}), \quad \boldsymbol{\mu}_{\boldsymbol{r}_f} = \boldsymbol{v}_f(\mathbf{Z}). \tag{A.4}$$

The Covariance block matrices $\boldsymbol{\Sigma}_{\boldsymbol{yy}}$, $\boldsymbol{\Sigma}_{\boldsymbol{rr}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{ry}}$ are :

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{yy}} &= \mathbb{E}\left[(\boldsymbol{y}_f - \boldsymbol{\mu}_{\boldsymbol{y}_f})(\boldsymbol{y}_f - \boldsymbol{\mu}_{\boldsymbol{y}_f})^H\right] = \mathbf{I} \\
\boldsymbol{\Sigma}_{\boldsymbol{rr}} &= \mathbb{E}_{\boldsymbol{r}}\left[(\boldsymbol{r}_f - \boldsymbol{\mu}_{\boldsymbol{r}_f})(\boldsymbol{r}_f - \boldsymbol{\mu}_{\boldsymbol{r}_f})^H\right] = {\sigma^{(m)}}^2\mathbf{I}_{4B} + \sigma_f^{2^{(m)}}\mathbf{W}^{(m)}\mathbf{W}^{(m)^H} \\
\boldsymbol{\Sigma}_{\boldsymbol{yr}} &= \boldsymbol{\Sigma}_{\boldsymbol{ry}}^H = \mathbb{E}\left[(\boldsymbol{y}_f - \boldsymbol{\mu}_{\boldsymbol{y}_f})(\boldsymbol{r}_f - \boldsymbol{\mu}_{\boldsymbol{r}_f})^H\right] \\
&= \mathbb{E}\left[(\boldsymbol{y}_f - \mathrm{vec}(\mathbf{I}))(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right] \\
&= \mathbb{E}\left[(\boldsymbol{y}_f - \mathrm{vec}(\mathbf{I}))(\sigma_f^{(m)}\mathbf{W}^{(m)}\boldsymbol{y}_f + \boldsymbol{n}_f)^H\right] \\
&= \mathbb{E}\left[\boldsymbol{y}_f(\sigma_f^{(m)}\mathbf{W}^{(m)}\boldsymbol{y}_f)^H\right] + \mathbb{E}\left[\boldsymbol{y}_f\boldsymbol{n}_f^H\right] - \mathrm{vec}(\mathbf{I})(\sigma_f^{(m)}\mathbf{W}^{(m)}\mathbb{E}\left[\boldsymbol{y}_f\right])^H - \mathrm{vec}(\mathbf{I})\mathbb{E}\left[\boldsymbol{n}_f^H\right]
\end{aligned} \tag{A.5}$$

Since $\boldsymbol{y}_f$ and $\boldsymbol{n}_f$ are considered independently distributed and $\mathbb{E}\left[\boldsymbol{n}_f\right] = \mathbf{0}$, $\mathbb{E}\left[\boldsymbol{y}_f \boldsymbol{n}_f^H\right] = \mathbf{0}$ and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{yr}} = \boldsymbol{\Sigma}_{\boldsymbol{ry}}^H &= \mathbb{E}\left[\boldsymbol{y}_f (\sigma_f^{(m)} \mathbf{W}^{(m)} \boldsymbol{y}_f)^H\right] - \mathrm{vec}(\mathbf{I})(\sigma_f^{(m)} \mathbf{W}^{(m)} \mathbb{E}\left[\boldsymbol{y}_f\right])^H \\ &= (\mathbb{E}\left[\boldsymbol{y}_f \boldsymbol{y}_f^H\right] - \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^H)\sigma_f^{(m)} \mathbf{W}^{(m)^H} = \sigma_f^{(m)} \mathbf{W}^{(m)^H}\end{aligned} \tag{A.6}$$

Using [35, p36] the analytical expression for $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{y}}$ are deduced,

$$\begin{aligned}\hat{\boldsymbol{y}}_f &= \boldsymbol{\mu}_{\boldsymbol{y}_f} + \boldsymbol{\Sigma}_{\boldsymbol{yr}} \boldsymbol{\Sigma}_{\boldsymbol{rr}}^{-1} (\boldsymbol{r}_f - \boldsymbol{\mu}_{\boldsymbol{r}_f}) \\ &= \mathrm{vec}(\mathbf{I}) + \sigma_f^{(m)} \mathbf{W}^{(m)^H} (\sigma^{(m)^2} \mathbf{I}_{4B} + \sigma_f^{2^{(m)}} \mathbf{W}^{(m)} \mathbf{W}^{(m)^H})^{-1} (\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}^{(m)})) \\ \hat{\boldsymbol{\Sigma}}_f &= \boldsymbol{\Sigma}_{\boldsymbol{yy}} - \boldsymbol{\Sigma}_{\boldsymbol{yr}} \boldsymbol{\Sigma}_{\boldsymbol{rr}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{ry}} \\ &= \mathbf{I} - \sigma_f^{(m)} \mathbf{W}^{(m)^H} (\sigma^{(m)^2} \mathbf{I}_{4B} + \sigma_f^{2^{(m)}} \mathbf{W}^{(m)} \mathbf{W}^{(m)^H})^{-1} \sigma_f \mathbf{W}^{(m)}.\end{aligned} \tag{A.7}$$

$\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f^H \boldsymbol{y}_f\right]$ and $\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f^H (\sigma_f^2 \mathbf{W}^H \mathbf{W}) \boldsymbol{y}_f\right]$ are then derived,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f^H \boldsymbol{y}_f\right] &= \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\mathrm{Tr}(\boldsymbol{y}_f \boldsymbol{y}_f^H)\right] = \mathrm{Tr}(\mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f \boldsymbol{y}_f^H\right]) \\ &= \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f \hat{\boldsymbol{y}}_f^H) \\ &= \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_f) + \hat{\boldsymbol{y}}_f^H \hat{\boldsymbol{y}}_f \\ \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f^H (\sigma_f^2 \mathbf{W}^H \mathbf{W}) \boldsymbol{y}\right] &= \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\mathrm{Tr}((\sigma_f^2 \mathbf{W}^H \mathbf{W}) \boldsymbol{y}_f \boldsymbol{y}_f^H)\right] \\ &= \mathrm{Tr}((\sigma_f^2 \mathbf{W}^H \mathbf{W}) \mathbb{E}_{\boldsymbol{y}|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{y}_f \boldsymbol{y}_f^H\right]) \\ &= \mathrm{Tr}((\sigma_f^2 \mathbf{W}^H \mathbf{W}) \hat{\boldsymbol{\Sigma}}_f + (\sigma_f^2 \mathbf{W}^H \mathbf{W}) \hat{\boldsymbol{y}}_f \hat{\boldsymbol{y}}_f^H) \\ &= \mathrm{Tr}((\sigma_f^2 \mathbf{W}^H \mathbf{W}) \hat{\boldsymbol{\Sigma}}_f) + \hat{\boldsymbol{y}}_f^H (\sigma_f^2 \mathbf{W}^H \mathbf{W}) \hat{\boldsymbol{y}}_f\end{aligned} \tag{A.8}$$

Finally Q_1 reads,

$$\begin{aligned}Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) &= A - F \log \mid \sigma^2 \mathbf{I}_{4B} \mid + \sum_{f=1}^{F} \frac{2}{\sigma^2} \Re((\boldsymbol{r}_f - \boldsymbol{v}_f)^H \sigma_f \mathbf{W} \hat{\boldsymbol{y}}_f) \\ &\quad - \frac{1}{\sigma^2} (\boldsymbol{r}_f - \boldsymbol{v}_f)^H (\boldsymbol{r}_f - \boldsymbol{v}_f) - \frac{\sigma_f^2}{\sigma^2} (\mathrm{Tr}(\mathbf{W}^H \mathbf{W} \hat{\boldsymbol{\Sigma}}_f) + \hat{\boldsymbol{y}}_f^H (\sigma_f^2 \mathbf{W}^H \mathbf{W}) \hat{\boldsymbol{y}}_f) \\ &\quad - \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_f) - \hat{\boldsymbol{y}}_f^H \hat{\boldsymbol{y}}_f + 2\mathrm{vec}(\mathbf{I})^H \hat{\boldsymbol{y}}_f - \mathrm{vec}(\mathbf{I})^H \mathrm{vec}(\mathbf{I}) \\ &= A - F \log \mid \sigma^2 \mathbf{I}_{4B} \mid - \sum_f \Bigg(\frac{1}{\sigma^2} \| \boldsymbol{r}_f - \boldsymbol{v}_f - \sigma_f \mathbf{W} \hat{\boldsymbol{y}}_f \|_2^2 \\ &\quad + \| \hat{\boldsymbol{y}}_f - \mathrm{vec}(\mathbf{I}) \|_2^2 + \frac{\sigma_f^2}{\sigma^2} \mathrm{Tr}(\mathbf{W}^H \mathbf{W}^H \hat{\boldsymbol{\Sigma}}_f) + \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_f) \Bigg)\end{aligned} \tag{A.9}$$

Appendix B. Maximization of $Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})$

We compute the derivative of $Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})$ w.r.t σ_f, for $f \in [1; F]$:

$$\frac{\partial Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})}{\partial \sigma_f} = -\frac{2\sigma_f}{\sigma^2} \mathrm{Tr}\big(\mathbf{W}^H \mathbf{W} (\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f^H \hat{\boldsymbol{y}}_f)\big) + \frac{2}{\sigma^2} \Re\Big((\boldsymbol{r}_f - \boldsymbol{v}_f(\mathbf{Z}))^H \mathbf{W} \hat{\boldsymbol{y}}_f \Big) \tag{B.1}$$

We solve for $\dfrac{\partial Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})}{\partial \sigma_f} = 0$ and we obtain

$$\hat{\sigma}_f = \frac{\Re\Big((\boldsymbol{r}_f - \boldsymbol{v}_f)^H \mathbf{W}\hat{\boldsymbol{y}}_f\Big)}{\mathrm{Tr}\big(\mathbf{W}^H\mathbf{W}(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)\big)} \tag{B.2}$$

$Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})$ can be written in function of $\mathbf{W}$,

$$\begin{aligned} Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) &= A + \frac{1}{\sigma^2}\sum_{f=1}^{F}\Big(2\Re\big((\boldsymbol{r}_f - \boldsymbol{v}_f)^H\sigma_f\mathbf{W}\hat{\boldsymbol{y}}_f\big) - \mathrm{Tr}(\sigma_f^2\mathbf{W}^H\mathbf{W}\hat{\boldsymbol{\Sigma}}_f) - \hat{\boldsymbol{y}}_f^H\sigma_f^2\mathbf{W}^H\mathbf{W}\hat{\boldsymbol{y}}_f\Big) \\ &= A + \frac{1}{\sigma^2}\sum_{f=1}^{F}\Big(\mathrm{Tr}\left(\mathbf{W}\sigma_f\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right) + \mathrm{Tr}\left(\mathbf{W}^H\sigma_f\hat{\boldsymbol{y}}_f^H(\boldsymbol{r}_f - \boldsymbol{v}_f)\right) \\ &\quad - \mathrm{Tr}\left(\sigma_f^2\mathbf{W}^H\mathbf{W}(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)\right)\Big) \end{aligned} \tag{B.3}$$

The derivatives of Q_1 w.r.t $\mathbf{W}$ is computed using Wirtinger derivative [33]

$$\begin{aligned} \frac{\partial}{\partial\mathbf{W}}\Big(\mathrm{Tr}\left(\mathbf{W}\sigma_f\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right)\Big) &= \left(\sigma_f\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right)^T \\ \frac{\partial}{\partial\mathbf{W}}\Big(\mathrm{Tr}\left(\sigma_f\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \mathbf{v_f})^H\mathbf{W}^H\right)\Big) &= \mathbf{0}_{4B\times r} \\ \frac{\partial}{\partial\mathbf{W}}\Big(\mathrm{Tr}\left(\sigma_f^2\mathbf{W}^H\mathbf{W}(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)\right)\Big) &= \sigma_f^2\mathbf{W}^*(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)^T. \end{aligned} \tag{B.4}$$

Consequently,

$$\begin{aligned} \frac{\partial}{\partial\mathbf{W}}(Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})) &= \sum_{f=1}^{F}\Big(\left(\sigma_f\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right)^T - \sigma_f^2\mathbf{W}^*(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)^T\Big) \\ &= \sum_{f=1}^{F}\sigma_f\left(\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right)^T - \mathbf{W}^*\sum_{f=1}^{F}\sigma_f^2\left(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H\right)^T \end{aligned} \tag{B.5}$$

Finally, solving for $\dfrac{\partial Q_1(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})}{\partial\mathbf{W}} = 0$ leads to,

$$\begin{aligned} \mathbf{W}^* &= \Big(\sum_{f=1}^{F}\sigma_f\left(\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \boldsymbol{v}_f)^H\right)^T\Big)\Big(\sum_{f=1}^{F}\sigma_f^2(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H)^T\Big)^{-1} \\ \mathbf{W} &= \Big(\sum_{f=1}^{F}\sigma_f\left(\hat{\boldsymbol{y}}_f(\boldsymbol{r}_f - \mathbf{v_f})^H\right)^H\Big)\Big(\sum_{f=1}^{F}\sigma_f^2\left(\hat{\boldsymbol{\Sigma}}_f + \hat{\boldsymbol{y}}_f\hat{\boldsymbol{y}}_f^H\right)^H\Big)^{-1} \end{aligned} \tag{B.6}$$

Appendix C. Proof of proposition 2

From (28),

$$\boldsymbol{u}_i \sim CN(\boldsymbol{v}_i(\mathbf{Z}), \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}}) \text{ with } \sum_{i=1}^{D}\beta_i = 1$$

The log-likelihood of the complete data can be written

$$\log(p(\boldsymbol{u}_i;\boldsymbol{\theta})) = C_{u_i} - \log \mid \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \mid -\frac{1}{\beta_i\sigma^2}(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z}))^H(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z})). \tag{C.1}$$

Thus, Q_2 can be expressed as follows,

$$\begin{aligned}
Q_2(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) &= \sum_{i=1}^{D} \mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r};\boldsymbol{\theta}^{(m)}} \left[\log(p(\boldsymbol{u}_i;\boldsymbol{\theta}))\right] \\
&= \sum_{i=1}^{D} \mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r}} \left[C_{u_i} - \log \mid \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \mid -\frac{1}{\beta_i\sigma^2}(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z}))^H(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z}))\right] \\
&= \sum_{i=1}^{D} C_{u_i} - \log \mid \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \mid -\frac{1}{\beta_i\sigma^2}\mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r}} \left[(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z}))^H(\boldsymbol{r}_i - \boldsymbol{v}_i(\mathbf{Z}))\right] \\
&= \sum_{i=1}^{D} C_{u_i} - \log \mid \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \mid -\frac{1}{\beta_i\sigma^2}\mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r}} \left[\boldsymbol{u}_i^H\boldsymbol{u}_i\right] + \frac{2}{\beta_i\sigma^2}\Re\Big(\boldsymbol{v}_i(\mathbf{Z})^H\mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r}}\left[\boldsymbol{u}_i\right]\Big) \\
&- \frac{1}{\beta_i\sigma^2}\boldsymbol{v}_i^H\boldsymbol{v}_i.
\end{aligned} \tag{C.2}$$

Let us note $\hat{\boldsymbol{u}}_i = \mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[\boldsymbol{u}_i\right]$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i} = \mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r};\boldsymbol{\theta}^{(m)}}\left[(\boldsymbol{u}_i - \hat{\boldsymbol{u}}_i)(\boldsymbol{u}_i - \hat{\boldsymbol{u}}_i)^H\right]$ the mean and covariance of the complete data given the oberved data and the parameter vector. We consider $\boldsymbol{z}_i = [\boldsymbol{u}_i, \boldsymbol{r}_i]$ being a multivariate Gaussian law. The mean of $\boldsymbol{u}_i$ and $\boldsymbol{r}$ are respectively

$$\boldsymbol{\mu}_{\boldsymbol{u_i}} = \boldsymbol{v}_i(\mathbf{Z}),\ \boldsymbol{\mu}_{\boldsymbol{r}} = \boldsymbol{v}(\mathbf{Z}).$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{u}_i} &= \mathbb{E}_{\boldsymbol{u}_i}\left[(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})^H\right] = \mathbb{E}_{\boldsymbol{u}_i}\left[(\boldsymbol{u}_i - \boldsymbol{v}_i(Z))(\boldsymbol{u}_i - \boldsymbol{v}_i(Z))^H\right] = \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \\
\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{r}} &= \mathbb{E}_{\boldsymbol{r}}\left[(\boldsymbol{r} - \boldsymbol{\mu}_{\boldsymbol{r}})(\boldsymbol{r} - \boldsymbol{\mu}_{\boldsymbol{r}})^H\right] = \mathbb{E}_{\boldsymbol{r}}\left[(\boldsymbol{r} - \boldsymbol{v}(\mathbf{Z}))(\boldsymbol{r} - \boldsymbol{v}(\mathbf{Z}))^H\right] = (\sigma^2\mathbf{I} + \boldsymbol{\Psi}\boldsymbol{\Psi}^H) \\
\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{u}_\mathbf{i}} &= \boldsymbol{\Sigma}_{\boldsymbol{u}_\mathbf{i}\boldsymbol{r}}{}^H = \mathbb{E}\left[(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})(\boldsymbol{r} - \boldsymbol{\mu}_{\boldsymbol{r}})^H\right] \\
&= \mathbb{E}\left[(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z}))(\sum_{k=1}^{D}(\boldsymbol{u}_k - \boldsymbol{v}_k(\mathbf{Z})) + \boldsymbol{\Psi}\boldsymbol{y})^H\right] \\
&= \sum_{k=1}^{D}\mathbb{E}\left[(\boldsymbol{u}_i - \boldsymbol{v}_i(\mathbf{Z}))(\boldsymbol{r}_k - \boldsymbol{v}_k)^H\right]
\end{aligned} \tag{C.3}$$

Since the $\boldsymbol{u}_i$ are considered independant from one another

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{u}_i} = \boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{r}}^H &= \boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{u}_i} + \sum_{k=1,k\neq i}^{D}\mathbb{E}\left[(\boldsymbol{u}_i - \boldsymbol{v}_i)\right]\mathbb{E}\left[(\boldsymbol{u}_k - \boldsymbol{v}_k)\right] \\
&= \boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{u}_i}
\end{aligned} \tag{C.4}$$

Using [35, p36] we can compute the analytical expression for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i}$ and $\hat{\boldsymbol{u}}_i$ as

$$\begin{aligned}
\hat{\boldsymbol{u}}_i &= \boldsymbol{\mu}_{\boldsymbol{u}_i} + \boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{r}}\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{r}}^{-1}(\boldsymbol{r} - \mu_{\boldsymbol{r}}) = \boldsymbol{v}_i(\mathbf{Z}) + \beta_i\sigma^2(\sigma^2\mathbf{I} + \boldsymbol{\Psi}\boldsymbol{\Psi}^H)^{-1}(\boldsymbol{r} - \boldsymbol{v}(Z)) \\
\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i} &= \boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{u}_i} + \boldsymbol{\Sigma}_{\boldsymbol{u}_i\boldsymbol{r}}\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{r}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{u}_i} = \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} - \beta_i^2\sigma^4(\sigma^2\mathbf{I} + \boldsymbol{\Psi}\boldsymbol{\Psi}^H)^{-1}
\end{aligned} \tag{C.5}$$

We can then express

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r}}\left[\boldsymbol{u}_i\right] &= \hat{\boldsymbol{u}}_i \\
\mathbb{E}_{\boldsymbol{u}_i|\boldsymbol{r}}\left[\boldsymbol{u}_i^H\boldsymbol{u}_i\right] &= \mathrm{Tr}\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i}\right) + \hat{\boldsymbol{u}}_i^H\hat{\boldsymbol{u}}_i.
\end{aligned} \tag{C.6}$$

Finally,

$$Q_2(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{D} C_{u_i} - \log \mid \beta_i\sigma^2\mathbf{I}_{4FN_{\mathcal{B}}} \mid -\frac{1}{\beta_i\sigma^2}\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 - \frac{1}{\beta_i\sigma^2}\mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i}). \tag{C.7}$$

Appendix D. Maximization of $Q_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)})$

We compute the derivatives of Q_2 w.r.t σ^2:

$$\begin{aligned}\frac{\partial Q_2}{\sigma^2} &= \sum_{i=1}^{D} -\frac{4FN_{\mathcal{B}}}{\sigma^2} + \frac{1}{\beta_i \sigma^4}\left(\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 + \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i})\right) \\ &= \frac{1}{\sigma^4}\sum_{i=1}^{D} -4FN_{\mathcal{B}}\sigma^2 + \frac{1}{\beta_i}\left(\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 + \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i})\right) \\ &= \frac{1}{\sigma^4}\left(-4FN_{\mathcal{B}}D\sigma^2 + \sum_{i=1}^{D}\frac{1}{\beta_i}\left(\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 + \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i})\right)\right)\end{aligned} \tag{D.1}$$

The update for σ^2 is then derived from D.1,

$$\sigma^2 = \frac{1}{4FN_{\mathcal{B}}D}\sum_{i=1}^{D}\frac{1}{\beta_i}\left(\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2 + \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}_i})\right) \tag{D.2}$$

In order to derive the update for $\mathbf{Z}$, Q_2 is express to highlight the dependence with $\mathbf{Z}$,

$$Q_2(\mathbf{Z}) = A - \sum_{i=1}^{D}\frac{1}{\beta_i\sigma^2}\|\hat{\boldsymbol{u}}_i - \boldsymbol{v}_i(\mathbf{Z})\|_F^2, \tag{D.3}$$

where A is a constant independant of $\mathbf{Z}$. Since $\hat{\boldsymbol{u}}_i$ and $\boldsymbol{v}_i(\mathbf{Z}_i)$ can both be divided into F sub-vectors, they can be written as $\hat{\boldsymbol{u}}_i = [\hat{\boldsymbol{u}}_{i,1}^T, ..., \hat{\boldsymbol{u}}_{i,F}^T]^T$ and $\boldsymbol{v}_i(\mathbf{Z}) = [\boldsymbol{v}_{i,1}(\mathbf{Z})^T, ..., \boldsymbol{v}_{i,F}(\mathbf{Z})^T]^T$. Thus Q_2 can be expressed as follows,

$$Q_2(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) = A - \sum_{i=1}^{D}\sum_{f=1}^{F}\frac{1}{\beta_i\sigma^2}\|\hat{\boldsymbol{u}}_{if} - \boldsymbol{v}_{i,f}(\mathbf{Z})\|_F^2.$$

We consider

$$\phi(\mathbf{Z}_i) = \sum_{f=1}^{F}\frac{1}{\beta_i\sigma^2}\|\hat{\boldsymbol{u}}_{if} - \boldsymbol{v}_{i,f}(\mathbf{Z})\|_F^2$$

where $\boldsymbol{v}_{i,f}$ is the i^{th} source noise free component. Although, $\boldsymbol{v}_{i,f}$ can be partitioned in $N_{\mathcal{B}} = \frac{N(N-1)}{2}$ 4-dimension vectors $\boldsymbol{v}_{i,f} = [\boldsymbol{v}_{i,f,12}^T, \boldsymbol{v}_{i,f,13}^T, ..., \boldsymbol{v}_{i,f,N(N-1)}^T]^T$ and $\boldsymbol{v}_{i,f,pq} = \left(\mathbf{B}_f^* \otimes \mathbf{B}_f\right)(\mathbf{Z}_{i,q}^* \otimes \mathbf{Z}_{i,p})\mathrm{vec}(\mathbf{C}_{i,f})$. We note as well $\hat{\boldsymbol{u}}_{i,f,pq}$ the 4×1 vector such that $\hat{\boldsymbol{u}}_{i,f} = [\hat{\boldsymbol{u}}_{i,f,12}, \hat{\boldsymbol{u}}_{i,f,13}..., \hat{\boldsymbol{u}}_{i,f,(N-1)N}]$. Thus $\phi(\mathbf{Z}_i)$ can now be written

$$\phi(\mathbf{Z}_i) = \sum_{f=1}^{F}\sum_{(p,q)\in[1;N]^2|p<q}\|\hat{\boldsymbol{u}}_{i,f,pq} - \boldsymbol{v}_{i,f,pq}(\mathbf{Z}_i)\|_F^2$$

where $\mathbf{Z}_i$ is a $2KN \times 2$ matrix that can be expressed using $2K \times 2$ matrix

$$\mathbf{Z}_i = \begin{bmatrix}\mathbf{Z}_{i,1}\\ \mathbf{Z}_{i,2}\\ ...\\ \mathbf{Z}_{i,N}\end{bmatrix} \quad \text{where } \mathbf{Z}_{i,p} \in \mathcal{M}_{2K,2} \text{ and } \mathbf{Z}_{i,p} = \begin{bmatrix}\mathbf{Z}_{1,i,p}\\ \mathbf{Z}_{2,i,p}\\ ...\\ \mathbf{Z}_{K,i,p}\end{bmatrix} \quad \text{where } \mathbf{Z}_{k,i,p} \in \mathcal{M}_{2,2}$$

Let us recall that

$$\begin{aligned}\boldsymbol{v}_{i,f,pq} &= \left((\mathbf{B}_f^*\mathbf{Z}_{i,q}^*) \otimes (\mathbf{B}_f\mathbf{Z}_{i,p})\right)\mathrm{vec}(\mathbf{C}_{i,f}) \\ &= \mathrm{vec}(\mathbf{B}_f\mathbf{Z}_{i,p}\mathbf{C}_{i,f}\mathbf{Z}_{i,q}^H\mathbf{B}_f^H)\end{aligned} \tag{D.4}$$

By considering $\hat{\mathbf{U}}_{i,f,pq} = \text{unvec}(\hat{\boldsymbol{u}}_{i,f,pq})$, we can express $\phi(\mathbf{Z}_i)$ as

$$\phi(\mathbf{Z}_i) = \sum_{f=1}^{F} \sum_{(p,q)\in[1;N]^2|p<q} \|\hat{\mathbf{U}}_{i,f,pq} - \mathbf{B}_f \mathbf{Z}_{i,p} \mathbf{C}_{i,f} \mathbf{Z}_{i,q}^H \mathbf{B}_f^H\|_F^2. \tag{D.5}$$

$\phi(\mathbf{Z}_i)$ is minimized w.r.t to $\mathbf{Z}_i$ with a block coordinate descent [34] by minimizing recursevely $\phi(\mathbf{Z}_i)$ w.r.t $\mathbf{Z}_{i,p}$ for a given $\mathbf{Z}_{i,q}$ with $p \neq q$. Let $p \in [1:N]$, we note

$$f_{i,f,pq}(\mathbf{Z}_{i,p}) = \|\hat{\mathbf{U}}_{i,f,pq} - \mathbf{B}_f \mathbf{Z}_{i,p} \mathbf{C}_{i,f} \mathbf{Z}_{i,q}^H \mathbf{B}_f^H\|_F^2$$

$$g_{i,f,pq}(\mathbf{Z}_{i,p}) = \|\hat{\mathbf{U}}_{i,f,qp} - \mathbf{B}_f \mathbf{Z}_{i,q} \mathbf{C}_{i,f} \mathbf{Z}_{i,p}^H \mathbf{B}_f^H\|_F^2.$$

The cost function $\phi(\mathbf{Z}_{i,p})$ becomes,

$$\phi(\mathbf{Z}_{i,p}) = A + \sum_{f=1}^{F} \sum_{q=p+1}^{N} f_{i,f,pq}(\mathbf{Z}_{i,p}) + \sum_{f=1}^{F} \sum_{q=1}^{p-1} g_{i,f,pq}(\mathbf{Z}_{i,p}). \tag{D.6}$$

The partial derivatives of f and g w.r.t $\mathbf{Z}_{i,p}$ are computed using [33]

$$\frac{\partial f_{i,f,pq}}{\mathbf{Z}_{i,p}} = \mathbf{B}_f^T (\hat{\mathbf{U}}_{i,f,pq}^* - \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{F}_{i,f,q}^H) \mathbf{F}_{i,f,q}$$

$$\frac{\partial g_{i,f,pq}}{\mathbf{Z}_{i,p}} = \mathbf{B}_f^T (\hat{\mathbf{U}}_{i,f,qp}^T - \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{G}_{i,f,q}^H) \mathbf{G}_{i,f,q}.$$

Thus, the partial derivative of $\phi(\mathbf{Z}_{i,p})$ w.r.t $\mathbf{Z}_{i,p}$ can be expressed,

$$\begin{aligned}
\frac{\partial \phi}{\mathbf{Z}_{i,p}} &= \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{B}_f^T (\hat{\mathbf{U}}_{i,f,pq}^* - \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{F}_{i,f,q}^H) \mathbf{F}_{i,f,q} \\
&+ \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{B}_f^T (\hat{\mathbf{U}}_{i,f,qp}^T - \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{G}_{i,f,q}^H) \mathbf{G}_{i,f,q} \\
&= \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{B}_f^T \hat{\mathbf{U}}_{i,f,pq}^* \mathbf{F}_{i,f,q} + \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{B}_f^T \hat{\mathbf{U}}_{i,f,qp}^T \mathbf{G}_{i,f,q} \\
&- \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{B}_f^T \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{F}_{i,f,q}^H \mathbf{F}_{i,f,q} \\
&- \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{B}_f^T \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{G}_{i,f,q}^H \mathbf{G}_{i,f,q}.
\end{aligned} \tag{D.7}$$

Let us note

$$\mathbf{T}_{i,p} = \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{B}_f^T \hat{\mathbf{U}}_{i,f,pq}^* \mathbf{F}_{i,f,q} + \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{B}_f^T \hat{\mathbf{U}}_{i,f,qp}^T \mathbf{G}_{i,f,q}.$$

The partial derivative of $\phi(\mathbf{Z}_{i,p})$ w.r.t $\mathbf{Z}_{i,p}$ becomes

$$\frac{\partial \phi}{\mathbf{Z}_{i,p}} = \mathbf{T}_{i,p} - \sum_{f=1}^{F} \sum_{q=p+1}^{N} \mathbf{B}_f^T \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{F}_{i,f,q}^H \mathbf{F}_{i,f,q} - \sum_{f=1}^{F} \sum_{q=1}^{p-1} \mathbf{B}_f^T \mathbf{B}_f^* \mathbf{Z}_{i,p}^* \mathbf{G}_{i,f,q}^H \mathbf{G}_{i,f,q}. \tag{D.8}$$

Setting the partial derivative of $\phi(\mathbf{Z}_{i,p})$ w.r.t $\mathbf{Z}_{i,p}$ to zero is equivalent to setting the partial derivative of $\text{vec}(\phi(\mathbf{Z}_{i,p}))$ w.r.t $\text{vec}(\mathbf{Z}_{i,p})$ to zero,

$$\begin{aligned}\text{vec}(\frac{\partial \phi}{\mathbf{Z}_{i,p}}) &= \text{vec}(\mathbf{T}_{i,p}) - \sum_{f=1}^{F}\sum_{q=p+1}^{N}\text{vec}(\mathbf{B}_f^T\mathbf{B}_f^*\mathbf{Z}_{i,p}^*\mathbf{F}_{i,f,q}^H\mathbf{F}_{i,f,q}) \\ &\quad - \sum_{f=1}^{F}\sum_{q=1}^{p-1}\text{vec}(\mathbf{B}_f^T\mathbf{B}_f^*\mathbf{Z}_{i,p}^*\mathbf{G}_{i,f,q}^H\mathbf{G}_{i,f,q}) \\ &= \text{vec}(\mathbf{T}_i) - \sum_{f=1}^{F}\sum_{q=p+1}^{N}\Big(\mathbf{F}_{i,f,q}^T\mathbf{F}_{i,f,q}^* \otimes \mathbf{B}_f^T\mathbf{B}_f^*\Big)\text{vec}(\mathbf{Z}_{i,p}^*) \\ &\quad - \sum_{f=1}^{F}\sum_{q=1}^{p-1}\Big(\mathbf{G}_{i,f,q}^T\mathbf{G}_{i,f,q}^* \otimes \mathbf{B}_f^T\mathbf{B}_f^*\Big)\text{vec}(\mathbf{Z}_{i,p}^*).\end{aligned} \tag{D.9}$$

Finally, we obtain a solution for $\mathbf{Z}_{i,p}$ such that $\text{vec}(\frac{\partial \phi}{\mathbf{Z}_{i,p}}) = 0$,

$$\text{vec}(\mathbf{Z}_{i,p}^*) = (\mathbf{M}_{i,p} + \tilde{\mathbf{M}}_{i,p})^{-1}\text{vec}(\mathbf{T}_{i,p}),$$

$$\text{with } \mathbf{M}_{i,p} = \sum_{f=1}^{F}\sum_{q=p+1}^{N}\Big(\mathbf{F}_{i,f,q}^T\mathbf{F}_{i,f,q}^* \otimes \mathbf{B}_f^T\mathbf{B}_f^*\Big)$$

$$\tilde{\mathbf{M}}_{i,p} = \sum_{f=1}^{F}\sum_{q=1}^{p-1}\Big(\mathbf{G}_{i,f,q}^T\mathbf{G}_{i,f,q}^* \otimes \mathbf{B}_f^T\mathbf{B}_f^*\Big)$$

$$\mathbf{T}_{i,p} = \sum_{f=1}^{F}\sum_{q=p+1}^{N}\mathbf{B}_f^T\hat{\mathbf{U}}_{i,f,pq}^*\mathbf{F}_{i,f,q} + \sum_{f=1}^{F}\sum_{q=1}^{p-1}\mathbf{B}_f^T\hat{\mathbf{U}}_{i,f,qp}^T\mathbf{G}_{i,f,q}$$

$$\mathbf{F}_{i,f,q} = \mathbf{B}_f^*\mathbf{Z}_{i,q}^*\mathbf{C}_{i,f}^T$$

$$\mathbf{G}_{i,f,q} = \mathbf{B}_f^*\mathbf{Z}_{i,q}^*\mathbf{C}_{i,f}^*.$$

References

[1] K. Jansky, "Electrical Disturbances Apparently Of Extraterrestrial Origin," *Proceedings of the IEEE*, vol. 86, pp. 1510–1515, 1933.

[2] A. Alberdi, J. Gómez Fernández, E. H. T. Collaboration, *et al.*, "First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole," 2019.

[3] A. Thompson, J. Moran, and G. Swenson, *Interferometry and synthesis in radio astronomy*. Springer Nature, 2017.

[4] O. Smirnov, "Revisiting the radio interferometer measurement equation-I. A full-sky Jones formalism," *Astronomy & Astrophysics*, vol. 527, p. A106, 2011.

[5] J. Hamaker, J. Bregman, and R. Sault, "Understanding radio polarimetry. I. Mathematical foundations," *Astronomy and Astrophysics Supplement Series*, vol. 117, no. 1, pp. 137–147, 1996.

[6] A. Boonstra and A. Van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 25–38, 2003.

[7] S. Wijnholds and A. Van Der Veen, "Multisource self-calibration for sensor arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3512–3522, 2009.

[8] S. Salvini and S. Wijnholds, "Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications," *Astronomy & Astrophysics*, vol. 571, p. A97, 2014.

[9] C. Tasse, "Nonlinear Kalman filters for calibration in radio interferometry," *Astronomy & Astrophysics*, vol. 566, p. A127, 2014.

[10] S. Kazemi and S. Yatawatta, "Robust radio interferometric calibration using the t-distribution," *Monthly Notices of the Royal Astronomical Society*, vol. 435, pp. 597–605, Oct. 2013.

[11] V. Ollier, M. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Robust calibration of radio interferometers in non-Gaussian environment," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5649–5660, 2017.
[12] U. Sob, H. Bester, O. Smirnov, J. Kenyon, and T. Grobler, "Radio interferometric calibration using a complex Student's t-distribution and Wirtinger derivatives," *Monthly Notices of the Royal Astronomical Society*, vol. 491, no. 1, pp. 1026–1042, 2020.
[13] Y. Mhiri, M. El Korso, L. Bacharach, A. Breloy, and P. Larzabal, ""Expectation-Maximization Based Direction of Arrival Estimation Under a Mixture of Noise"," in *2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 1890–1893, 2021.
[14] S. Yatawatta, "Stochastic calibration of radio interferometers," *Monthly Notices of the Royal Astronomical Society*, vol. 493, no. 4, pp. 6071–6078, 2020.
[15] W. Baan, "Implementing RFI mitigation in radio science," *Journal of Astronomical Instrumentation*, vol. 8, no. 01, p. 1940010, 2019.
[16] A. Offringa *et al.*, "A LOFAR RFI detection pipeline and its first results, in proceedings of RFI mitigation workshop," *PoS (RFI2010)*, vol. 36, 2010.
[17] A. Leshem and A. van der Veen, "Introduction to interference mitigation techniques in radio astronomy," in *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*, p. 201, 2000.
[18] A. Boonstra, S. Wijnholds, S. van der Tol, and B. Jeffs, "Calibration, Sensitivity and RFI Mitigation Requirements for LOFAR," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5, (Philadelphia, Pennsylvania, USA), pp. 869–872, IEEE, 2005.
[19] J. Lemmon, "Wideband model of man-made HF noise and interference," *Radio Science*, vol. 32, no. 2, pp. 525–539, 1997.
[20] M. Lavielle, *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014.
[21] J. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on signal processing*, vol. 42, no. 10, pp. 2664–2677, 1994.
[22] S. Wijnholds and A. Boonstra, "A multisource calibration method for phased array radio telescopes," in *Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006.*, pp. 200–204, IEEE, 2006.
[23] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Lampropoulos, A. De Bruyn, L. Koopmans, and J. Noordam, "Radio interferometric calibration using the SAGE algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 2, pp. 1656–1666, 2011.
[24] S. Yatawatta, "Distributed radio interferometric calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 449, no. 4, pp. 4506–4514, 2015.
[25] S. Yatawatta, "Polarization-based online interference mitigation in radio interferometry," in *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 1961–1965, IEEE, 2021.
[26] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez, "Asymptotic properties of robust complex covariance matrix estimates," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3348–3356, 2013.
[27] P. Fridman and W. Baan, "RFI mitigation methods in radio astronomy," *Astronomy & Astrophysics*, vol. 378, no. 1, pp. 327–344, 2001.
[28] M. Brossard, M. El Korso, M. Pesavento, R. Boyer, P. Larzabal, and S. Wijnholds, "Parallel multi-wavelength calibration algorithm for radio astronomical arrays," *Signal Processing*, vol. 145, pp. 258–271, 2018.
[29] R. Kozick and B. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3520–3535, 2000.
[30] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
[31] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of statistics*, pp. 95–103, 1983.
[32] Z. Wang, Q. Gu, Y. Ning, and H. Liu, "High Dimensional Expectation-Maximization Algorithm: Statistical Optimization and Asymptotic Normality," *arXiv:1412.8729 [stat]*, Jan. 2015. arXiv: 1412.8729.
[33] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2740–2746, 2007.
[34] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *The annals of applied statistics*, vol. 1, no. 2, pp. 302–332, 2007.
[35] T. Anderson, *An introduction to multivariate statistical analysis*. Wiley New York, 2003.