URL: http://dx.doi.org/10.1016/j.sigpro.2009.09.008

Robust approaches to remote calibration of a transmitting array ☆

Olivier Besson [a,*], Stéphanie Bidon [a], Cécile Larue de Tournemine [b]

[a] University of Toulouse, ISAE/TéSA, Department of Electronics, Optronics and Signal, Toulouse, France
[b] Thales Alenia Space, Toulouse, France

Keywords:
Array processing
Calibration
Robust estimation

A B S T R A C T

We consider the problem of estimating the gains and phases of the RF channels of a M-element transmitting array, based on a calibration procedure where M orthogonal signals are sent through M orthogonal beams and received on a single antenna. The received data vector obeys a linear model of the type $\boldsymbol{y} = A\boldsymbol{F}\boldsymbol{g} + \boldsymbol{n}$ where A is an unknown complex scalar accounting for propagation loss and $\boldsymbol{g}$ is the vector of unknown complex gains. In order to improve the performance of the least-squares (LS) estimator at low signal to noise ratio (SNR), we propose to exploit knowledge of the nominal value of $\boldsymbol{g}$, viz $\overline{\boldsymbol{g}}$. Towards this end, two approaches are presented. First, a Bayesian approach is advocated where A and $\boldsymbol{g}$ are considered as random variables, with a non-informative prior distribution for A and a Gaussian prior distribution for $\boldsymbol{g}$. The posterior distributions of the unknown random variables are derived and a Gibbs sampling strategy is presented that enables one to generate samples distributed according to these posterior distributions, leading to the minimum mean-square error (MMSE) estimator. A second approach consists in solving a constrained least-squares problem in which $\boldsymbol{h} = A\boldsymbol{g}$ is constrained to be close to a scaled version of $\overline{\boldsymbol{g}}$. This second approach yields a closed-form solution, which amounts to a linear combination of $\overline{\boldsymbol{g}}$ and the LS estimator. Numerical simulations show that the two new estimators significantly outperform the conventional LS estimator, especially at low SNR.

1. Introduction

In many array processing applications, it is desired to design and maintain a specified array beampattern, obtained by properly weighting the signals received at or transmitted from the various antennas, in order to achieve spatially selective filtering. This is especially so in multibeam satellite communications systems [1] where the area to be covered is divided into several ''spots'' wherein a minimal gain should be guaranteed for all users inside the spot while ensuring a sufficient isolation between spots, so as to minimize inter-beam interference. However, due for instance to severe temperature conditions, it is hardly feasible to maintain the same gains and phases for all RF channels of the array and, unavoidably, disparities between the channels appear leading to degradation of the array beampattern (deformation of the mainlobe and increase of sidelobe levels). Therefore, it is necessary to regularly re-calibrate the array in order for the nominal mainlobe–sidelobes specifications to be fulfilled constantly over time. This task usually requires as a pre-requisite estimation of all RF complex gains followed by proper modification of the weights to be applied to each channel.

While the literature about calibration of a receiving array is abundant, see e.g. [2–10], the case of a transmitting array, which is the one of primary interest in the present paper, has received much less attention.

☆ This work was supported by Thales Alenia Space.
* Corresponding author.
E-mail addresses: olivier.besson@isae.fr (O. Besson), sbidon@isae.fr (S. Bidon), Cecile.Larue-de-Tournemine@thalesaleniaspace.com (C. Larue de Tournemine).

In [11], calibration of a spaceborne phased-array is studied based on transmission of time-multiplexed signals over orthogonal beams. The PN-Gating method is presented and studied in [12,13] to calibrate internally each transmit/receive module of an active phased array antenna. The technique relies on scrambling the signal of each operating module with a pseudo-noise sequence. Refs. [14,15] present a very general framework for calibration of a transmitting array. The principle is to transmit M signals—with M the number of array elements—orthogonal in the time domain over M orthogonal beams, so as to benefit from both time and space diversities. The signal is then received on a single antenna and the complex gains of each antenna are estimated using a linear least-squares approach, since the received data obeys a linear model. The method is simple and performs well, at least at high signal to noise ratio (SNR). However, it does not perform well at low SNR. In order to remedy this problem, we propose to exploit knowledge of the nominal value of $\boldsymbol{g}$ in the calibrated case, viz $\overline{\boldsymbol{g}}$. More precisely, we wish to take advantage of the fact that $\boldsymbol{h}$ is "close" to $A\overline{\boldsymbol{g}}$ but, at the same time, one should take into account the fact that A *is unknown*. In order to use the former a priori knowledge while alleviating the latter problem, two approaches are proposed. First, we consider $\boldsymbol{g}$ as a random variable with a Gaussian distribution centered around $\overline{\boldsymbol{g}}$. Since no information is assumed to be available for A, the latter is assumed to be random with a non-informative prior distribution. Within this Bayesian framework, the minimum mean-square error (MMSE) estimates of A and $\boldsymbol{g}$ are obtained using a Markov-chain Monte-Carlo (MCMC) approach. An alternative frequentist approach is also considered where $\boldsymbol{h}$ is obtained from a constrained least-squares (CLS) problem.

The paper is organized as follows. In Section 2, the model of [14,15] is briefly reviewed. In Section 3, we derive the minimum mean-square error estimates of the parameters of interest. In Section 4, we present the CLS problem and derive a closed-form solution for it. Numerical simulations study the performance as well as the robustness of the estimators in Section 5.

2. Data model

The principle of the method proposed in [14,15] is to transmit M orthogonal waveforms $c_k(t)$, $k = 1, \ldots, M$ over M orthogonal beams, each of them corresponding to a weight vector $\boldsymbol{w}_k$, see Fig. 1 for a pictorial representation of this scheme. The complex envelope of the signal $\boldsymbol{s}(t) = [s_1(t)\ s_2(t)\ \cdots\ s_M(t)]^T$ transmitted at time t by the array is given by

$$\boldsymbol{s}(t) = \sum_{k=1}^{M} \boldsymbol{w}_k^* c_k(t) = \boldsymbol{W}^* \boldsymbol{c}(t) \tag{1}$$

where $\boldsymbol{c}(t) = [c_1(t)\ c_2(t)\ \cdots\ c_M(t)]^T$ and $\boldsymbol{W} = [\boldsymbol{w}_1\ \boldsymbol{w}_2\ \cdots\ \boldsymbol{w}_M]$. Without loss of generality the weight vectors $\boldsymbol{w}_k$ are assumed to have unit norm, i.e. $\|\boldsymbol{w}_k\| = 1$. Let $\boldsymbol{g} = [g_1\ g_2\ \cdots\ g_M]^T$ denote the vector of the unknown complex gains of the array, and $\boldsymbol{D}_{\boldsymbol{g}} = \mathrm{diag}(\boldsymbol{g})$. Let the receiver consist of a single antenna, with known location,

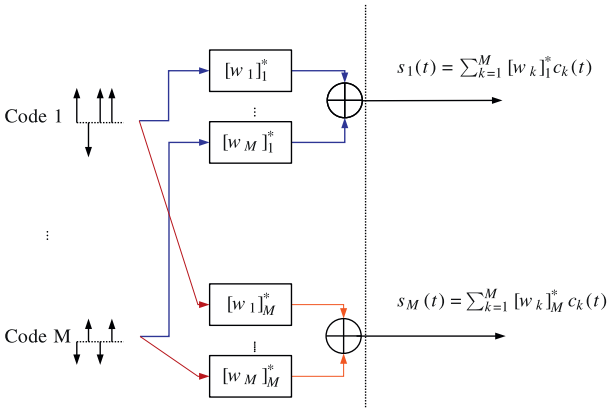

Fig. 1. Principle of the calibration procedure.

and let $\boldsymbol{a}_r$ denote the (known) array steering vector for the receiver. We assume herein that its elements have a constant modulus (which is set to 1). The received signal can thus be written as

$$y(t) = A\boldsymbol{a}_r^T \boldsymbol{D}_{\boldsymbol{g}} \boldsymbol{W}^* \boldsymbol{c}(t) + n(t) = A\boldsymbol{c}^T(t) \boldsymbol{W}^H \boldsymbol{D}_r \boldsymbol{g} + n(t) \tag{2}$$

where $\boldsymbol{D}_r = \mathrm{diag}(\boldsymbol{a}_r)$, $n(t)$ stands for the additive noise, and A is an unknown complex scalar which accounts for the propagation loss between the array and the receiver, and the receiver antenna gain. We assume that N samples are collected at the receiver and stacked in a vector $\boldsymbol{y} = [y(1)\ y(2)\ \cdots\ y(N)]^T$. Let

$$\boldsymbol{C} = \begin{bmatrix} c_1(1) & c_2(1) & \cdots & c_M(1) \\ c_1(2) & c_2(2) & \cdots & c_M(2) \\ \vdots & \vdots & & \vdots \\ c_1(N) & c_2(N) & \cdots & c_M(N) \end{bmatrix} = [\boldsymbol{c}_1\ \ \boldsymbol{c}_2\ \ \cdots\ \ \boldsymbol{c}_M] \tag{3}$$

be the $N \times M$ signal matrix, whose m th column corresponds to the signals sent on antenna number m. Then, using (2) it follows that

$$\boldsymbol{y} = A\boldsymbol{C}\boldsymbol{W}^H \boldsymbol{D}_r \boldsymbol{g} + \boldsymbol{n} = A\boldsymbol{F}\boldsymbol{g} + \boldsymbol{n} \tag{4}$$

where $\boldsymbol{n} = [n(1)\ \ldots\ n(N)]^T$ stands for the noise vector and $\boldsymbol{F} \triangleq \boldsymbol{C}\boldsymbol{W}^H \boldsymbol{D}_r$. Through the paper, we assume that $\boldsymbol{n}$ is a zero-mean complex Gaussian vector, whose covariance matrix is $\sigma^2 \boldsymbol{I}$. We also assume that the waveforms are orthogonal, i.e.

$$\boldsymbol{c}_k^H \boldsymbol{c}_\ell = \sum_{t=1}^{N} c_k^*(t) c_\ell(t) = NP_c \delta_{k,\ell}$$

where P_c stands for their (common) power, and that the beams are orthogonal, i.e. $\boldsymbol{w}_k^H \boldsymbol{w}_\ell = \delta_{k,\ell}$. In such a case, $\boldsymbol{F}^H \boldsymbol{F} = NP_c \boldsymbol{D}_r^H \boldsymbol{D}_r = NP_c \boldsymbol{I} = \gamma \boldsymbol{I}$.

The model in (4) forms the basis of our study. A first observation regarding this model is that, if A and $\boldsymbol{g}$ are considered as unknown deterministic parameters, they cannot be identified unambiguously since there exists a scaling ambiguity between them, namely $\boldsymbol{h} = A\boldsymbol{g} = (\eta A)(\eta^{-1}\boldsymbol{g})$. For calibration purposes, this may not be a problem as, usually, the beamformer weights $\boldsymbol{w}$ are first properly scaled to compensate for gain and phase errors,

and then normalized to meet an external constraint, for example on power consumption by transmitters, desired radiated power levels, or dynamic range. In [14,15], it is proposed to estimate $\boldsymbol{h}$ in a least-squares sense as

$$\hat{\boldsymbol{h}}_{\mathrm{ls}} = \underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 = (\boldsymbol{F}^H\boldsymbol{F})^{-1}\boldsymbol{F}^H\boldsymbol{y} = \gamma^{-1}\boldsymbol{F}^H\boldsymbol{y}. \tag{5}$$

This estimator is simple and performs rather well, at least at high signal to noise ratio. Moreover, it is known that its variance is minimal when $\boldsymbol{F}^H\boldsymbol{F}$ is proportional to the identity matrix [16], which validates the use of time and space orthogonal signals. However, its performance degrades at low SNR and our goal is to improve over (5) in this regime. Towards this end, a natural means is to make use of the a priori knowledge available. In our case, the nominal value $\overline{\boldsymbol{g}}$ of the complex gain vector $\boldsymbol{g}$ is usually known, and this information should be taken in advantage in order to improve estimation performance. Although $\boldsymbol{g}$ is known to be close to $\overline{\boldsymbol{g}}$, we only know that $\boldsymbol{h}$ is close to $A\overline{\boldsymbol{g}}$ but A is otherwise unknown. In other words, one does not have access to a known vector $\overline{\boldsymbol{h}}$ that would be close to $\boldsymbol{h}$ and could be used to estimate directly $\boldsymbol{h}$. Therefore, our problem consists of estimating $\boldsymbol{g}$ given knowledge of $\overline{\boldsymbol{g}}$, or estimating $\boldsymbol{h}$ knowing that $\boldsymbol{h}$ should be close to a scaled version of $\overline{\boldsymbol{g}}$. In order to solve this problem, we first investigate a Bayesian approach in which $\boldsymbol{g}$ is random, with a Gaussian distribution around $\overline{\boldsymbol{g}}$. Then, we consider a constrained leats-squares approach where $\boldsymbol{h}$ is constrained to be close, up to a scaling factor, to $\overline{\boldsymbol{g}}$.

3. Bayesian approach

As indicated above, our first approach consists in assuming that A and $\boldsymbol{g}$ are random variables with some prior distributions. More precisely, we assume here that $\boldsymbol{g}$ is a complex Gaussian vector, with mean $\overline{\boldsymbol{g}}$ and covariance matrix $\sigma_g^2\boldsymbol{I}$, i.e. $\boldsymbol{g}\sim\mathcal{CN}(\overline{\boldsymbol{g}}, \sigma_g^2\boldsymbol{I})$, so that the prior distribution of $\boldsymbol{g}$ is

$$f_g(\boldsymbol{g}) = \pi^{-M}\sigma_g^{-2M}e^{-\sigma_g^{-2}\|\boldsymbol{g}-\overline{\boldsymbol{g}}\|^2}. \tag{6}$$

Regarding A we wish not to make any assumption about it (i.e. not use any knowledge) and therefore we consider a non-informative prior for A, namely Jeffreys prior [17]. This approach is robust as no statistical model for A is invoked. In Appendix A, we show that Jeffreys prior for A is given by

$$f_a(A) \propto \frac{(1+\gamma\sigma^{-2}\sigma_g^2[1+2M\sigma_g^2\|\overline{\boldsymbol{g}}\|^{-2}]|A|^2)^{1/2}}{(1+\gamma\sigma^{-2}\sigma_g^2|A|^2)^{3/2}}. \tag{7}$$

As expected, this prior distribution only depends on $\overline{\boldsymbol{g}}$ and σ_g^2. In order to estimate $\boldsymbol{g}$, a natural approach is to look for its MMSE estimate which, as a pre-requisite, necessitates deriving the posterior distribution of $\boldsymbol{g}$, conditionally to $\boldsymbol{y}$. Since the joint posterior distribution of A and $\boldsymbol{g}$ is given by

$$f(A,\boldsymbol{g}|\boldsymbol{y}) \propto f(\boldsymbol{y}|A,\boldsymbol{g})f_g(\boldsymbol{g})f_a(A) \propto e^{-\sigma^{-2}\|\boldsymbol{y}-A\boldsymbol{F}\boldsymbol{g}\|^2}e^{-\sigma_g^{-2}\|\boldsymbol{g}-\overline{\boldsymbol{g}}\|^2}f_a(A) \tag{8}$$

it follows that the posterior distribution of $\boldsymbol{g}$, conditionally to $\boldsymbol{y}$, is

$$\begin{aligned} f(\boldsymbol{g}|\boldsymbol{y}) &= \int f(A,\boldsymbol{g}|\boldsymbol{y})\,dA \\ &\propto e^{-\sigma_g^{-2}\|\boldsymbol{g}-\overline{\boldsymbol{g}}\|^2} \times \int e^{-\sigma^{-2}\|\boldsymbol{y}-A\boldsymbol{F}\boldsymbol{g}\|^2}f_a(A)\,dA \\ &\propto e^{-\sigma_g^{-2}\|\boldsymbol{g}-\overline{\boldsymbol{g}}\|^2}e^{C_A^{-1}|\mu_A|^2} \times \int e^{-C_A^{-1}|A-\mu_A|^2}f_a(A)\,dA \end{aligned} \tag{9}$$

where

$$C_A = \sigma_g^2(\boldsymbol{g}^H\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{g})^{-1} \tag{10a}$$

$$\mu_A = \sigma^{-2}C_A\boldsymbol{g}^H\boldsymbol{F}^H\boldsymbol{y}. \tag{10b}$$

The MMSE estimator of $\boldsymbol{g}$ corresponds to the mean of $f(\boldsymbol{g}|\boldsymbol{y})$. Unfortunately, it seems intractable to obtain a closed-form expression for the integral in (9) and therefore deriving $f(\boldsymbol{g}|\boldsymbol{y})$ appears to be impossible. As a consequence, obtaining $\int \boldsymbol{g}f(\boldsymbol{g}|\boldsymbol{y})\,d\boldsymbol{g}$ or even generating samples distributed according to $f(\boldsymbol{g}|\boldsymbol{y})$ is not feasible. In order to solve this problem, a convenient alternative is to resort to a Gibbs-sampling strategy that only requires the conditional posterior distributions $f(\boldsymbol{g}|A,\boldsymbol{y})$ and $f(A|\boldsymbol{g},\boldsymbol{y})$. Indeed, the principle of the Gibbs sampler is to generate iteratively samples drawn from $f(A|\boldsymbol{g},\boldsymbol{y})$ and samples drawn from $f(\boldsymbol{g}|A,\boldsymbol{y})$, which is easier to do, as shown now. From the assumptions made, we have

$$f(\boldsymbol{g}|A,\boldsymbol{y}) \propto e^{-\sigma^{-2}\|\boldsymbol{y}-A\boldsymbol{F}\boldsymbol{g}\|^2-\sigma_g^{-2}\|\boldsymbol{g}-\overline{\boldsymbol{g}}\|^2} \propto e^{-(\boldsymbol{g}-\boldsymbol{\mu}_g)^H\boldsymbol{C}_g^{-1}(\boldsymbol{g}-\boldsymbol{\mu}_g)} \tag{11}$$

with

$$\boldsymbol{C}_g = (\sigma^{-2}|A|^2\boldsymbol{F}^H\boldsymbol{F} + \sigma_g^{-2}\boldsymbol{I})^{-1} \tag{12a}$$

$$\boldsymbol{\mu}_g = \boldsymbol{C}_g(\sigma^{-2}A^*\boldsymbol{F}^H\boldsymbol{y} + \sigma_g^{-2}\overline{\boldsymbol{g}}) \tag{12b}$$

and therefore

$$\boldsymbol{g}|A,\boldsymbol{y}\sim\mathcal{CN}(\boldsymbol{\mu}_g, \boldsymbol{C}_g). \tag{13}$$

Consequently, $\boldsymbol{g}$, conditionally to A and $\boldsymbol{y}$, is Gaussian distributed and therefore, it is relatively simple to generate samples from this distribution. Let us turn now to $f(A|\boldsymbol{g},\boldsymbol{y})$. From (8), we have that

$$\begin{aligned} f(A|\boldsymbol{g},\boldsymbol{y}) &\propto e^{-\sigma^{-2}\|\boldsymbol{y}-A\boldsymbol{F}\boldsymbol{g}\|^2}f_a(A) \\ &\propto e^{-C_A^{-1}|A-\mu_A|^2} \times \frac{(1+\gamma\sigma^{-2}\sigma_g^2[1+2M\sigma_g^2\|\overline{\boldsymbol{g}}\|^{-2}]|A|^2)^{1/2}}{(1+\gamma\sigma^{-2}\sigma_g^2|A|^2)^{3/2}}. \end{aligned} \tag{14}$$

It turns out that $f(A|\boldsymbol{g},\boldsymbol{y})$ does not belong to a familiar class of distributions and hence generating samples drawn according to $f(A|\boldsymbol{g},\boldsymbol{y})$ appears problematic. In order to get round this difficulty, we propose to use an hybrid Metropolis-within-Gibbs sampling strategy [18,19]. Briefly stated, the principle is the following. Since it is not possible to draw samples from $f(A|\boldsymbol{g},\boldsymbol{y})$, the idea is to draw samples from a proposal distribution and to accept or reject this candidate with a given probability. The proposal distribution should, as much as possible, be close to the target distribution. Looking at the different terms in (14), we observed that the first term is the most influent: the second term does not vary much and hence $f(A|\boldsymbol{g},\boldsymbol{y})$ can be fairly well approximated by its first term.

Table 1
Metropolis-within-Gibbs sampler.

Input: $\boldsymbol{y}$
1: generate initial value $A^{(0)}$
2: **for** $n = 1, \ldots, N_{bi} + N_r$ **do**
3: generate $\boldsymbol{g}^{(n)}$ according to $f(\boldsymbol{g}|A^{(n-1)}, \boldsymbol{y})$ in (11)
4: generate A_c according to $f_p(A|\boldsymbol{g}^{(n)}, \boldsymbol{y})$ in (15)
5: compute $r = \frac{f(A_c|\boldsymbol{g}^{(n)},\boldsymbol{y}) \times f_p(A^{(n-1)}|\boldsymbol{g}^{(n)},\boldsymbol{y})}{f(A^{(n-1)}|\boldsymbol{g}^{(n)},\boldsymbol{y}) \times f_p(A_c|\boldsymbol{g}^{(n)},\boldsymbol{y})}$
6: Set $A^{(n)} = \begin{cases} A_c & \text{with probability } \min(r, 1) \\ A^{(n-1)} & \text{otherwise} \end{cases}$
7: **end for**
Output: sequence of random variables $A^{(n)}$ and $\boldsymbol{g}^{(n)}$

Therefore, we advocate as a proposal distribution a Gaussian distribution with mean μ_A and variance C_A, i.e.

$$f_p(A|\boldsymbol{g}, \boldsymbol{y}) \propto e^{-\sigma^{-2}\boldsymbol{g}^H\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{g}|A - \boldsymbol{g}^H\boldsymbol{F}^H\boldsymbol{y}/(\boldsymbol{g}^H\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{g})|^2}. \tag{15}$$

The additional interest of this choice is that it is easy to generate samples drawn from (15). The estimation procedure, referred to as a Metropolis-within-Gibbs sampler, is detailed in Table 1. Note that the numerator [resp. denominator] of r in line 5 of Table 1 is the last term of Eq. (14) evaluated at A_c [resp. $A^{(n-1)}$].

In Table 1, N_{bi} stands for the number of burn-in iterations and N_r is the number of samples which are effectively averaged. Note that there exist statistically sound criteria, such as the potential scale reduction factor [19,20], to select the values of N_{bi} and N_r that ensure convergence of the Gibbs sampler, see also [21,22] for an application to array processing. The above Metropolis-within-Gibbs sampler is known to generate random variables which are asymptotically distributed according to the posterior distributions $f(\boldsymbol{g}|\boldsymbol{y})$ and $f(A|\boldsymbol{y})$, and therefore a natural way to approximate the MMSE estimator is to average the N_r last values generated by the sampler, i.e.

$$\hat{A}_{\text{mmse}} = N_r^{-1} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} A^{(n)} \tag{16a}$$

$$\hat{\boldsymbol{g}}_{\text{mmse}} = N_r^{-1} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} \boldsymbol{g}^{(n)}. \tag{16b}$$

Once $\hat{A}_{\text{mmse}}$ and $\hat{\boldsymbol{g}}_{\text{mmse}}$ are available, an estimate of $\boldsymbol{h}$ can be obtained as $\hat{A}_{\text{mmse}} \times \hat{\boldsymbol{g}}_{\text{mmse}}$.

4. Constrained least-squares estimation

The second approach proposed in this paper considers the unknown variables as deterministic and amounts to constraining the solution $\boldsymbol{h}$ to be close to its nominal value $A\overline{\boldsymbol{g}}$. More precisely, we propose to estimate $\boldsymbol{h}$ (and A) by solving the following minimization problem:

$$\min_{A,\boldsymbol{h}} \|\boldsymbol{h} - A\overline{\boldsymbol{g}}\|^2 \text{ subject to } \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 \leq \varepsilon^2. \tag{17}$$

The criterion to be minimized guarantees that $\boldsymbol{h}$ will not be far from a scaled version of $\overline{\boldsymbol{g}}$ while the constraint imposes that the residual LS error be lower than a desired value. In (17), ε^2 is a user-defined parameter that yields a solution which lies in between a scaled version of $\overline{\boldsymbol{g}}$ and the LS solution $\hat{\boldsymbol{h}}_{\text{ls}}$. In fact, in order for (17) to have a solution, it is necessary that

$$\varepsilon^2 \geq \min_{\boldsymbol{h}} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 = \boldsymbol{y}^H\boldsymbol{P}_{\boldsymbol{F}}^{\perp}\boldsymbol{y} \triangleq \varepsilon_{\min}^2 \tag{18}$$

where $\boldsymbol{P}_{\boldsymbol{F}}^{\perp}$ stands for the orthogonal projector onto the subspace orthogonal to the range space $\mathcal{R}\{\boldsymbol{F}\}$ of $\boldsymbol{F}$. Accordingly, in order to avoid a solution aligned with $\overline{\boldsymbol{g}}$, one must enforce that

$$\varepsilon^2 < \min_{A} \|\boldsymbol{y} - A\boldsymbol{F}\overline{\boldsymbol{g}}\|^2 = \boldsymbol{y}^H\boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{y} \triangleq \varepsilon_{\max}^2 \tag{19}$$

with $\boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{g}}}^{\perp}$ the orthogonal projector onto the subspace orthogonal to $\mathcal{R}\{\boldsymbol{F}\overline{\boldsymbol{g}}\}$. Choosing $\varepsilon^2 \in [\varepsilon_{\min}^2, \varepsilon_{\max}^2[$ enables one to balance between a solution aligned with $\overline{\boldsymbol{g}}$ and the conventional LS estimate $\hat{\boldsymbol{h}}_{\text{ls}}$. Indeed, when $\varepsilon^2 = \varepsilon_{\min}^2$ there is only one vector that satisfies the constraint, namely $\hat{\boldsymbol{h}}_{\text{ls}}$, and therefore the latter is necessarily the solution to (17). On the other hand, when $\varepsilon^2 = \varepsilon_{\max}^2$ there exists a vector aligned with $\overline{\boldsymbol{g}}$, which drives the criterion down to 0 while satisfying the constraint: therefore, it is the solution. From these considerations, it is convenient to select ε^2 as

$$\varepsilon^2 = (1 - \alpha)\varepsilon_{\min}^2 + \alpha\varepsilon_{\max}^2 \tag{20}$$

where $\alpha \in [0, 1[$ is a user-defined parameter that sets how much confidence we place on the a priori information and on the information brought by the data. For α close to 1, the a priori information is deemed trustable while, when α approaches 0, it is discarded and only the data are used. The solution to the optimization problem in (17) is given in the next proposition.

Proposition 1. *The solution to* (17) *is given by*

$$\hat{\boldsymbol{h}}_{\text{cls}} = \boldsymbol{P}_{\overline{\boldsymbol{g}}}\hat{\boldsymbol{h}}_{\text{ls}} + \frac{\gamma\lambda_0}{1 + \gamma\lambda_0}\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\hat{\boldsymbol{h}}_{\text{ls}} \tag{21}$$

where

$$1 + \lambda_0\gamma = \sqrt{\frac{\varepsilon_{\max}^2 - \varepsilon_{\min}^2}{\varepsilon^2 - \varepsilon_{\min}^2}}. \tag{22}$$

Proof. see Appendix B.

It should be noted that the so-obtained estimator is given in *closed-form* and is thus very simple from a computational point of view. As expected, $\hat{\boldsymbol{h}}_{\text{cls}}$ is a linear combination of the a priori information through $\overline{\boldsymbol{g}}$ and of the information brought by the data $\boldsymbol{y}$ through $\hat{\boldsymbol{h}}_{\text{ls}}$.

The above result suggests looking for an optimal linear combiner (OLC) of $\boldsymbol{P}_{\overline{\boldsymbol{g}}}\hat{\boldsymbol{h}}_{\text{ls}}$ and $\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\hat{\boldsymbol{h}}_{\text{ls}}$ of the form

$$\hat{\boldsymbol{h}}_{\text{olc}} = \boldsymbol{P}_{\overline{\boldsymbol{g}}}\hat{\boldsymbol{h}}_{\text{ls}} + \beta\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\hat{\boldsymbol{h}}_{\text{ls}} \triangleq \boldsymbol{T}\hat{\boldsymbol{h}}_{\text{ls}}. \tag{23}$$

It is straightforward to show that

$$\begin{aligned}
&\mathcal{E}\{(\hat{\boldsymbol{h}}_{\text{olc}} - \boldsymbol{h})(\hat{\boldsymbol{h}}_{\text{olc}} - \boldsymbol{h})^H\} \\
&= \mathcal{E}\{(\boldsymbol{T}\hat{\boldsymbol{h}}_{\text{ls}} - \boldsymbol{h})(\boldsymbol{T}\hat{\boldsymbol{h}}_{\text{ls}} - \boldsymbol{h})^H\} \\
&= \boldsymbol{T}\mathcal{E}\{\hat{\boldsymbol{h}}_{\text{ls}}\hat{\boldsymbol{h}}_{\text{ls}}^H\}\boldsymbol{T}^H - \boldsymbol{T}\boldsymbol{h}\boldsymbol{h}^H - \boldsymbol{h}\boldsymbol{h}^H\boldsymbol{T} + \boldsymbol{h}\boldsymbol{h}^H \\
&= \boldsymbol{T}\left[\boldsymbol{h}\boldsymbol{h}^H + \frac{\sigma^2}{\gamma}\boldsymbol{I}\right]\boldsymbol{T}^H - \boldsymbol{T}\boldsymbol{h}\boldsymbol{h}^H - \boldsymbol{h}\boldsymbol{h}^H\boldsymbol{T} + \boldsymbol{h}\boldsymbol{h}^H \\
&= \frac{\sigma^2}{\gamma}\boldsymbol{T}\boldsymbol{T}^H + (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{h}\boldsymbol{h}^H(\boldsymbol{I} - \boldsymbol{T}).
\end{aligned} \tag{24}$$

Therefore, the mean-square error of $\hat{\boldsymbol{h}}_{\text{olc}}$ is given by

$$MSE(\hat{\boldsymbol{h}}_{\text{olc}}) = \frac{\sigma^2}{\gamma}\text{tr}\{\boldsymbol{T}\boldsymbol{T}^H\} + \boldsymbol{h}^H(\boldsymbol{I}-\boldsymbol{T})(\boldsymbol{I}-\boldsymbol{T})\boldsymbol{h}$$
$$= \frac{\sigma^2}{\gamma}[1+(M-1)\beta^2] + (1-\beta)^2\|\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}\|^2. \quad (25)$$

Minimizing the MSE in (25) yields the optimal value of β

$$\beta = \frac{\|\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}\|^2}{\|\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}\|^2 + \frac{(M-1)\sigma^2}{\gamma}}. \quad (26)$$

Of course, this optimal value depends on $\boldsymbol{h}$ which is unknown and therefore $\hat{\boldsymbol{h}}_{\text{olc}}$ is only an hypothetical estimator. However, it can serve as a reference and, in particular, it is of interest to compare $\hat{\boldsymbol{h}}_{\text{cls}}$ to $\hat{\boldsymbol{h}}_{\text{olc}}$.

5. Numerical examples

In this section, we compare the performances of the estimators derived previously, namely the LS estimator in (5), the CLS estimator in (21), the Bayesian estimator of Algorithm 1 and the OLC estimator in (23). We consider a uniform linear array with $M = 20$ elements spaced a half-wavelength apart. The steering vector for a receiver localized at angle θ is thus $\boldsymbol{a}(\theta) = [1\ e^{i\pi\sin\theta}\ \cdots\ e^{i\pi(M-1)\sin\theta}]^T$. The receiver is placed at the broadside of the array so that $\boldsymbol{a}_r = \boldsymbol{a}(0^\circ)$. The nominal antennas gain is $\overline{\boldsymbol{g}} = [1\ 1\ \cdots\ 1]^T$. Both the signals and the beams are orthogonal. The beams are in fact Fourier beams evenly spaced in spatial frequency. The signal to noise ratio is defined as

$$SNR = \frac{\|\overline{\boldsymbol{g}}\|^2|A|^2P_c}{\sigma^2}. \quad (27)$$

The estimators are evaluated in terms of their (normalized) mean-square error (MSE) defined as

$$MSE = \frac{1}{M}\mathcal{E}\{\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2\} \quad (28)$$

and the MSE is estimated from 500 independent Monte-Carlo trials. At each trial, $\boldsymbol{g}$ is randomly drawn as $g_k = \overline{g}_k\tilde{g}_k e^{i\tilde{\phi}_k}$ where the random variables $\tilde{g}_k$ and $\tilde{\phi}_k$ are independent. $10\log_{10}\tilde{g}_k$ is drawn from a Gaussian distribution with standard deviation $\sigma_{g-\text{dB}}$ while the phases $\tilde{\phi}_k$ are uniformly distributed over $[-\Delta_\phi, \Delta_\phi]$. Observe that as $\sigma_{g-\text{dB}}$ or Δ_ϕ increases, the true vector $\boldsymbol{g}$ may differ significantly from $\overline{\boldsymbol{g}}$, and hence the a priori knowledge is less accurate. This will of course impact the performance of all estimators, as illustrated below. We would like also to emphasize that, in this case, $\boldsymbol{g}$ is not Gaussian distributed and therefore the assumption in (6) does not hold. This enables one to test, in addition to its performance, the robustness of our Bayesian approach as the latter is fed with data that does not fulfill the assumptions on which it is based. Therefore, the Bayesian estimator derived previously does not correspond to the MMSE estimator and we will refer to it as MCMC in the figures. The Metropolis-within-Gibbs sampler of Algorithm 1 is used with $N_{bi} = 20$ and $N_r = 100$. As for the CLS approach, a value of $\alpha = 0.8$ is chosen to obtain ε^2 in (20). Finally, the OLC is implemented with β of (26) calculated from the exact value of $\boldsymbol{h}$.

We study the influence of the SNR, and the influence of $\sigma_{g-\text{dB}}$ and Δ_ϕ. The number of samples is set to $N = 32$. Figs. 2–4 deal with the case of small amplitude errors on $\boldsymbol{g}$ [$\sigma_{g-\text{dB}} = 1$ dB] and various phase errors, namely $\Delta_\phi = 15^\circ, 30^\circ, 45^\circ$ while the array is more significantly uncalibrated in Figs. 5–7 where $\sigma_{g-\text{dB}} = 3$ dB. Inspecting these figures, one can make the following observations:

(a) The MCMC approach provides the lowest MSE over all SNRs and orders of magnitude of the errors, among all estimators that can be implemented (i.e. omitting the OLC estimator). In particular, it achieves the desired

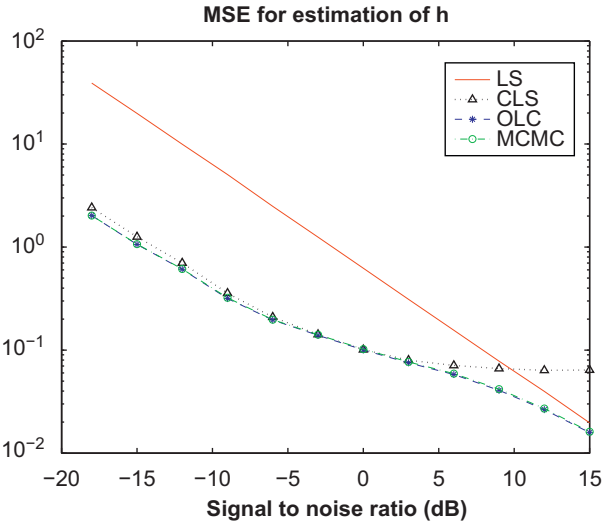

Fig. 2. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio. $N = 32$, $\sigma_{g-\text{dB}} = 1$ dB and $\Delta_\phi = 15^\circ$.

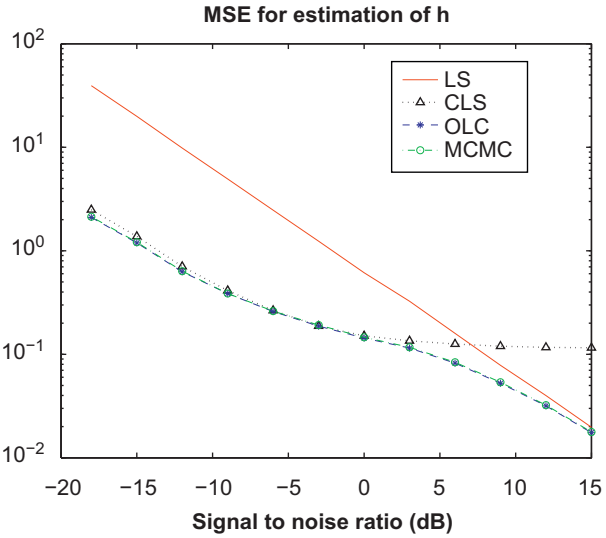

Fig. 3. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio. $N = 32$, $\sigma_{g-\text{dB}} = 1$ dB and $\Delta_\phi = 30^\circ$.

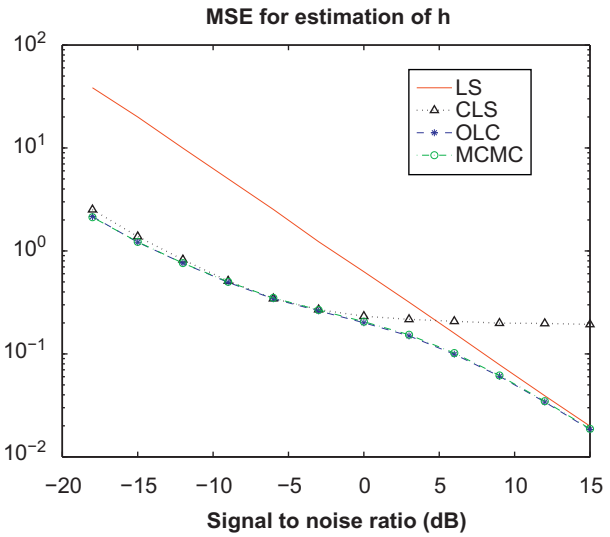

Fig. 4. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio. $N = 32$, $\sigma_{g-\mathrm{dB}} = 1\,\mathrm{dB}$ and $\Delta_\phi = 45^\circ$.

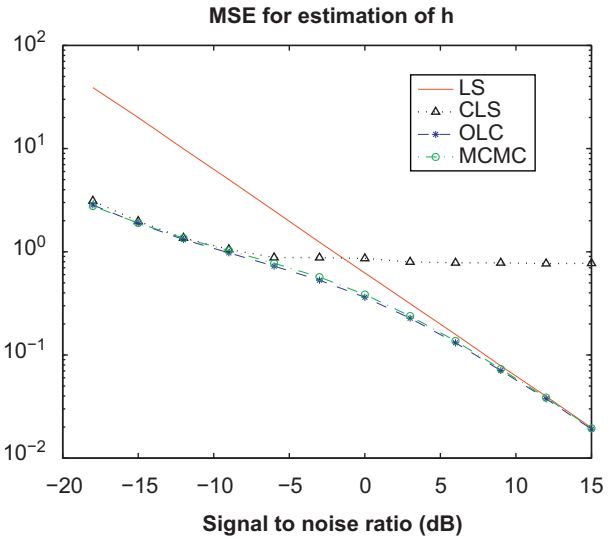

Fig. 5. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio. $N = 32$, $\sigma_{g-\mathrm{dB}} = 3\,\mathrm{dB}$ and $\Delta_\phi = 15^\circ$.

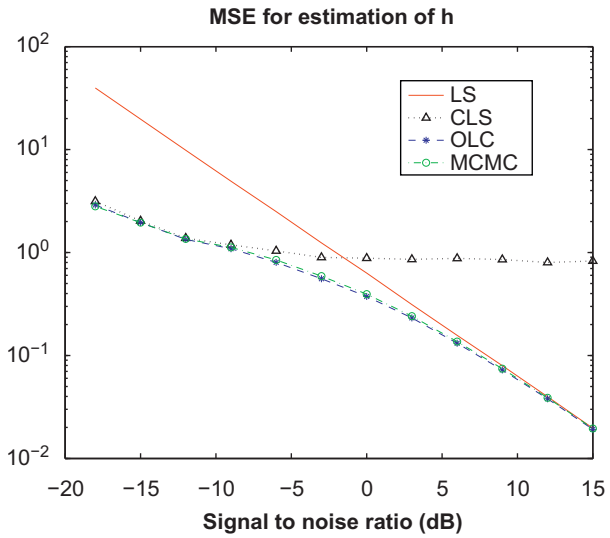

Fig. 6. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio. $N = 32$, $\sigma_{g-\mathrm{dB}} = 3\,\mathrm{dB}$ and $\Delta_\phi = 30^\circ$.

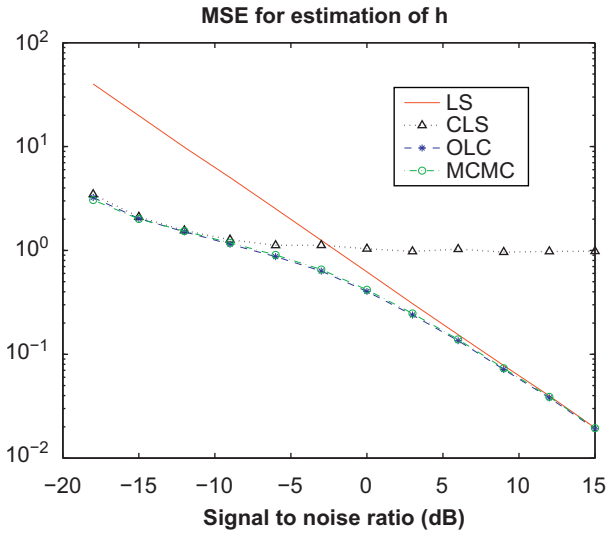

Fig. 7. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio. $N = 32$, $\sigma_{g-\mathrm{dB}} = 3\,\mathrm{dB}$ and $\Delta_\phi = 45^\circ$.

goal of significantly decreasing the LS estimator MSE at low SNR.

(b) The CLS estimator outperforms the conventional LS estimator at low signal to noise ratio, which was the main objective towards deriving new estimators. However, at high SNR the conventional LS estimator is seen to perform better than the CLS estimator. In fact, there exists a SNR threshold from which the MSE of the CLS estimator tends to be approximately constant and larger than that of the LS estimator. This threshold is smaller as the errors in $\boldsymbol{g}$ grow. This is logical since, with the rather high value of $\alpha = 0.8$ chosen, the CLS estimator grants too much importance to the a priori knowledge [and the latter is less and less reliable as $\sigma_{g-\mathrm{dB}}$ or Δ_ϕ increases] at medium to high SNR, and not enough to the data. Fig. 8 illustrates this phenomenon. There, we plot the MSE of the CLS estimator versus SNR for different values of α. It is clear from this figure that there does not exist a value of α that provides uniformly best performance, but that the choice of α is dictated by the SNR and the errors in $\boldsymbol{g}$. This fact is further investigated in Fig. 9. In fact, comparing (23) to (21), it is clear that the optimal value of β in (26) corresponds to an optimal value of λ_0 in (22), which itself amounts to an optimal value of ε^2 and hence of α in (20). Fig. 9 displays this optimal value of α [averaged over multiple random realizations of $\boldsymbol{g}$] versus SNR and $\sigma_{g-\mathrm{dB}}$, Δ_ϕ. As expected, the more reliable the a priori knowledge the larger α. Accordingly, α should decrease when SNR increases.

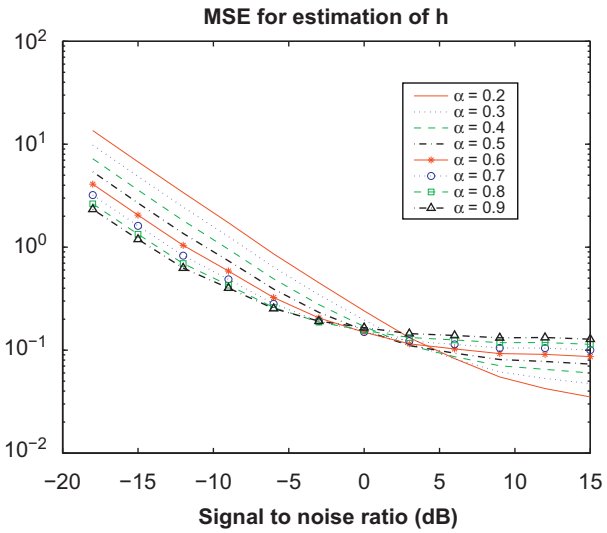

Fig. 8. MSE for estimation of $\boldsymbol{h}$ versus signal to noise ratio for different values of α. $N = 32$, $\sigma_{g-\mathrm{dB}} = 1$ dB and $\Delta_\phi = 30^\circ$.

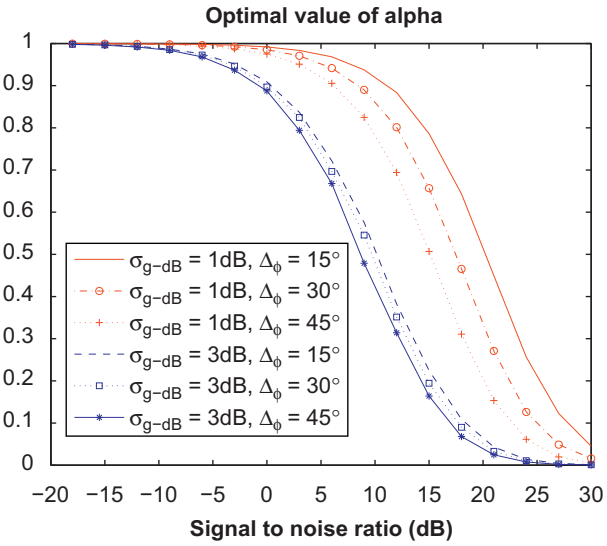

Fig. 9. Optimal value of α versus SNR and $\sigma_{g-\mathrm{dB}}$, Δ_ϕ. $N = 32$.

Therefore, the CLS estimator can perform well if one has a good idea of the operating SNR and of the accuracy of $\overline{\boldsymbol{g}}$. In fact, if we knew $\boldsymbol{h}$— which is unfortunately impossible—a simple linear combination of $\boldsymbol{P}_{\overline{\boldsymbol{g}}}\hat{\boldsymbol{h}}_{\mathrm{ls}}$ and $\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\hat{\boldsymbol{h}}_{\mathrm{ls}}$ could result in a quasi optimal solution, as the OLC estimator achieves the same performance as the MCMC for all SNRs and all values of $\sigma_{g-\mathrm{dB}}$ and Δ_ϕ.

To summarize this section, the Bayesian approach yields the lowest MSE whatever SNR or $\sigma_{g-\mathrm{dB}}$ and Δ_ϕ. The CLS estimator has the advantage of being very simple, more accurate than the LS estimator at low SNR but selecting the value of α is a delicate issue.

6. Conclusions

In this paper, we considered the problem of estimating the gains and phases of a transmitting array from the observation of signals received on a single antenna. New solutions, which are based on exploiting knowledge of the gains in the calibrated case, were proposed to improve over the conventional least-squares estimator, especially at low signal to noise ratio. Two different approaches were presented. A Bayesian approach was first investigated where the gain vector $\boldsymbol{g}$ is considered as a random variable. The MMSE estimator was derived and implemented through a Metropolis-within-Gibbs sampler. Its performance was shown to be very good, at the price of some computational complexity. A very different approach based on constrained least-squares was also presented, which results in a simple estimator. It consists of a weighted linear combination of the a priori knowledge and the information brought by the data. The method performs well as low SNR but degrades at high SNR, mainly because the choice of the weights is delicate. Both methods enable however to significantly improve over the LS estimator at low SNR.

Appendix A. Derivation of Jeffreys prior

In this appendix, we derive Jeffreys prior for the complex random variable $A = A_R + iA_I$ where A_R and A_I stand for the real and imaginary parts of A, respectively. Let us consider the generic model

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{n} \tag{29}$$

where $\boldsymbol{x}$ is drawn from a complex Gaussian distribution with mean $\overline{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{C}$, i.e. $\boldsymbol{x}\sim\mathcal{CN}(\overline{\boldsymbol{x}}, \boldsymbol{C})$, and $\boldsymbol{n}$ is a zero-mean complex Gaussian vector with covariance matrix $\sigma^2\boldsymbol{I}$, independent of $\boldsymbol{x}$. The model in (4) corresponds to $\overline{\boldsymbol{x}} = \boldsymbol{F}\overline{\boldsymbol{g}}$ and $\boldsymbol{C} = \sigma_g^2\boldsymbol{F}\boldsymbol{F}^H$. Under these assumptions, one has $\boldsymbol{y}\sim\mathcal{CN}(A\overline{\boldsymbol{x}}, \boldsymbol{\Gamma} = |A|^2\boldsymbol{C} + \sigma^2\boldsymbol{I})$. Therefore, the distribution of $\boldsymbol{y}$ conditionally to A is given by

$$f(\boldsymbol{y}|A) = \pi^{-N}|\boldsymbol{\Gamma}|^{-1}e^{-(\boldsymbol{y}-A\overline{\boldsymbol{x}})^H\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}-A\overline{\boldsymbol{x}})}. \tag{30}$$

In order to derive Jeffreys prior for A_R, A_I, we need to derive the Fisher information matrix (FIM) as Jeffreys prior is proportional to the square root of the determinant of the Fisher information matrix. The latter is obtained by differentiating twice the log-likelihood function

$$\Lambda(\boldsymbol{y}|A) = -N\ln\pi - \ln|\boldsymbol{\Gamma}| - (\boldsymbol{y} - A\overline{\boldsymbol{x}})^H\boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - A\overline{\boldsymbol{x}}) \tag{31}$$

with respect to $[A_R \;\; A_I]^T$, and by taking expectation. Using the fact that

$$\frac{\partial\boldsymbol{\Gamma}}{\partial A_R} = 2A_R\boldsymbol{C};\quad \frac{\partial\boldsymbol{\Gamma}^{-1}}{\partial A_R} = -2A_R\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1};\quad \frac{\partial\boldsymbol{\Gamma}}{\partial A_I} = 2A_I\boldsymbol{C};$$
$$\frac{\partial\boldsymbol{\Gamma}^{-1}}{\partial A_I} = -2A_I\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1}$$

it is straightforward to show that

$$\frac{\partial\ln|\boldsymbol{\Gamma}|}{\partial A_R} = \mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial A_R}\right\} = 2A_R\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} \tag{32a}$$

$$\frac{\partial\ln|\boldsymbol{\Gamma}|}{\partial A_I} = \mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial A_I}\right\} = 2A_I\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\}. \tag{32b}$$

Differentiating again leads to

$$\frac{\partial^2 \ln|\boldsymbol{\Gamma}|}{\partial A_R^2} = 2\,\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} - 4A_R^2\,\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} \tag{33a}$$

$$\frac{\partial^2 \ln|\boldsymbol{\Gamma}|}{\partial A_I^2} = 2\,\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} - 4A_I^2\,\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} \tag{33b}$$

$$\frac{\partial^2 \ln|\boldsymbol{\Gamma}|}{\partial A_R \partial A_I} = -4A_R A_I\,\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\}. \tag{33c}$$

For the sake of convenience, let us temporarily note $\boldsymbol{z} = \boldsymbol{y} - A\overline{\boldsymbol{x}}$ and $\boldsymbol{u} = \boldsymbol{\Gamma}^{-1}\boldsymbol{z}$. Since

$$\frac{\partial \boldsymbol{z}}{\partial A_R} = -\overline{\boldsymbol{x}};\quad \frac{\partial \boldsymbol{u}}{\partial A_R} = -2A_R\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} - \boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}};\quad \frac{\partial \boldsymbol{z}}{\partial A_I} = -i\overline{\boldsymbol{x}};$$
$$\frac{\partial \boldsymbol{u}}{\partial A_R} = -2A_I\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} - i\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}},$$

it follows that

$$\frac{\partial \boldsymbol{z}^H\boldsymbol{u}}{\partial A_R} = -\overline{\boldsymbol{x}}^H\boldsymbol{u} - \boldsymbol{u}^H\overline{\boldsymbol{x}} - 2A_R\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u} \tag{34a}$$

$$\frac{\partial \boldsymbol{z}^H\boldsymbol{u}}{\partial A_I} = i\overline{\boldsymbol{x}}^H\boldsymbol{u} - i\boldsymbol{u}^H\overline{\boldsymbol{x}} - 2A_I\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u}. \tag{34b}$$

Using the fact that

$$\frac{\partial \boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u}}{\partial A_R} = -\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} - \boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} - 4A_R\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} \tag{35a}$$

$$\frac{\partial \boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u}}{\partial A_I} = i\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} - i\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} - 4A_I\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} \tag{35b}$$

one obtains the second-order derivatives as

$$\frac{\partial^2 \boldsymbol{z}^H\boldsymbol{u}}{\partial A_R^2} = 2\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} + 4A_R\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} + 4A_R\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} + 8A_R^2\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} - 2\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u} \tag{36a}$$

$$\frac{\partial^2 \boldsymbol{z}^H\boldsymbol{u}}{\partial A_I^2} = 2\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} - 4iA_I\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} + 4iA_I\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} + 8A_I^2\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} - 2\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u} \tag{36b}$$

$$\frac{\partial^2 \boldsymbol{z}^H\boldsymbol{u}}{\partial A_R \partial A_I} = -2iA\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u} + 2iA^*\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} + 8A_RA_I\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u}. \tag{36c}$$

Now, since $\mathcal{E}\{\boldsymbol{u}\} = \boldsymbol{0}$ and

$$\mathcal{E}\{\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{u}\} = \mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} \tag{37a}$$

$$\mathcal{E}\{\boldsymbol{u}^H\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{u}\} = \mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\} \tag{37b}$$

it follows that the FIM is given by

$$\boldsymbol{I}(A_R, A_I) = \begin{pmatrix} 2\alpha + 4A_R^2\beta & 4A_RA_I\beta \\ 4A_RA_I\beta & 2\alpha + 4A_I^2\beta \end{pmatrix} \tag{38}$$

where $\alpha \triangleq \overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}}$ and $\beta \triangleq \mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{C}\}$. The previous equation holds for any matrix $\boldsymbol{C}$. Let us now consider our case where $\boldsymbol{C} = \sigma_g^2\boldsymbol{F}\boldsymbol{F}^H$. Using some matrix inversion lemma, it is straightforward to show that

$$\overline{\boldsymbol{x}}^H\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{x}} = \sigma^{-2}\overline{\boldsymbol{g}}^H[(\boldsymbol{F}^H\boldsymbol{F})^{-1} + \sigma^{-2}|A|^2\sigma_g^2\boldsymbol{I}]^{-1}\overline{\boldsymbol{g}} \tag{39a}$$

$$\boldsymbol{\Gamma}^{-1}\boldsymbol{C} = \sigma^{-2}\sigma_g^2\boldsymbol{F}[\boldsymbol{I} + \sigma^{-2}|A|^2\sigma_g^2\boldsymbol{F}^H\boldsymbol{F}]^{-1}\boldsymbol{F}^H. \tag{39b}$$

Therefore, with $\boldsymbol{T}^{-1} = (\boldsymbol{F}^H\boldsymbol{F})^{-1} + \sigma^{-2}\sigma_g^2|A|^2\boldsymbol{I}$, one obtains

$$\alpha = \sigma^{-2}\overline{\boldsymbol{g}}^H\boldsymbol{T}\overline{\boldsymbol{g}} \tag{40a}$$

$$\beta = (\sigma^{-2}\sigma_g^2)^2\mathrm{tr}\{\boldsymbol{T}^H\boldsymbol{T}\}. \tag{40b}$$

In the particular case where $\boldsymbol{F}^H\boldsymbol{F} = \gamma\boldsymbol{I}$, one finally gets

$$\alpha = \frac{\gamma\sigma^{-2}\|\overline{\boldsymbol{g}}\|^2}{1 + \gamma\sigma^{-2}\sigma_g^2|A|^2} \tag{41a}$$

$$\beta = \frac{M(\gamma\sigma^{-2}\sigma_g^2)^2}{[1 + \gamma\sigma^{-2}\sigma_g^2|A|^2]^2}. \tag{41b}$$

Jeffreys prior is obtained as the square-root of the determinant of $\boldsymbol{I}(A_R, A_I)$ and the latter can be written as

$$|\boldsymbol{I}(A_R, A_I)| = 4\alpha[\alpha + 2\beta|A|^2] \propto \frac{1 + \gamma\sigma^{-2}\sigma_g^2[1 + 2M\sigma_g^2\|\overline{\boldsymbol{g}}\|^{-2}]|A|^2}{[1 + \gamma\sigma^{-2}\sigma_g^2|A|^2]^3}. \tag{42}$$

Appendix B. Proof of Proposition 1

In this appendix, we show how to solve the constrained LS problem in (17) which we recast here for the sake of convenience

$$\min_{A,\boldsymbol{h}} \|\boldsymbol{h} - A\overline{\boldsymbol{g}}\|^2 \text{ subject to } \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 \le \varepsilon^2. \tag{43}$$

Let us first minimize the criterion with respect to (w.r.t.) A. It is well known that

$$\min_{A} \|\boldsymbol{h} - A\overline{\boldsymbol{g}}\|^2 = \boldsymbol{h}^H\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h} \tag{44}$$

so that we are left with the following optimization problem:

$$\min_{\boldsymbol{h}}\ \boldsymbol{h}^H\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h} \text{ subject to } \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 \le \varepsilon^2. \tag{45}$$

To begin with, we show that the inequality constraint in (45) is in fact an equality constraint. To see this, let us temporarily denote by $\boldsymbol{h}_0$ the solution and let us assume that $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}_0\|^2 < \varepsilon^2$. Let also note

$$A_1 = \underset{A}{\mathrm{argmin}}\|\boldsymbol{y} - A\boldsymbol{F}\overline{\boldsymbol{g}}\|^2 = \frac{\overline{\boldsymbol{g}}^H\boldsymbol{F}^H\boldsymbol{y}}{\overline{\boldsymbol{g}}^H\boldsymbol{F}^H\boldsymbol{F}\overline{\boldsymbol{g}}} \tag{46a}$$

$$\boldsymbol{h}_1 = A_1\overline{\boldsymbol{g}} = \frac{\overline{\boldsymbol{g}}^H\boldsymbol{F}^H\boldsymbol{y}}{\overline{\boldsymbol{g}}^H\boldsymbol{F}^H\boldsymbol{F}\overline{\boldsymbol{g}}}\overline{\boldsymbol{g}}. \tag{46b}$$

From the assumption made in (19), we clearly have $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}_1\|^2 = \varepsilon_{\max}^2 > \varepsilon^2$. Let

$$\boldsymbol{h}(\mu) = \boldsymbol{h}_0 + \mu(\boldsymbol{h}_1 - \boldsymbol{h}_0);\quad \mu \in [0, 1]. \tag{47}$$

Eq. (47) defines a trajectory from $\boldsymbol{h}_0$ to $\boldsymbol{h}_1$ and therefore, there exists at least one value of $\mu \in]0, 1[$ such that $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}(\mu)\|^2 = \varepsilon^2$. Now, for any $\mu \in]0, 1[$,

$$\boldsymbol{h}(\mu)^H\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}(\mu) = [(1-\mu)\boldsymbol{h}_0 + \boldsymbol{h}_1]^H\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}[(1-\mu)\boldsymbol{h}_0 + \boldsymbol{h}_1] = (1-\mu)^2\boldsymbol{h}_0\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}_0 < \boldsymbol{h}_0\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}_0. \tag{48}$$

Therefore, there would exist a vector $\boldsymbol{h}(\mu)$ such that $\boldsymbol{h}(\mu)^H\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}(\mu) < \boldsymbol{h}_0\boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}\boldsymbol{h}_0$ and that would satisfy the constraint, which is in contradiction with the fact that $\boldsymbol{h}_0$ is the solution. Therefore, the inequality constraint in (45) is necessarily an equality constraint and hence we are finally

left with the problem of solving

$$\min_{\boldsymbol{h}} \boldsymbol{h}^H \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp} \boldsymbol{h} \text{ subject to } \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 = \varepsilon^2. \tag{49}$$

In order to solve (49) a Lagrange multiplier technique is used. The Lagrangian can be written as

$$L(\boldsymbol{h}, \lambda) = \boldsymbol{h}^H \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp} \boldsymbol{h} + \lambda[\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}\|^2 - \varepsilon^2] \tag{50}$$

where $\lambda > 0$ is the real-valued Lagrange multiplier. One can rewrite (50) as

$$\begin{aligned} L(\boldsymbol{h}, \lambda) = {} & [\boldsymbol{h} - \lambda \boldsymbol{Q}^{-1} \boldsymbol{F}^H \boldsymbol{y}]^H \boldsymbol{Q} [\boldsymbol{h} - \lambda \boldsymbol{Q}^{-1} \boldsymbol{F}^H \boldsymbol{y}] \\ & + \lambda(\boldsymbol{y}^H \boldsymbol{y} - \varepsilon^2) - \lambda^2 \boldsymbol{y}^H \boldsymbol{F} \boldsymbol{Q}^{-1} \boldsymbol{F}^H \boldsymbol{y} \end{aligned} \tag{51}$$

with $\boldsymbol{Q} = \lambda \boldsymbol{F}^H \boldsymbol{F} + \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}$. For a given λ, $L(\boldsymbol{h}, \lambda)$ is thus minimum for

$$\boldsymbol{h}(\lambda) = \lambda(\lambda \boldsymbol{F}^H \boldsymbol{F} + \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp})^{-1} \boldsymbol{F}^H \boldsymbol{y}. \tag{52}$$

Let us find a more convenient expression for the above $\boldsymbol{h}(\lambda)$. Observing that $\boldsymbol{F}^H \boldsymbol{F} = \gamma \boldsymbol{I}$ and $\hat{\boldsymbol{h}}_{\mathrm{ls}} = \gamma^{-1} \boldsymbol{F}^H \boldsymbol{y}$, it follows that:

$$\begin{aligned} \boldsymbol{h}(\lambda) &= \lambda[\lambda\gamma \boldsymbol{I} + \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp}]^{-1} \boldsymbol{F}^H \boldsymbol{y} \\ &= \lambda[(1 + \lambda\gamma) \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp} + \lambda\gamma \boldsymbol{P}_{\overline{\boldsymbol{g}}}]^{-1} \boldsymbol{F}^H \boldsymbol{y} \\ &= \lambda[(1 + \lambda\gamma)^{-1} \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp} + (\lambda\gamma)^{-1} \boldsymbol{P}_{\overline{\boldsymbol{g}}}] \boldsymbol{F}^H \boldsymbol{y} \\ &= \boldsymbol{P}_{\overline{\boldsymbol{g}}} \hat{\boldsymbol{h}}_{\mathrm{ls}} + \frac{\lambda\gamma}{1 + \lambda\gamma} \boldsymbol{P}_{\overline{\boldsymbol{g}}}^{\perp} \hat{\boldsymbol{h}}_{\mathrm{ls}}. \end{aligned} \tag{53}$$

In order to obtain λ, we enforce the constraint that $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}(\lambda)\|^2 = \varepsilon^2$. Towards this end, let $\overline{\boldsymbol{G}}_{\perp}$ denote an orthornormal basis for the space orthogonal to $\overline{\boldsymbol{g}}$ and observe that

$$\begin{aligned} \boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}(\lambda) &= \boldsymbol{y} - \frac{\lambda}{1 + \lambda\gamma} \boldsymbol{F}\overline{\boldsymbol{G}}_{\perp} \overline{\boldsymbol{G}}_{\perp}^H \boldsymbol{F}^H \boldsymbol{y} - \gamma^{-1} \frac{\boldsymbol{F}\overline{\boldsymbol{g}}\overline{\boldsymbol{g}}^H \boldsymbol{F}^H \boldsymbol{y}}{\overline{\boldsymbol{g}}^H \overline{\boldsymbol{g}}} \\ &= \boldsymbol{y} - \frac{\lambda\gamma}{1 + \lambda\gamma} \boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{G}}_{\perp}} \boldsymbol{y} - \boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{g}}} \boldsymbol{y} \\ &= [\boldsymbol{P}_{\boldsymbol{F}} + \boldsymbol{P}_{\boldsymbol{F}}^{\perp}] \boldsymbol{y} - \frac{\lambda\gamma}{1 + \lambda\gamma} \boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{G}}_{\perp}} \boldsymbol{y} \\ &\quad - [\boldsymbol{P}_{\boldsymbol{F}} - \boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{G}}_{\perp}}] \boldsymbol{y} \\ &= \boldsymbol{P}_{\boldsymbol{F}}^{\perp} \boldsymbol{y} + \frac{1}{1 + \lambda\gamma} \boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{G}}_{\perp}} \boldsymbol{y}. \end{aligned} \tag{54}$$

It ensues that

$$\begin{aligned} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}(\lambda)\|^2 &= \|\boldsymbol{P}_{\boldsymbol{F}}^{\perp} \boldsymbol{y}\|^2 + \frac{1}{(1 + \lambda\gamma)^2} \|\boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{G}}_{\perp}} \boldsymbol{y}\|^2 \\ &= \|\boldsymbol{P}_{\boldsymbol{F}}^{\perp} \boldsymbol{y}\|^2 + \frac{1}{(1 + \lambda\gamma)^2} [\|\boldsymbol{P}_{\boldsymbol{F}} \boldsymbol{y}\|^2 - \|\boldsymbol{P}_{\boldsymbol{F}\overline{\boldsymbol{g}}} \boldsymbol{y}\|^2] \\ &= \|\boldsymbol{P}_{\boldsymbol{F}}^{\perp} \boldsymbol{y}\|^2 + \frac{1}{(1 + \lambda\gamma)^2} [\|\boldsymbol{P}_{\overline{\boldsymbol{F}\boldsymbol{g}}}^{\perp} \boldsymbol{y}\|^2 - \|\boldsymbol{P}_{\boldsymbol{F}}^{\perp} \boldsymbol{y}\|^2] \\ &= \varepsilon_{\min}^2 + \frac{1}{(1 + \lambda\gamma)^2} [\varepsilon_{\max}^2 - \varepsilon_{\min}^2]. \end{aligned} \tag{55}$$

Therefore, $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{h}(\lambda)\|^2 = \varepsilon^2$ leads to

$$(1 + \lambda\gamma)^2 = \frac{\varepsilon_{\max}^2 - \varepsilon_{\min}^2}{\varepsilon^2 - \varepsilon_{\min}^2}. \tag{56}$$

The solution $\hat{\boldsymbol{h}}_{\mathrm{cls}}$ to the minimization problem is thus $\boldsymbol{h}(\lambda)$ in (53) with λ given in (56), which concludes the proof.

References

[1] A. Jacomb-Hood, E. Lier, Multibeam active phased arrays for communications satellites, IEEE Microwave Magazine 1 (4) (2000) 40–47.

[2] D.R. Fuhrmann, Estimation of sensor gain and phase, IEEE Transactions on Signal Processing 42 (1) (1994) 77–87.

[3] V.C. Soon, L. Tong, Y.F. Huang, R. Liu, A subspace method for estimating gains and phases, IEEE Transactions on Signal Processing 42 (4) (1994) 973–976.

[4] M. Viberg, A.L. Swindlehurst, A Bayesian approach to auto-calibration for parametric array processing, IEEE Transactions on Signal Processing 42 (12) (1994) 3495–3507.

[5] C.-Y. Tseng, D.D. Feldman, L.J. Griffiths, Steering vector estimation in uncalibrated arrays, IEEE Transactions on Signal Processing 43 (6) (1995) 1397–1412.

[6] A.J. Weiss, B. Friedlander, Almost blind steering vector estimation using second-order moments, IEEE Transactions on Signal Processing 44 (4) (1996) 1024–1027.

[7] D. Astély, A.L. Swindlehurst, B. Ottersten, Spatial signature estimation for uniform linear arrays with unknown receiver gains and phases, IEEE Transactions on Signal Processing 47 (8) (1999) 2128–2138.

[8] B. Flanagan, K.L. Bell, Array self-calibration with large sensor position errors, Signal Processing 81 (10) (2001) 2201–2214.

[9] A.-J. Boonstra, A.-J. van der Veen, Gain calibration methods for radio telescope arrays, IEEE Transactions on Signal Processing 51 (11) (2003) 25–38.

[10] S. van der Tol, B. Jeffs, A.-J. van der Veen, Self-calibration for the lofar radio astronomical array, IEEE Transactions on Signal Processing 55 (9) (2007) 4497–4510.

[11] S.D. Silverstein, Application of orthogonal codes to the calibration of active phased array antennas for communication satellites, IEEE Transactions on Signal Processing 45 (1) (1997) 206–218.

[12] D. Hounam, M. Schwerdt, M. Zink, Active antenna module characterisation by pseudo-noise gating, in: Proceedings of the 25th ESA Workshop on Satellite Antenna Technology, Noordwijk, The Netherlands, 2002.

[13] B. Bräutigam, M. Schwerdt, M. Bachmann, M. Stangl, Individual T/R module characterisation of the TerraSAR-X active phased array antenna by calibration pulse sequences with orthogonal codes, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 2007, pp. 5202–5205.

[14] M. Oodo, R. Miura, A remote calibration for DBF transmitting array antennas by using synchronous orthogonal codes, in: Proceedings of the IEEE International Symposium on Antennas Propagation, vol. 2, 1999, pp. 1428–1431.

[15] M. Oodo, R. Miura, A remote calibration for a transmitting array antenna by using synchronous orthogonal codes, IEICE Transactions on Communications E84-B (7) (2001) 1808–1815.

[16] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[17] C.P. Robert, The Bayesian Choice—From Decision-Theoretic Foundations to Computational Implementation, Springer, New York, 2007.

[18] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, second ed., Springer, New York, 2004.

[19] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., Chapman & Hall, New York, 2004.

[20] S.P. Brooks, A. Gelman, General methods for monitoring convergence of iterative simulations, Journal of Computational and Graphical Statistics 7 (4) (1998) 434–455.

[21] S. Bidon, O. Besson, J.-Y. Tourneret, A Bayesian approach to adaptive detection in non-homogeneous environments, IEEE Transactions on Signal Processing 56 (1) (2008) 205–217.

[22] O. Besson, S. Bidon, J.-Y. Tourneret, Covariance matrix estimation with heterogeneous samples, IEEE Transactions on Signal Processing 56 (3) (2008) 909–920.