# Gain Calibration Methods for Radio Telescope Arrays

Albert-Jan Boonstra and Alle-Jan van der Veen, *Senior Member, IEEE*

***Abstract*—In radio telescope arrays, the complex receiver gains and sensor noise powers are initially unknown and have to be calibrated. Gain calibration can enhance the quality of astronomical sky images and, moreover, improve the effectiveness of array signal processing techniques for interference mitigation and spatial filtering. A challenging aspect is that the signal-to-noise ratio (SNR) is usually well below 0 dB, even for the brightest sky sources. The calibration method considered here consists of observing a single point source and extracting the gain and noise parameters from the estimated covariance matrix. We present several closed-form and iterative identification algorithms for this. Weighted versions of the algorithms are proven to be asymptotically efficient. The algorithms are validated by simulations and application to experimental data observed at the Westerbork Synthesis Radio Telescope (WSRT).**

***Index Terms*—Array calibration, covariance matching, factor analysis, radio astronomy, weighted least squares.**

## I. Introduction

IN CONTRAST to most communications systems, in radio astronomy the sources of interest are usually much weaker than the instantaneous system noise levels, with signal to noise (SNR) levels much below 0 dB even for the strongest sources. Integration times of several hours to more than ten hours are needed to obtain sky images with acceptable sensitivity. Astronomical sources can be broadband or narrowband, and accurate estimates of the telescope gains are necessary to compute the astronomical source power distribution, in radio astronomy also known as the surface brightness.

For an array of telescopes (see Fig. 1), not only the gain of the main beam of each telescope needs to be estimated but the phase differences between the telescopes, and the power of the noise present at each receiver need to be estimated as well. Two techniques are widely used. In imaging applications, a "blind" iterative self-calibration technique [1], [2] starts from an initial estimate of the parameters and adapts them until the resulting image matches a prior parametric model of the field of interest (usually a point source model). A second standard technique is to obtain dedicated calibration observations of a part of the sky that contains a single known, relatively strong point source, and this is the technique that we consider in this paper.

Manuscript received April 2, 2002; revised August 29, 2002. This work was supported in part by the NOEMI project of the STW under Contract DEL77-4476. The associate editor coordinating the review of this paper and approving it for publication was Dr. Olivier Besson.

A.-J. Boonstra is with the Department of Electrical Engineering/DIMES, Delft University of Technology, Delft, The Netherlands. He is also with ASTRON, Foundation for Research in Astronomy, Dwingeloo, The Netherlands (e-mail: boonstra@astron.nl).

A.-J. van der Veen is with the Department of Electrical Engineering/DIMES, Delft University of Technology, Delft, The Netherlands (e-mail: allejan@cas.et.tudelft.nl).

Digital Object Identifier 10.1109/TSP.2002.806588

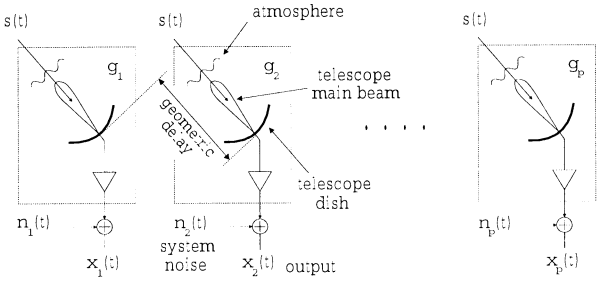


Fig. 1. Radio telescope array.

Several algorithms to estimate the parameters from these observations have existed for a long time [3]–[5], and until recently, their performance was considered satisfactory. For example, a typical 12-hr duration observation at the Westerbork Synthesis Radio Telescope (WSRT), which is a 3-km linear array of 14 telescope dishes of 25 m diameter located in The Netherlands, is usually calibrated with two dedicated short calibration observations prior to and after the 12-hr observation run. At 1420 MHz and under optimal atmospheric and ionospheric conditions, the resulting gain phase accuracy is about 5°, and the gain magnitude accuracy about a few percent. The gain estimates are assumed to be valid for any direction in the sky. If a better accuracy is required or if the atmospheric and ionospheric conditions are unstable, intermediate calibration measurements need to be carried out, and calibration sources closer to the astronomical source of interest can be selected. Since these are usually weaker, longer integration periods are needed to reach sufficient accuracy. Observation time is expensive, and there is a growing need for data efficient estimation algorithms.

The advent of a new generation of radio telescopes such as the square kilometer array (SKA) radio telescope [6] has sparked new interest in the issue. SKA will be a phased array of about $10^6$ elements with receivers that are not as matched and much noisier than the classical arrays. Hence, array calibration is both harder and more important. A proposed low-frequency array (LOFAR) will suffer from atmospheric disturbances that can vary within minutes, and calibration will have to be online and work with short data sets; hence, efficient algorithms are needed. A second reason for renewed interest in gain calibration issues is the recent attention for radio frequency interference (RFI) suppression techniques. Advanced array signal processing techniques such as spatial filtering are very adequate in this context [7], but most algorithms rely on spatially white noise models that are valid only after accurate calibration. In this paper, we will study calibration techniques valid for the 14-dish Westerbork array, but the results are more generally applicable.

As mentioned before, we consider the standard procedure for estimating the complex gain and noise power of each tele-

scope, which starts by pointing the telescopes at a relatively strong astronomical source (e.g., SNR $= -20$ dB). Each telescope output signal is the sum of the telescope system noise (uncorrelated among the telescopes) and the astronomical source flux, which is correlated, multiplied by the telescope gain. The source flux is the same for each of the telescopes, but the telescope gain and noise power usually are not. The gain consists of the combined effect of atmospheric disturbances, telescope geometry, receiver characteristics, and electronic (amplifier) gain. The noise consists mostly of thermal system noise, and differs between the receivers by several decibels.

The output of the back-end processing stage is a sequence of covariance matrices formed by cross-correlation of all telescope outputs. In this paper, we aim to estimate the complex gain factors and the system noise power from a recorded covariance matrix, assuming that we observe a single astronomical source whose flux is known from tables. We present four algorithms to extract these parameters.

The telescope gains are frequency dependent; therefore, the received frequency band is divided into subbands, and the parameter estimation is repeated for each frequency. We assume that the subbands are sufficiently narrow so that the geometric delay compensation (i.e., the delay of the astronomical signal across the array) can be carried out via phase rotations. Under the narrowband assumption, it is sufficient to discuss the parameter estimation at a single frequency.[1]

As we will show in Section II, the calibration problem is essentially reduced to estimating the parameters of a $p \times p$ covariance matrix of the form

$$\mathbf{R} = \mathbf{g}\mathbf{g}^H + \mathbf{D} \tag{1}$$

where $\mathbf{g}$ is a gain vector, and $\mathbf{D}$ is a diagonal matrix representing the noise covariance. This problem is not entirely unknown in signal processing. There is some recent literature on direction-of-arrival estimation in colored noise, with models of the form $\mathbf{R} = \mathbf{a}(\theta)\mathbf{a}(\theta)^H + \mathbf{D}$, where $\mathbf{a}(\theta)$ is an array response vector corresponding to the direction of arrival of a source. This problem and generalizations to multiple sources have been considered, for example, in [8] and [9]. There is also a significant body of earlier research that considers more general structured noise models; see [10], [11], and references therein. Our problem differs in that we consider an *unstructured* complex vector $\mathbf{g}$ and a *diagonal* noise covariance matrix $\mathbf{D}$. This enables some interesting closed-form estimators.

The model (1) also appears in the statistical literature as a (rank-1) *factor analysis* problem [12], [13]. Factor analysis is a mature field that has seen much activity in the 1960s through the 1980s. Although they are quite relevant to array signal processing, the results are apparently little known in this field. Many results also need to be translated to the complex domain.

Our main contribution in this paper is that we give several new algorithms to estimate the gain and noise parameters, including simple closed-form algorithms. We start with posing the data model, deriving the Cramér–Rao bound (CRB), and formulating a maximum likelihood (ML) estimation problem (Section II). This does not directly lead to useful algorithms since the number of parameters to be estimated can be large [$(3p-1)$, where the number of sensors $p$ is in the order of 10–60 for classical arrays and potentially much more in future arrays]. We then present an asymptotically efficient Least Squares (LS) cost function (Section III-A) and subsequently derive several iterative and closed-form algorithms (Sections III-B–E). The algorithms are first verified using simulations (Section IV) and then applied to experimental data collected at WSRT (Section V).

*Notation:* The complex conjugate (Hermitian) transpose is denoted by $H$, the transpose operator by $t$, the complex conjugate by overbar $\bar{}$, and the matrix pseudo inverse (Moore–Penrose inverse) by $\dagger$. The Kronecker matrix product is represented by $\otimes$, $\odot$ is the element-wise matrix multiplication (Hadamard product), and $\circ$ is the Khatri–Rao product, which is a column-wise Kronecker product $(\mathbf{A} \circ \mathbf{B} = (\mathbf{a}_1 \otimes \mathbf{b}_1 \;\; \mathbf{a}_2 \otimes \mathbf{b}_2 \cdots))$. An estimated value is denoted by $\hat{}$, $\mathrm{diag}(\cdot)$ and either converts a vector into a diagonal matrix (with the vector placed on the main diagonal) or, if applied to a matrix, sets the off-diagonal elements to zero, and $\mathrm{vecdiag}(\cdot)$ returns the main diagonal of a matrix stacked into a vector. Finally, $\delta_{ij}$ is the Kronecker delta function, $\mathbf{e}_i$ is the $i$th unit vector (all zeros except for the $i$th element), $\mathbf{1}$ is a vector containing ones, and $\mathbf{I}_c$ is the complement of $\mathbf{I}$: $\mathbf{I}_c = \mathbf{1}\mathbf{1}^t - \mathbf{I}$.

Define the selection matrix $\mathbf{J}$: $(p^2 - p) \times p^2$ such that for any $p \times p$ matrix $\mathbf{A}$, $\mathrm{vec}'(\mathbf{A}) = \mathbf{J}\mathrm{vec}(\mathbf{A})$, where $\mathrm{vec}'$ is the vectorizaton omitting the diagonal entries. Note that $\mathbf{J}$ is easily constructed from a $p^2 \times p^2$ identity matrix by removing its rows $1, (p+1)+1, 2(p+1)+1, \ldots, p^2$. Then, $\mathbf{J}^H\mathbf{J} = \mathbf{I} \otimes \mathbf{I} - (\mathbf{I} \circ \mathbf{I})(\mathbf{I} \circ \mathbf{I})^H$.

## II. Data Model and Preliminary Results

### A. Data Model Description

Consider a telescope array as in Fig. 1, and assume that during the calibration observation, the telescopes are pointed at a single radio source in the sky, placed in the center of the field of view. Let $x_i(t)$ be the complex baseband signal at a certain frequency $f$ at the output of the receiver of element $i$ at time $t$. We assume that the frequency bin is sufficiently narrow for the maximal propagation delay of a signal across the array to be much smaller than the inverse bandwidth so that it may be represented by a phase shift. Then, $x_i(t)$ can be modeled as

$$x_i(t) = \gamma_i e^{\imath\phi_i} a_i s(t) + n_i(t)$$

where $\gamma_i$ is the overall amplitude gain of the receiver system and the atmospheric disturbances, $\phi_i$ is the corresponding phase shift, $s(t)$ is the flux of the impinging source signal, $a_i = e^{-\imath 2\pi f \tau_i}$ is the phase shift due to the propagation delay $\tau_i$ of $s(t)$ across the array, as compared to its arrival at the first element, and $n_i$ is the system noise.

Assuming that we have $p$ elements, we can stack the $x_i(t)$ into a vector $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_p(t)]^t$. Similarly, we define

$$\mathbf{g} = \left[\gamma_1 e^{\imath\phi_1} a_1, \ldots, \gamma_p e^{\imath\phi_p} a_p\right]^t$$

[1] If a model is assumed for the frequency behavior of the parameters, then the information from different frequencies should obviously be combined. This is outside the scope of the paper.

and $\mathbf{n}(t) = [n_1(t), n_2(t), \ldots, n_p(t)]^t$. We thus arrive at the data model

$$\mathbf{x}(t) = \mathbf{g}s(t) + \mathbf{n}(t).$$

We make the following model assumptions.

1) The source signal is zero mean Gaussian, temporally independent identically distributed (i.i.d.) with signal power $\sigma_s^2 = \mathrm{E}\{|s(t)|^2\}$ known from tables.
2) The noise signal is zero mean Gaussian, temporally i.i.d., spatially uncorrelated among the sensors, with unknown power $\mathrm{E}\{|n_i|^2\} = \sigma_i^2$ and independent of the source signal.
3) The gains $\gamma_i$ and phases $\phi_i$ are unknown constants.
4) The geometry and looking directions of the telescopes are known, so the phase shifts $a_i$ are known.

Assume that we have collected $N$ independent samples $\mathbf{x}_n = \mathbf{x}(nT), n = 0, \ldots, N-1$, where $T$ is the sample period. Since the signals are Gaussian, all information on the parameters is located in the second-order covariance. Let the true covariance matrix $\mathbf{R}$ and its sample estimate $\hat{\mathbf{R}}$ be

$$\mathbf{R} = \mathrm{E}\{\mathbf{x}_n\mathbf{x}_n^H\}, \qquad \hat{\mathbf{R}} = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{x}_n\mathbf{x}_n^H. \tag{2}$$

Since the noise is uncorrelated to the signal, we obtain

$$\mathbf{R} = \sigma_s^2\mathbf{g}\mathbf{g}^H + \mathbf{D} \tag{3}$$

where $\mathbf{D} = \mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \mathrm{diag}[\sigma_1^2, \ldots, \sigma_p^2]$ is the noise covariance matrix, which is a diagonal matrix containing the noise powers.

Our objective in this paper will be, given $\hat{\mathbf{R}}$, to estimate $\mathbf{g}$ and $\mathbf{D}$. Since $\sigma_s^2$ is known from tables, we can make it equal to 1 without loss of generality. In addition, the $a_i$ are known from the geometry and look direction of the telescopes; hence, without loss of generality, we make $a_i = 1$ as well.

We will use the following real-valued parametrization of the model. Define a factorization of $\mathbf{g}$ into a magnitude vector $\boldsymbol{\gamma}$ and a phase vector $e^{\imath\boldsymbol{\phi}}$ as

$$\mathbf{g} = \boldsymbol{\gamma} \odot e^{\imath\boldsymbol{\phi}}.$$

It is clear that the phases in $\mathbf{g}$ are underdetermined so that we define the phase of the first entry to be zero ($\phi_1 = 0$). The parameter vector to estimate is thus

$$\boldsymbol{\theta} = \left[\boldsymbol{\gamma}^t,\ \boldsymbol{\phi}^t,\ \mathbf{d}^t\right]^t = [\gamma_1, \ldots, \gamma_p, \phi_2, \ldots, \phi_p, d_1, \ldots, d_p]^t$$

where $d_i = \sigma_i^2$.

### B. Cramér–Rao Lower Bound

The CRB gives a lower bound to the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$. In our situation, we assume that the source signal and the channel noise are complex independent Gaussian distributed with zero mean and that they satisfy the model in (3) with $\sigma_s = 1$. Following standard techniques [14], [15], the CRB is known to be given by the diagonal entries of

$$\mathbf{C}_{\mathrm{CRB}} = \frac{1}{N}\mathbf{M}^{-1}$$

where $\mathbf{M}$ is the (scaled) Fisher information matrix (FIM), which can be written as (e.g., [16])

$$\mathbf{M} = \mathbf{F}_0^H\left(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}\right)\mathbf{F}_0. \tag{4}$$

Here, $\mathbf{F}_0 := \mathbf{F}(\boldsymbol{\theta}_0)$ is the Jacobian evaluated at the true value of the parameters $\boldsymbol{\theta}_0$

$$\begin{aligned}\mathbf{F}(\boldsymbol{\theta}) &= [\mathbf{F}_\gamma,\ \mathbf{F}_\phi,\ \mathbf{F}_d](\boldsymbol{\theta})\\ \mathbf{F}_\gamma(\boldsymbol{\theta}) &= \frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \boldsymbol{\gamma}^t}(\boldsymbol{\theta}) = \left[\frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \gamma_1}, \ldots, \frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \gamma_p}\right](\boldsymbol{\theta}).\end{aligned} \tag{5}$$

$\mathbf{F}_\phi(\boldsymbol{\theta})$ and $\mathbf{F}_d(\boldsymbol{\theta})$ are expressed similarly.

Further expansion of $\mathbf{M}$, which is presented in Appendix A, shows that the gain phase parameters are decoupled from the other parameters, which hints that they can be estimated separately (indeed, we will derive such an algorithm). Moreover, $\mathbf{M}$ is independent of $\boldsymbol{\phi}$; the estimation accuracy bounds are independent of the particular values of the phases.

### C. ML Formulation

In principle, asymptotically efficient estimates of the model parameters $\boldsymbol{\theta}$, or $\mathbf{g}$ and $\mathbf{D}$, can be obtained via an ML formulation. Since all signal waveforms are i.i.d. Gaussian sequences, the derivation is standard, and ML parameter estimates for $N$ independent samples are obtained by minimizing the negative log likelihood function

$$\left\{\hat{\mathbf{g}}, \hat{\mathbf{D}}\right\} = \underset{\mathbf{g},\,\mathbf{D}}{\mathrm{argmin}}\ \ln|\mathbf{R}_\theta| + \mathrm{tr}\left(\mathbf{R}_\theta^{-1}\hat{\mathbf{R}}\right)$$

where $\mathbf{R}_\theta$ is a function of $\boldsymbol{\theta}$, and $\hat{\mathbf{R}}$ is the sample covariance matrix defined in (2). This would lead to ML optimal, asymptotically efficient, estimates.

Even if the model is somewhat simpler than in [11], it does not seem possible to solve this minimization problem in closed form, and we have to resort to numerical optimization methods, e.g., Newton–Raphson or scoring methods [15], [17], [13]. Stable implementations of such methods are complicated; global convergence is not guaranteed, and a good initial point is needed. The computational complexity is dominated by the repeated evaluation of second-order derivatives.

In the following section, we propose algorithms that are based on the least squares (LS) optimization of the model errors. Although there is no guarantee that the solution converges to the global optimum here either, the advantage of the LS algorithms is reduced computational complexity. We also derive simple closed-form approximate solutions to the LS cost function.

## III. Gain Decomposition Algorithms

We start by describing the least squares cost function, and subsequently, we derive four least squares estimation algorithms: two iterative and two in closed form.

### A. Generalized Least Squares Formulation

As discussed in [11] and following techniques written in detail in [18, chap. 9], a weighted least squares covariance matching approach is known to lead to estimates that, for a large number of samples, are equivalent to ML estimates and, hence, are asymptotically efficient and reach the CRB.

Thus, define the least squares covariance model fitting problem as, for given $\hat{\mathbf{R}}$, an estimation of $\mathbf{g}$ and $\mathbf{D}$ via

$$\left\{\hat{\mathbf{g}}, \hat{\mathbf{D}}\right\} = \underset{\mathbf{g},\,\mathbf{D}}{\operatorname{argmin}} \left\|\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D}\right\|_F^2. \tag{6}$$

Equivalently, by writing $\mathbf{g}$ and $\mathbf{D}$ in terms of their parametrization $\boldsymbol{\theta}$, we consider the minimization of the cost function $\kappa(\boldsymbol{\theta})$:

$$\kappa(\boldsymbol{\theta}) = \left\|\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D}\right\|_F^2 = \tfrac{1}{2}\,\mathbf{f}(\boldsymbol{\theta})^H\mathbf{f}(\boldsymbol{\theta}) \tag{7}$$

where $\mathbf{f}(\boldsymbol{\theta}) = \operatorname{vec}(\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D})$. A more general weighted least squares problem is obtained, for a Hermitian weighting matrix $\mathbf{W}$, as the optimization of

$$\kappa_{\mathbf{W}}(\boldsymbol{\theta}) = \tfrac{1}{2}\,\mathbf{f}(\boldsymbol{\theta})^H\mathbf{W}\mathbf{f}(\boldsymbol{\theta}) = \tfrac{1}{2}\,\mathbf{f}_{\mathbf{W}}(\boldsymbol{\theta})^H\mathbf{f}_{\mathbf{W}}(\boldsymbol{\theta}) \tag{8}$$

where $\mathbf{f}_{\mathbf{W}}(\boldsymbol{\theta}) := \mathbf{W}^{1/2}\mathbf{f}(\boldsymbol{\theta})$. The weighting can be used to obtain estimators with a reduced variance. The optimal weight is known to be the inverse of the asymptotic covariance of the residuals (cf. [11]), $\mathrm{E}\{\mathbf{f}(\boldsymbol{\theta}_0)\mathbf{f}(\boldsymbol{\theta}_0)^H\}$, where $\boldsymbol{\theta}_0$ is the "true" value of the parameters, or $\mathbf{f}(\boldsymbol{\theta}_0) = \operatorname{vec}(\hat{\mathbf{R}} - \mathbf{R})$ (corresponding to the true covariance matrix $\mathbf{R}$). Since all sources are Gaussian, we find

$$\mathbf{W}_{opt} = \left(\overline{\mathbf{R}} \otimes \mathbf{R}\right)^{-1} = \overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}. \tag{9}$$

Let $\hat{\boldsymbol{\theta}} = \operatorname{argmin} \kappa(\boldsymbol{\theta})$. Before we discuss algorithms to compute $\hat{\boldsymbol{\theta}}$, we summarize the statistical performance of this LS estimator because they follow from [18, ch. 9] and from similar results, e.g., in [11], [15], and [19]. It is well known that the (W)LS estimator is asymptotically unbiased and consistent. The asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is Gaussian, and the large sample covariance matrix $\mathbf{C} = \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^H\}$ of the parameter estimate $\hat{\boldsymbol{\theta}}$ is given by

$$\mathbf{C} = \frac{1}{N}\,\mathbf{F}_0^{\dagger}\left(\overline{\mathbf{R}} \otimes \mathbf{R}\right)\mathbf{F}_0^{\dagger H}$$

where $\mathbf{F}_0 = \mathbf{F}(\boldsymbol{\theta}_0)$ is the Jacobian in (5) evaluated at $\boldsymbol{\theta}_0$. For weighted estimates $\hat{\boldsymbol{\theta}}_{\mathbf{W}}$, the large sample covariance matrix is derived as

$$\mathbf{C}_{\mathbf{W}} = \frac{1}{N}\left(\mathbf{F}_0{}^H\mathbf{W}\mathbf{F}_0\right)^{-1}\mathbf{F}_0^H\mathbf{W}\left(\overline{\mathbf{R}} \otimes \mathbf{R}\right)\mathbf{W}\mathbf{F}_0 \cdot\left(\mathbf{F}_0^H\mathbf{W}\mathbf{F}_0\right)^{-1}.$$

At the optimal weight (9)

$$\mathbf{C}_{\mathbf{W}_{opt}} = \frac{1}{N}\left(\mathbf{F}_0^H\left(\overline{\mathbf{R}} \otimes \mathbf{R}\right)^{-1}\mathbf{F}_0\right)^{-1} = \mathbf{C}_{\mathrm{CRB}}.$$

Hence, at the optimal weight (or a consistent estimate of it), the WLS estimator is asymptotically efficient.

### B. Gauss–Newton Iterations (GNLS)

Assuming that we have a good initial point for $\boldsymbol{\theta}$, the minimization of the (weighted) LS cost function (6) can be carried out using the Gauss–Newton method [17].

Let $\mathbf{F}(\boldsymbol{\theta})$ denote the Jacobian (5). For the unweighted cost function, the gradient at $\boldsymbol{\theta}$ is

$$\mathbf{p}(\boldsymbol{\theta}) = \operatorname{Re}\left(\mathbf{F}(\boldsymbol{\theta})^H\mathbf{f}(\boldsymbol{\theta})\right) = \mathbf{F}(\boldsymbol{\theta})^H\mathbf{f}(\boldsymbol{\theta}). \tag{10}$$

Note that due to the Hermitian symmetries, the product $\mathbf{F}^H\mathbf{f}$ is already real. The Hessian of the cost function at $\boldsymbol{\theta}$ is given by

$$\mathbf{H}(\boldsymbol{\theta}) = \operatorname{Re}\left(\mathbf{F}^H(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})\right) = \mathbf{F}(\boldsymbol{\theta})^H\mathbf{F}(\boldsymbol{\theta}). \tag{11}$$

The Gauss–Newton update step is then

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \mu_k\mathbf{H}(\boldsymbol{\theta}_k)^{-1}\mathbf{p}(\boldsymbol{\theta}_k) = \boldsymbol{\theta}_k - \mu_k\mathbf{F}_k^{\dagger}\mathbf{f}_k \tag{12}$$

where $\mathbf{F}_k = \mathbf{F}(\boldsymbol{\theta}_k)$, and $\mathbf{f}_k = \mathbf{f}(\boldsymbol{\theta}_k)$. $\mu_k \in (0, 1]$ is a step size; with a good initial point, we can take $\mu_k = 1$, but in practice, a line search would be necessary to ensure proper convergence. For numerical stability, the pseudo-inverse of $\mathbf{F}_k$ can be regularized by incorporating a certain threshold on its singular values. A similar iteration can be derived for the weighted cost function.

An initial point for the GNLS recursion is needed, and can often be obtained from an SVD of $\hat{\mathbf{R}} - \operatorname{diag}(\hat{\mathbf{R}})$, because the astronomical source power is usually much smaller than the noise powers (or $\operatorname{diag}(\mathbf{g}\mathbf{g}^H) \ll \mathbf{D}$). Another possibility is applying one of the closed-form algorithms described later in Sections III-D and -E. The initial point can also be used to generate a consistent estimate of the optimal weight (since the latter depends on the true $\mathbf{R}$ and is unknown).

In the CRB derivations, it was shown that the gain magnitudes and the gain phases are decoupled. Similar to the derivation in the Appendix, we can show that $\mathbf{F}_{\gamma}^H\mathbf{F}_{\phi} = 0$, and $\mathbf{F}_d^H\mathbf{F}_{\phi} = 0$, which somewhat simplifies the Hessian. As in [11], it is also possible to concentrate the cost function, eliminating $\mathbf{D}$ and the scaling of $\mathbf{g}$, but the remaining parameters are more strongly coupled and the derivatives more complex to evaluate. In our experience, the complexity of the resulting Gauss–Newton scheme is higher, and the convergence is not better.

### C. Minimization Using Alternating Least Squares (ALS)

*Unweighted ALS Algorithm:* A straightforward technique to try to optimize a cost function over many parameters is to alternatingly minimize over a subset, keeping the remaining parameters fixed. In our case, assume at the $k$th iteration that we have an estimate $\hat{\mathbf{D}}_k$. The next step is to minimize the LS cost function (6) with respect to the gain vector only.

$$\hat{\mathbf{g}}_k = \underset{\mathbf{g}}{\operatorname{argmin}} \left\|\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \hat{\mathbf{D}}_k\right\|_F^2. \tag{13}$$

The minimum is found from the eigenvalue decomposition $\hat{\mathbf{R}} - \hat{\mathbf{D}}_k = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where the matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$ contains the eigenvectors $\mathbf{u}_i$, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_i$. The gain estimate minimizing (13) is given by

$$\hat{\mathbf{g}}_k = \mathbf{u}_1\sqrt{\lambda_1} \tag{14}$$

where $\lambda_1$ is the largest eigenvalue, and $\mathbf{u}_1$ is the corresponding eigenvector. The second step is minimizing (6) with respect to the system noise matrix $\mathbf{D}$, keeping the gain vector fixed

$$\hat{\mathbf{D}}_{k+1} = \underset{\mathbf{D}}{\operatorname{argmin}} \left\| \hat{\mathbf{R}} - \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H - \mathbf{D} \right\|^2 \tag{15}$$

where $\mathbf{D}$ is constrained to be diagonal with non-negative entries. The minimum is obtained by subtracting $\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H$ from $\hat{\mathbf{R}}$ and discarding all off-diagonal elements

$$\hat{\mathbf{D}}_{k+1} = \operatorname{diag}\left( \hat{\mathbf{R}} - \mathbf{g}_k \mathbf{g}_k^H \right).$$

The condition that the diagonal elements of $\hat{\mathbf{D}}_{k+1}$ should be positive can be implemented by subsequently setting the negative entries at zero. The two minimizations steps (13) and (15) are repeated until the model error (7) converges. Since each of the minimizing steps in the iteration loop reduces the model error, we obtain monotonic convergence to a local minimum. Although the iteration is very simple to implement, simulations indicate that convergence is usually very slow, especially in the absence of a reasonable initial point.

*Weighted ALS:* The optimal weight for the LS cost function is $\mathbf{W} = \overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}$. Due to this Kronecker structure, the WLS cost function (8) can also be written as

$$\left\{ \hat{\mathbf{g}}, \hat{\mathbf{D}} \right\} = \underset{\mathbf{g},\, \mathbf{D}}{\operatorname{argmin}} \left\| \mathbf{W}_c \left( \hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D} \right) \mathbf{W}_c \right\|_F^2 \tag{16}$$

where $\mathbf{W}_c = \mathbf{R}^{-1/2}$. As before, if at the $k$th iteration we have an estimate $\hat{\mathbf{D}}_k$ of $\mathbf{D}$, then an estimate of $\mathbf{g}$ follows from

$$\hat{\mathbf{g}}_k = \underset{\mathbf{g}}{\operatorname{argmin}} \left\| \mathbf{W}_c \left( \hat{\mathbf{R}} - \hat{\mathbf{D}}_k \right) \mathbf{W}_c - (\mathbf{W}_c \mathbf{g})(\mathbf{W}_c \mathbf{g})^H \right\|_F^2. \tag{17}$$

After computing the eigenvalue decomposition $\mathbf{W}_c(\hat{\mathbf{R}} - \hat{\mathbf{D}}_k)\mathbf{W}_c = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, the estimate $\hat{\mathbf{g}}_k$ follows as

$$\hat{\mathbf{g}}_k = \mathbf{W}_c^{-1} \mathbf{u}_1 \sqrt{\lambda_1}. \tag{18}$$

The second step is minimizing the cost function with respect to $\mathbf{D}$ while keeping $\mathbf{g}$ fixed:

$$\hat{\mathbf{D}}_{k+1} = \underset{\mathbf{D}}{\operatorname{argmin}} \left\| \mathbf{W}_c \left( \hat{\mathbf{R}} - \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H \right) \mathbf{W}_c - \mathbf{W}_c \mathbf{D} \mathbf{W}_c \right\|^2. \tag{19}$$

Letting $\mathbf{d} = \operatorname{vecdiag}(\mathbf{D})$ (using several Kronecker relations [20])

$$\begin{aligned} \hat{\mathbf{d}}_{k+1} &= \underset{\mathbf{d}}{\operatorname{argmin}} \left\| \left( \overline{\mathbf{W}}_c \otimes \mathbf{W}_c \right) \operatorname{vec}\left( \hat{\mathbf{R}} - \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H \right) - \left( \overline{\mathbf{W}}_c \circ \mathbf{W}_c \right) \mathbf{d} \right\|^2 \\ &= \left( \overline{\mathbf{W}}_c^{\,2} \odot \mathbf{W}_c^2 \right)^{-1} \operatorname{vecdiag}\left( \mathbf{W}_c^2 \left( \hat{\mathbf{R}} - \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H \right) \mathbf{W}_c^2 \right). \end{aligned}$$

This is a closed-form solution for $\hat{\mathbf{D}}_{k+1}$. Note that unless $\mathbf{W}_c$ is diagonal, in general, the result is not equal to $\operatorname{diag}(\hat{\mathbf{R}} - \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H)$.

The optimal weight $\mathbf{W}_c = \mathbf{R}^{-1/2}$ depends on the true covariance matrix $\mathbf{R}$, which is unknown. Asymptotically, the same results are obtained by replacing $\mathbf{R}$ by a consistent estimate, for example, $\hat{\mathbf{R}}$ or the result of one of the closed-form estimates in the following sections. At the optimum, the statistical properties of the ALS and W-ALS estimators are as described in Section III-A.

### *D. Closed Form Using Logarithmic Least Squares (LOGLS)*

*Unweighted LOGLS Algorithm:* An alternative closed form estimate (as in use at the WSRT [3] since 1980) is obtained by minimizing the mean squared error of the *logarithms* of the model. As we will show, taking the logarithm has several effects. The equations become linear in the parameters as the products of gains become sums:

$$\ln(r_{ij}) = \ln(\gamma_i) + \ln(\gamma_j) + \imath(\phi_i - \phi_j) \bmod 2\pi\imath \qquad (i \neq j).$$

It is seen that a least squares model fitting can be applied to the gain magnitude and gain phase separately. Unfortunately, a modulo $2\pi$ phase ambiguity is introduced because of the complex properties of the logarithm. This makes phase unwrapping necessary in the decomposition algorithm.

In a matrix formulation, we minimize a LS cost function after taking the element-wise logarithm

$$\left\{ \hat{\mathbf{g}}, \hat{\mathbf{D}} \right\} = \underset{\mathbf{g},\, \mathbf{D},\, \mathbf{k}}{\operatorname{argmin}} \left\| \ln\left( \hat{\mathbf{R}} \right) - \ln\left( \mathbf{g}\mathbf{g}^H + \mathbf{D} \right) + \mathbf{K} 2\pi\imath \right\|_F^2 \tag{20}$$

where $\mathbf{K}$ is a nuisance parameter matrix of integers. For any estimate $\hat{\mathbf{g}}$, the optimal estimate of $\hat{\mathbf{D}}$ is still given by $\hat{\mathbf{D}} = \operatorname{diag}(\hat{\mathbf{R}} - \hat{\mathbf{g}}\hat{\mathbf{g}}^H)$. Substituting this back into the cost function shows that the main diagonal of the argument to the Frobenius norm is equal to zero. Thus, the cost function is compressed as

$$\hat{\mathbf{g}} = \underset{\mathbf{g},\, \mathbf{k}}{\operatorname{argmin}} \left\| \mathbf{J} \left[ \operatorname{vec} \ln\left( \hat{\mathbf{R}} \right) - \ln(\overline{\mathbf{g}} \otimes \mathbf{g}) + \mathbf{k} 2\pi\imath \right] \right\|^2 \tag{21}$$

where $\mathbf{k} = \operatorname{vec}(\mathbf{K})$. Note that

$$\begin{aligned} \ln(\overline{\mathbf{g}} \otimes \mathbf{g}) &= \mathbf{1} \otimes \ln(\mathbf{g}) + \ln(\overline{\mathbf{g}}) \otimes \mathbf{1} \\ &= \mathbf{1} \otimes \ln(\boldsymbol{\gamma}) + \ln(\boldsymbol{\gamma}) \otimes \mathbf{1} + \imath(\mathbf{1} \otimes \boldsymbol{\phi} - \boldsymbol{\phi} \otimes \mathbf{1}) \\ &= [\mathbf{1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{1}] \ln(\boldsymbol{\gamma}) + \imath[\mathbf{1} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{1}]\boldsymbol{\phi}. \end{aligned} \tag{22}$$

Defining $\mathbf{v}_R = \operatorname{vec}(\operatorname{Re} \ln(\hat{\mathbf{R}}))$ and $\mathbf{v}_I = \operatorname{vec}(\operatorname{Im} \ln(\hat{\mathbf{R}}))$, it is seen that the optimization problem separates into independent optimizations over $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ (corresponding to the real and imaginary components), namely

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \, \| \mathbf{J}\mathbf{v}_R - \mathbf{J}[\mathbf{1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{1}] \ln(\boldsymbol{\gamma}) \|^2 \\ \hat{\boldsymbol{\phi}} &= \underset{\boldsymbol{\phi},\, \mathbf{k};\phi_1 = 0}{\operatorname{argmin}} \, \| \mathbf{J}(\mathbf{v}_I + \mathbf{k} 2\pi) - \mathbf{J}[\mathbf{1} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{1}]\boldsymbol{\phi} \|^2. \end{aligned}$$

Each cost function is linear and easily optimized in closed form. For the first cost function, we obtain, using $\mathbf{J}^H\mathbf{J} = \mathbf{I} \otimes \mathbf{I} - (\mathbf{I} \circ \mathbf{I})(\mathbf{I} \circ \mathbf{I})^H$ and several Kronecker relations [20]

$$\ln(\hat{\boldsymbol{\gamma}}) = (\mathbf{J}[\mathbf{1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{1}])^{\dagger} \mathbf{J}\mathbf{v}_R = \mathbf{B}^{-1}\mathbf{q} \tag{23}$$

where $\mathbf{B} = 2((p-2)\mathbf{I} + \mathbf{1}\mathbf{1}^t)$, $\mathbf{q} = 2(\operatorname{Re} \ln(\hat{\mathbf{R}}) \odot \mathbf{I}_c)\mathbf{1}$, and $\mathbf{I}_c = \mathbf{1}\mathbf{1}^t - \mathbf{I}$. An explicit expression of $\mathbf{B}^{-1}$ is easily obtained

using Woodbury's identity. Note also that $\mathbf{B}^{-1}\mathbf{q}$ can be computed very efficiently. Similarly, solving the second cost function is reduced to solving

$$\mathbf{C}_e\hat{\boldsymbol{\phi}} = \mathbf{v}_e + 2\pi\mathbf{k}_e$$

where

$$\mathbf{C}_e = \begin{bmatrix}\mathbf{C}\\ \mathbf{e}_1^t\end{bmatrix}, \quad \mathbf{v}_e = \begin{bmatrix}\mathbf{v}_I\\ \phi_1\end{bmatrix}, \quad \mathbf{k}_e = \begin{bmatrix}\mathbf{k}\\ 0\end{bmatrix}$$

and $\mathbf{C} = \mathbf{1}\otimes\mathbf{I} - \mathbf{I}\otimes\mathbf{1}$. The last row in $\mathbf{C}_e$ implements the chosen phase uniqueness constraint $\phi_1 = 0$. The system can be solved as

$$\hat{\boldsymbol{\phi}} = \left(\mathbf{C}_e^H\mathbf{C}_e\right)^{-1}\mathbf{C}_e^H(\mathbf{v}_e + 2\pi\mathbf{k}_e) \tag{24}$$

once the integers $\mathbf{k}_e$ are known. Suppose we have an initial estimate $\boldsymbol{\phi}^\circ$ of $\boldsymbol{\phi}$, for example, from the first column $\hat{\mathbf{r}}_1$ of $\hat{\mathbf{R}}$ as $\boldsymbol{\phi}^\circ = \operatorname{Im}\ln(\hat{\mathbf{r}}_1)$. Then, $\mathbf{k}$ (and thus $\mathbf{k}_e$) is obtained by rounding the entries of $\mathbf{C}\boldsymbol{\phi}^\circ - \mathbf{v}_I$ to the nearest multiple of $2\pi$: $\mathbf{k} = \operatorname{round}((2\pi)^{-1}(\mathbf{C}\boldsymbol{\phi}^\circ - \mathbf{v}_I))$. Using Woodbury's identity and several Kronecker relations, (24) can be worked out further, producing

$$\left(\mathbf{C}_e^H\mathbf{C}_e\right)^{-1} = \frac{1}{2}\left(\mathbf{E} + \frac{(\mathbf{E}\mathbf{1})(\mathbf{E}\mathbf{1})^H}{1-\operatorname{tr}(\mathbf{E})}\right)$$

$$\mathbf{C}_e^H(\mathbf{v}_e + 2\pi\mathbf{k}_e) = 2\left[\left(\operatorname{Im}\ln\left(\hat{\mathbf{R}}\right) + 2\pi\mathbf{K}\right)\odot\mathbf{I}_c\right]\mathbf{1}$$

where $\mathbf{E} = (1/2)(2p\mathbf{I} + \mathbf{e}_1\mathbf{e}_1^t)$, and $\mathbf{K} = \operatorname{unvec}(\mathbf{k})$.

*Weighted LOGLS:* The LOGLS method has zero bias and is consistent, as the logarithm function is a smooth monotonous transformation, but the logarithm operator prevents the LOGLS method from being statistically efficient. In general, there exists no weighting matrix that makes the LOGLS method asymptotically efficient, but there exist special cases for which asymptotical efficiency can be reached. This will be shown by comparing the LOGLS cost function with the weighted LS cost function, the latter leading to asymptotically efficient estimates. We start from the compressed LOGLS cost function (21). After replacing the selection matrix $\mathbf{J}$ by $\mathbf{J}^H\mathbf{J}$ (which does not change the norm and is more convenient since $\mathbf{J}^H\mathbf{J}$ is diagonal), we introduce a weighting matrix $\mathbf{W}_{\log}$ to obtain

$$\hat{\mathbf{g}} = \underset{\mathbf{g},\,\mathbf{k}}{\operatorname{argmin}}\left\|\mathbf{W}_{\log}^{1/2}\mathbf{J}^H\mathbf{J}\operatorname{vec}\left[\ln\left(\hat{\mathbf{R}}\right) - \ln\left(\mathbf{g}\mathbf{g}^H\right) + \mathbf{k}2\pi\imath\right]\right\|^2. \tag{25}$$

Assuming that $\hat{\mathbf{g}}$ is close to $\mathbf{g}$ and that none of the gain amplitudes $\gamma$ are zero, we can use the Taylor approximation $\ln(x)\approx x-1$. Let $\ominus$ denote an element-wise matrix division; then

$$\begin{aligned}
&\mathbf{J}^H\mathbf{J}\operatorname{vec}\left[\ln\left(\hat{\mathbf{R}}\right) - \ln\left(\mathbf{g}\mathbf{g}^H\right)\right]\\
&\quad= \mathbf{J}^H\mathbf{J}\operatorname{vec}\left[\ln\left(\hat{\mathbf{R}}\ominus\left(\mathbf{g}\mathbf{g}^H\right)\right)\right]\\
&\quad\approx \mathbf{J}^H\mathbf{J}\operatorname{vec}\left[\left(\hat{\mathbf{R}}\ominus\left(\mathbf{g}\mathbf{g}^H\right) - \mathbf{I}\right)\right]\\
&\quad= \mathbf{J}^H\mathbf{J}\operatorname{vec}\left[(\operatorname{diag}(\mathbf{g}))^{-1}\left(\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H\right)(\operatorname{diag}(\overline{\mathbf{g}}))^{-1}\right]\\
&\quad= \mathbf{J}^H\mathbf{J}, (\operatorname{diag}(\overline{\mathbf{g}}\otimes\mathbf{g}))^{-1}\operatorname{vec}\left(\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H\right).
\end{aligned}$$

Note that for the cost function, the phase of the diagonal matrix $\operatorname{diag}(\overline{\mathbf{g}}\otimes\mathbf{g})^{-1}$ is irrelevant; therefore, it can be replaced by $(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1}$, where $\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma})$. The weighted LOGLS cost function close to the optimum can thus be written as

$$\min_{\mathbf{g}}\left\|\mathbf{W}_{\log}^{1/2}\mathbf{J}^H\mathbf{J}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1}\operatorname{vec}\left(\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H\right)\right\|^2.$$

If we compare this to the cost function for the weighted LS, $\min_{\mathbf{g}}\|\mathbf{W}_{ls}^{1/2}\mathbf{J}^H\mathbf{J}\operatorname{vec}(\hat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H)\|^2$, we see that these two cost functions give the same solution if

$$\mathbf{W}_{\log}^{1/2}\mathbf{J}^H\mathbf{J}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1} = \mathbf{W}_{ls}^{1/2}\mathbf{J}^H\mathbf{J}. \tag{26}$$

Because $\mathbf{J}^H\mathbf{J}$ is singular, a solution is possible only in special cases, for example, diagonal weighting matrices. For low SNRs, $\text{SNR} = \mathbf{g}^H\mathbf{g}/\operatorname{tr}(\mathbf{D}) \ll 1$, which is true for most radio astronomical observations, the $\mathbf{W}_{ls}$ weighting matrix is close to diagonal $\mathbf{W}_{ls} = \mathbf{D}^{-1}\otimes\mathbf{D}^{-1}$, and in this case, the choice $\mathbf{W}_{\log}^{1/2} = (\mathbf{D}^{-(1/2)}\otimes\mathbf{D}^{-(1/2)})(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}) = \mathbf{D}^{-(1/2)}\boldsymbol{\Gamma}\otimes\mathbf{D}^{-(1/2)}\boldsymbol{\Gamma}$ satisfies (26). Here, we used the diagonal and commutative properties of $\mathbf{J}^H\mathbf{J}$. Defining $\mathbf{W} = \mathbf{D}^{-(1/2)}\boldsymbol{\Gamma}$, the weighted LOGLS cost function can be expressed as

$$\min_{\mathbf{g},\mathbf{k}}\left\|\mathbf{J}(\mathbf{W}\otimes\mathbf{W})\left[\operatorname{vec}\ln\left(\hat{\mathbf{R}}\right) - \ln(\overline{\mathbf{g}}\otimes\mathbf{g}) + 2\pi\mathbf{k}\imath\right]\right\|^2.$$

Further defining $\mathbf{v}_R' = \operatorname{vec}(\operatorname{Re}\mathbf{W}\ln(\hat{\mathbf{R}})\mathbf{W})$, $\mathbf{v}_I' = \operatorname{vec}(\operatorname{Im}\mathbf{W}\ln(\hat{\mathbf{R}})\mathbf{W})$, and $\mathbf{k}' = (\mathbf{W}\otimes\mathbf{W})\mathbf{k}$, and noting that [viz. (22)]

$$\begin{aligned}
&(\mathbf{W}\otimes\mathbf{W})\ln(\overline{\mathbf{g}}\otimes\mathbf{g})\\
&\qquad= [\mathbf{w}\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{w}]\mathbf{W}\ln\gamma + \imath[\mathbf{w}\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{w}]\mathbf{W}\boldsymbol{\phi}
\end{aligned}$$

where $\mathbf{w} = \mathbf{W}\mathbf{1}$, it follows that we need to solve

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}}\left\|\mathbf{J}\mathbf{v}_R' - \mathbf{J}[\mathbf{w}\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{w}]\mathbf{W}\ln(\boldsymbol{\gamma})\right\|^2$$

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi},\,\mathbf{k};\phi_1=0}{\operatorname{argmin}}\left\|\mathbf{J}(\mathbf{v}_I' + \mathbf{k}'2\pi) - \mathbf{J}[\mathbf{w}\otimes\mathbf{I} - \mathbf{I}\otimes\mathbf{w}]\mathbf{W}\boldsymbol{\phi}\right\|^2.$$

It is clear that this can be done in closed form as before; we omit the details.

The conclusion is that for low SNRs, we can weight by $\mathbf{W}_{\log}$ and make the weighted LOGLS method asymptotically efficient to a good approximation.

### *E. Closed Form Using Column Ratios (COLR)*

*Unweighted COLR Algorithm:* As a last technique, we now set out to find a closed-form estimate of $\mathbf{g}$, which recovers $\mathbf{g}$ exactly when applied to $\mathbf{R}$ (hence asymptotically for $\hat{\mathbf{R}}$). The crux of this method is the observation that the off-diagonal entries of $\mathbf{g}\mathbf{g}^H$ are equal to those of $\mathbf{R}$ and known, so that we only need to reconstruct the diagonal entries of $\mathbf{g}\mathbf{g}^H$. We further note that $\mathbf{g}\mathbf{g}^H$ is rank 1; therefore, any submatrix of $\mathbf{R}$ that does not contain elements from the main diagonal is also rank 1. This property can be used to estimate the ratio between any pair of columns of $\mathbf{R}$ away from the diagonal and, subsequently, to estimate how the main diagonal of $\mathbf{R}$ has to be changed so that the resulting $\mathbf{R}'$ is rank 1 or $\mathbf{R}' = \mathbf{g}\mathbf{g}^H$. The gain vector $\mathbf{g}$ can then be extracted by an eigenvalue decomposition.

For any two elements $g_i$ and $g_j$ of the complex gain vector $\mathbf{g}$, define the ratio $\alpha_{ij} = g_i/g_j$. This ratio can be estimated from the data $\mathbf{R}$ by solving

$$\mathbf{c}_i = \alpha_{ij}\mathbf{c}_j$$

where $\mathbf{c}_i$ and $\mathbf{c}_j$ are the $i$th and $j$th column of the matrix $\mathbf{R}$, not including the entries $r_{ii}$, $r_{ij}$, $r_{ji}$, and $r_{jj}$ because $r_{ii}$ and $r_{jj}$ also depend on the unknown system noise $d_i$. Solving for $\alpha_{ij}$ in the least squares sense gives

$$\hat{\alpha}_{ij} = \left(\mathbf{c}_j^H\mathbf{c}_j\right)^{-1}\mathbf{c}_j^H\mathbf{c}_i = \frac{\sum_{k\neq i,j} r_{kj}^* r_{ki}}{\sum_{k\neq i,j} r_{kj}^* r_{kj}}.$$

We can subsequently estimate $|g_i|^2$ as $|g_i|^2 = \hat{\alpha}_{ij}r_{ij}$ for any choice of $j \neq i$. This estimate can be improved if all $(p-1)$ available column ratios are used, and the fact that $|g_i|^2$ is real

$$|g_i|^2 = \frac{1}{p-1}\,\mathrm{Re}\left(\sum_{\substack{j=1\\ j\neq i}}^{p} \hat{\alpha}_{ij}r_{ij}\right).$$

The next step is to form $\mathbf{R}'$ equal to $\mathbf{R}$ but with the diagonal entries replaced by the estimates of $|g_i|^2$ obtained above. The resulting matrix $\mathbf{R}'$ is an estimate of $\mathbf{g}\mathbf{g}^H$, and $\hat{\mathbf{g}}$ is found from the eigenvalue decomposition $\mathbf{R}' = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, similarly as in (14). $\mathbf{D}$ is obtained as $\mathbf{D} = \mathrm{diag}(\mathbf{R} - \mathbf{g}\mathbf{g}^H)$.

With measured data, we follow the same procedure but replace $\mathbf{R}$ by the sample estimate $\hat{\mathbf{R}}$. Although there is no claim that this procedure minimizes the LS cost function, its estimates are rather close and asymptotically (large $N$), the true $\mathbf{g}$ and $\mathbf{D}$ are obtained. In the simulations and in the experimental results, the column ratio method will be denoted by the acronym "COLR."

*Weighted COLR:* Because of the *ad hoc* derivation of the algorithm, it is difficult to establish the statistical efficiency and the weighting matrix that might achieve this. The analysis of the other three gain estimation methods suggests that a weighting matrix of the form $\mathbf{W}_{col}^{1/2} = (\mathbf{D}\otimes\mathbf{D})^{-(1/2)}$ might improve the statistical efficiency in some cases (namely, very nonuniform noise power). In the section on simulations, we show that this is indeed the case. Since $\mathbf{D}$ is initially unknown, we can construct the weighting matrix only after a first (unweighted) estimate of the parameters has been made. The COLR algorithm is subsequently applied to $\mathbf{D}^{-(1/2)}\hat{\mathbf{R}}\mathbf{D}^{-(1/2)}$. The resulting gain estimate $\hat{\mathbf{g}}_w$ is converted to the unweighted gain by $\hat{\mathbf{g}} = \mathbf{D}^{1/2}\hat{\mathbf{g}}_w$. The diagonal estimate $\hat{\mathbf{D}}$ is obtained as before as $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\mathbf{R}} - \hat{\mathbf{g}}\hat{\mathbf{g}}^H)$.

### F. Computational Complexity

When a fast update rate for the parameter estimation is needed or when a very large number of telescopes/antenna sensors is used, the computational complexity of the algorithms become important. This is especially the case for future generations of large radio telescopes, such as SKA and LOFAR, where the number of telescopes will be very large ($p \geq 100$). Table I lists the order of the number of multiplicative operations required for each of the algorithms and the additional number of multiplications for computing weighted estimates. The computations for determining the initial points for the GNLS and the ALS methods are not taken into account.

TABLE I
NUMBER OF MULTIPLICATIONS. ($p$ IS THE NUMBER OF SENSORS, $N_{it}$ IS THE NUMBER OF ITERATIONS)

| Method | unweighted | weighted (add'l) |
|---|---|---|
| GNLS | $34\,N_{it}p^3$ | $40\,N_{it}p^3$ |
| ALS | $8\,N_{it}p^3$ | $16\,N_{it}p^3$ |
| COLR | $20\,p^3$ | $2\,p^2$ |
| LOGLS | $2\,p^2$ | $16\,p^2$ |

It is seen that the GNLS method is the most complex, whereas the LOGLS method is computationally the fastest. For the GNLS method, the most demanding operation is the calculation of the inverse of the Hessian; for the ALS and the COLR methods, it is the repeated evaluation of the SVD. (In the table, we did not take into account the fact that faster estimators for the dominant singular vector exist.) The LOGLS algorithm only requires a few matrix–vector multiplications as the LOGLS estimators have simple closed-form expressions, and it is therefore the fastest method.

## IV. SIMULATIONS

First, we verify the performance of the estimation algorithms by computer simulations and then show results based on experimental data collected at the WSRT.

In all simulations, we use $p = 8$ telescope channels. Unless denoted otherwise, the gain magnitude $\gamma_i$ $(i = 1, \ldots, p)$ is kept fixed for each simulation. The $\gamma_i$ are chosen randomly in the interval $\gamma_0 - \Delta\gamma$ and $\gamma_0 + \Delta\gamma$, where $\gamma_0 = 1$ is the nominal gain magnitude, and $\Delta\gamma$ is a spreading parameter. The gain phase $\phi_i$ lies uniformly distributed between 0 and $\Delta\phi$, where $\Delta\phi \leq 2\pi$. The gain phase is randomly chosen and is also kept fixed during most simulations. The astronomical source is a zero mean complex Gaussian signal with unit power. The SNR is defined as the total received power due to the source, divided by the total noise power, or $\mathrm{SNR} = \mathbf{g}^H\mathbf{g}/\mathrm{tr}(\mathbf{D})$. Finally, the system noise magnitude $d_i$ lies uniformly distributed between $\mathrm{SNR}^{-1}(1-\Delta d)$ and $\mathrm{SNR}^{-1}(1+\Delta d)$, where $\Delta d$ is the noise spreading parameter $\Delta d \leq 1$.

For a typical online gain calibration measurement at a radio observatory, astronomical sources are used with SNRs in the range $-20$ to $-10$ dB, the integration time of the correlation data usually is several seconds to a few minutes for a typical frequency bin resolution of 10 to 100 kHz. Therefore, for most of the simulations, the following parameter settings are chosen: $\Delta\gamma = 0.1$, $\Delta d = 0.1$, $\Delta\phi = 2\pi$, and $N = 2^{15}$.

In the simulations, $N$ is the number of complex samples on which the covariance matrix is based and to which the gain decomposition algorithms are applied. Finally, $N_{sim}$ is the number of simulation runs from which the estimation standard deviation is derived. Choosing $N_{sim} = 256$ gives a reasonable standard deviation accuracy.

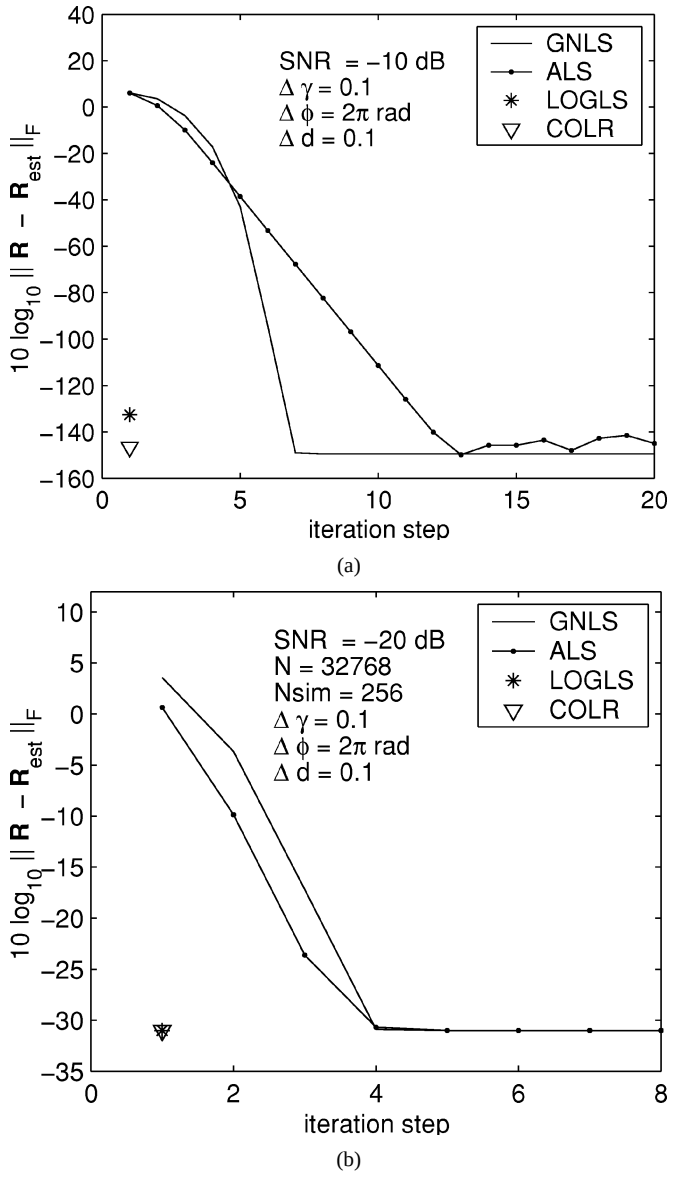


Fig. 2. Convergence speed of the cost function, weighted gain estimation algorithms (a) applied to the true $\mathbf{R}$ and (b) applied to finite sample covariance matrices.

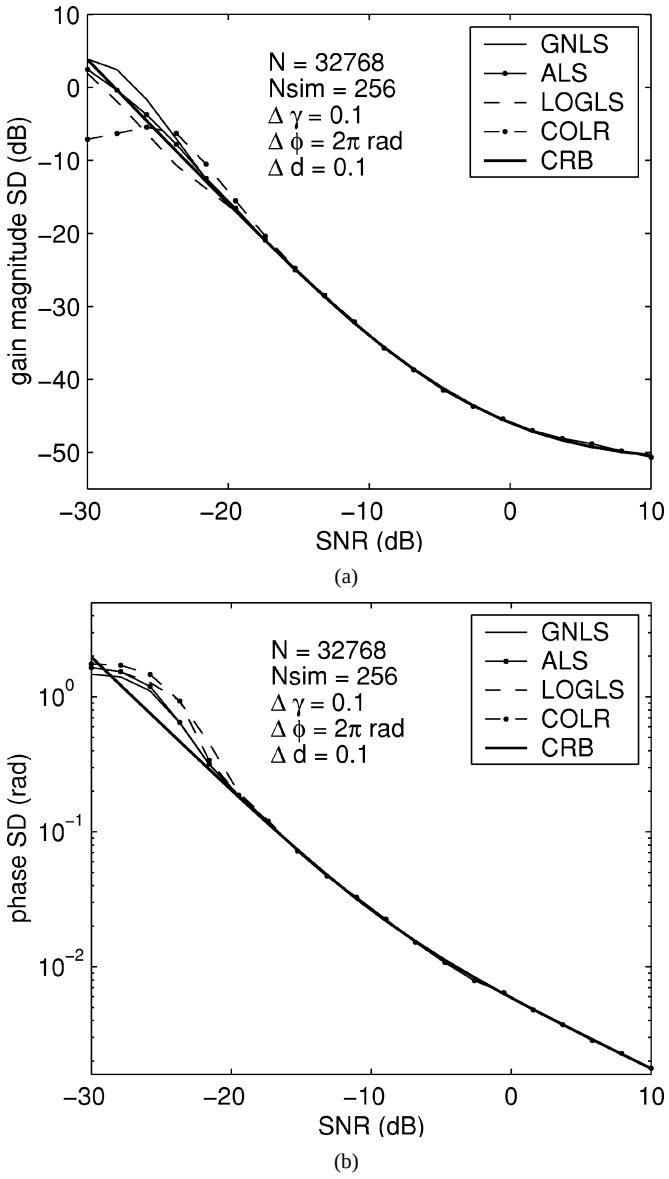


Fig. 3. Weighted gain estimate standard deviation versus SNR. (a) Gain magnitude SD. (b) Gain phase SD.

### A. Convergence of GNLS and ALS

Fig. 2(a) shows the convergence speed (Frobenius norm of the covariance matrix) of the weighted algorithms when applied to a true covariance matrix $\mathbf{R}$ (infinite sample case). The ALS method shows a linear decrease of the error; the Gauss–Newton curve shows a quadratic decrease in the estimation error, as is expected and well known from literature. The closed-form estimators produce the correct values in one step, up to computer accuracy. Fig. 2(b) shows the convergence speed for covariance matrices with noise. In this case, the GNLS method does not show quadratic convergence because after a few iteration steps, the estimation accuracy is dominated by the noise (the CRB is reached within a few iteration steps).

For the chosen SNRs and spread in gain and noise power, applying weights improves the convergence speed of the ALS method, which is reduced by a factor two. For the GNLS method, weighting does not have a significant effect on the convergence speed for the parameter settings used.

For the ALS and the GNLS methods, an initial point that is relatively close to the true values is needed. The results of the noniterative methods could be used as initial point. In the simulations, however, the true gain and noise values with small perturbations were used as initial points.

### B. Influence of SNR and Number of Samples

Fig. 3 shows the results of a gain estimation simulation in which the gain estimation standard deviation is plotted versus SNR for $N = 32\,768$ samples. The four estimator results are plotted together with the statistical efficiency bound (CRB). For the parameter range under consideration, all methods perform close to the bound, except for SNRs $< -15$ dB, in which case, the gain estimators are biased, as is shown in Fig. 4. The figures show that the LOGLS and COLR methods break down at slightly higher SNRs than the ALS and GNLS methods.

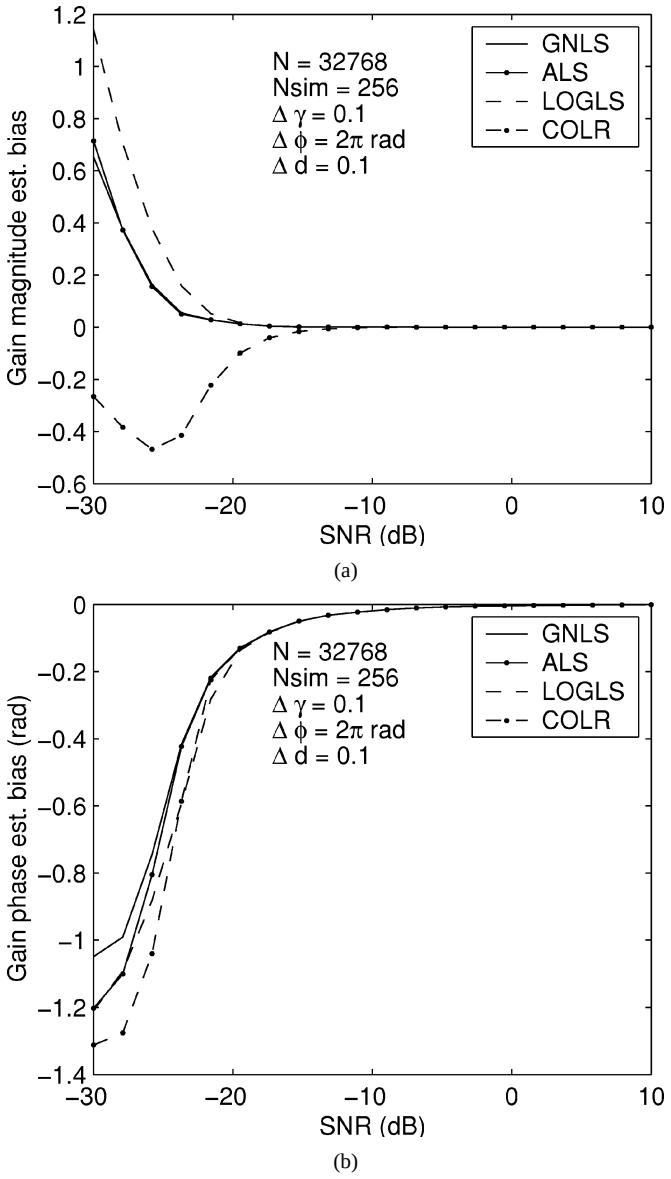


Fig. 4. Weighted gain estimate bias versus SNR. (a) Gain magnitude bias. (b) Gain phase bias.

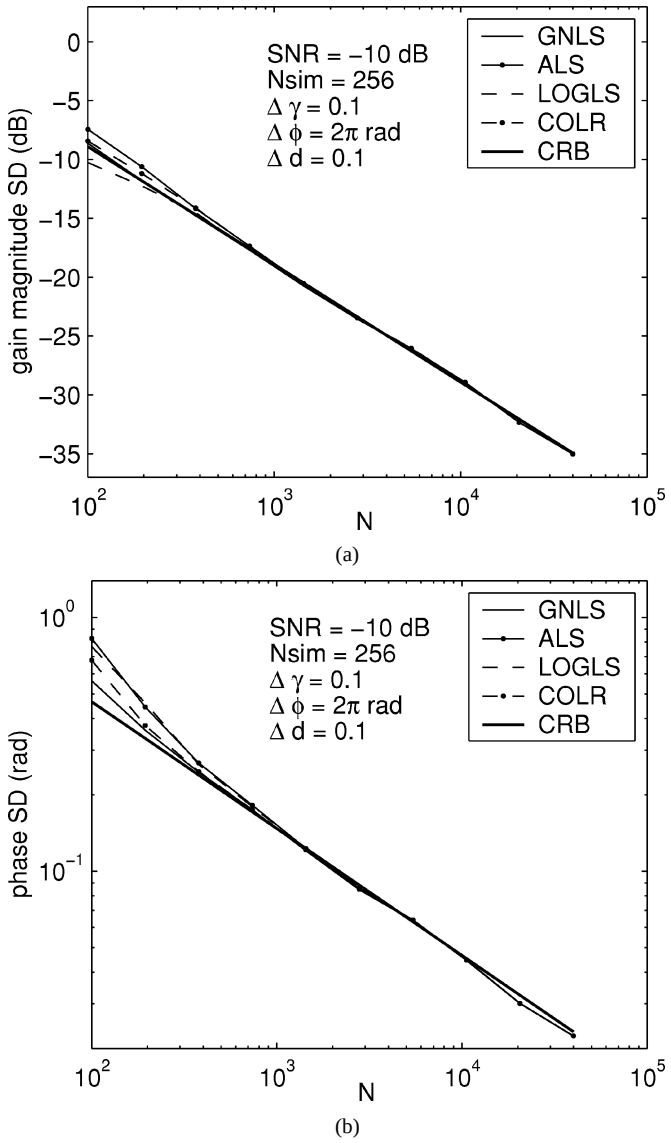


Fig. 5. Weighted gain estimate standard deviation versus number of samples. (a) Gain magnitude SD. (b) Gain phase SD.

For high SNR values, the variance of the gain estimates saturates toward a fixed value, determined by the number of samples. Indeed, at very large SNR, the direction of $\mathbf{g}$ can be estimated accurately even with a single sample, but its scaling $\|\mathbf{g}\|$, or the source power estimate, suffers from the finite sample effect.

In Fig. 5, the weighted gain estimates are plotted as a function of the number of observed time samples for an SNR of $-10$ dB. The performance of the four weighted methods do not differ much for the parameter setting used.

### C. Influence of Parameter Spread

We next investigate the influence of deviations of the gains and system noise values from their nominal values. The number of samples was fixed at $N = 2^{15}$ and the SNR at nominal $-10$ dB. The gain magnitude and noise parameter ranges are defined by $\boldsymbol{\gamma} = 10^{0.1\Delta\gamma\boldsymbol{\beta}_\gamma}$ and $\mathbf{d} = \mathrm{SNR}^{-1}10^{0.1\Delta d\boldsymbol{\beta}_d}$, where $\Delta\gamma$ and $\Delta d$ are spreading parameters with values ranging from 0 to 10 dB, and $\boldsymbol{\beta}_\gamma$ and $\boldsymbol{\beta}_d$ are random vectors with elements uniformly distributed in the interval $[-1, 1]$.

Figs. 6 and 7 show the gain estimation standard deviation as a function of gain spreading for the nonweighted and weighted simulations, respectively. For $\Delta\gamma = 0$ dB, all gains are unity; for $\Delta\gamma = 10$ dB, the gains vary between $-10$ and 10 dB (with respect to the nominal value of 1). The effect of the weighting on the statistical efficiency of the LOGLS method is dramatic for a large gain spreading parameter. The weighting does not influence the gain magnitude estimation of the COLR method. This is expected as the COLR weighting is a real diagonal matrix. The ALS and GNLS methods coincide with the CRB curve, except for a large gain spread parameter value in the nonweighted case.

Figs. 8 and 9 show the gain estimation standard deviation as a function of noise spreading for the nonweighted and weighted simulations, respectively. All nonweighted method standard deviations lie well above the CRB curves for $\Delta d > 2$ dB and

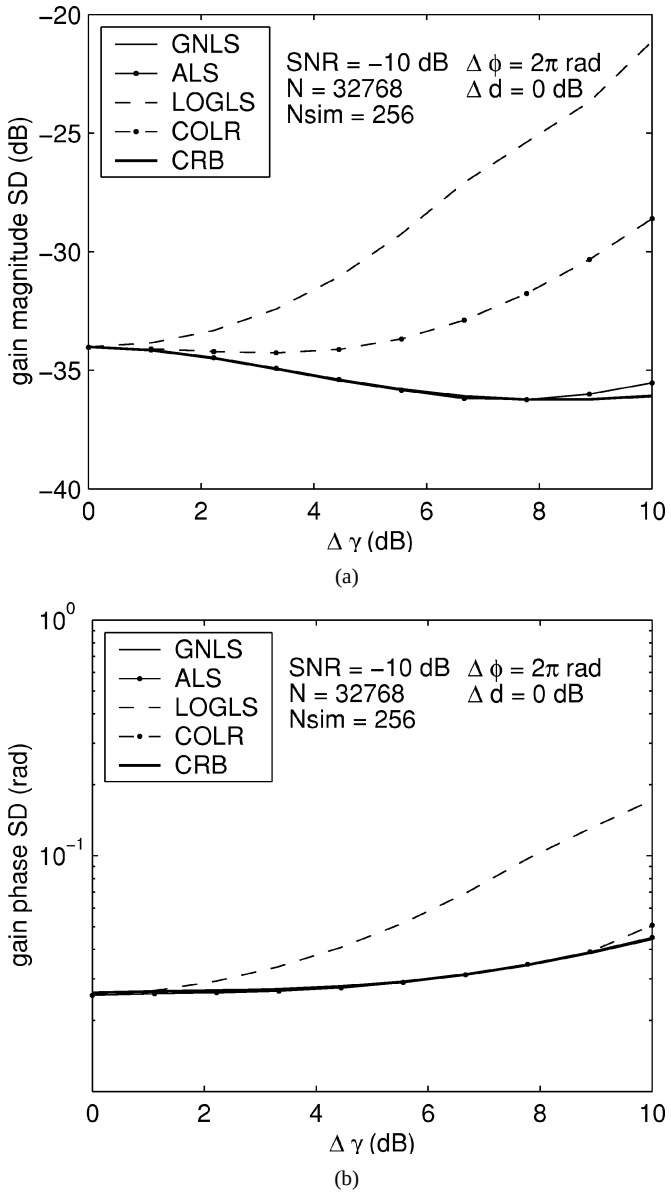


Fig. 6. Nonweighted gain estimation standard deviation versus dispersion of the gain magnitudes. (a) Gain magnitude SD. (b) Gain phase SD.

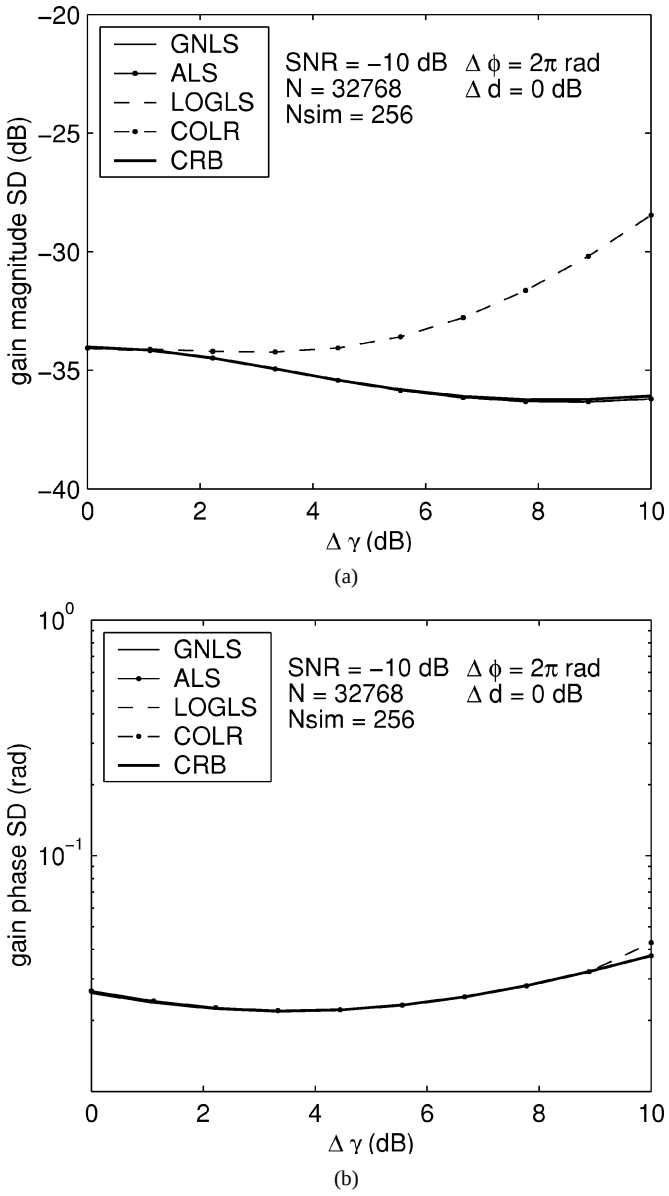


Fig. 7. Weighted gain estimation standard deviation versus dispersion of the gain magnitudes. (a) Gain magnitude SD. (b) Gain phase SD.

coincide with the CRB curves in the weighted case, except for the gain magnitude estimation for the COLR method.

We conclude that weighting does not have much influence if the gain and noise values of the different telescopes are approximately equal, as in this case the methods are already asymptotically efficient. The improvement, however, is large if gain and noise spread become significant. In that case, weighting makes the methods (except COLR) asymptotically efficient.

## V. Experimental Results

### A. Measurement Setup

So far, the accuracy of gain parameter estimations has been verified by means of simulations. In this section, we present an experimental verification based on collected radio telescope data. In the experiment, we observe a strong point source in the sky with a radio telescope interferometer array, in our case the WSRT. The point source requirement is that the source angular size is much smaller than the telescope main beam power.

The time series at the baseband digital output of the telescopes are segmented into short intervals, Fourier transformed, and subsequently spatially cross-correlated, resulting in complex covariance matrices for each of the frequency bins, to which the gain decomposition algorithms are applied.

In our experiments, we used $p = 8$ of the 14 WSRT telescopes, in single linear polarization mode, with a maximum distance (baseline) of 1 km. The telescopes tracked the strong astronomical point source "3C48" at a sky frequency of 1420.4 MHz with a receiver bandwidth of 1.25 MHz. An eight-channel data recorder, equipped with eight ADCs and $8 \times 32$ MBytes of memory, was connected to the telescope baseband IF system outputs. The time sample data was recorded on CDROM and processed offline. In our experiment, the earth-rotation related phase drift was compensated for, which means that during the experiment the telescope–interferometer phase was constant.

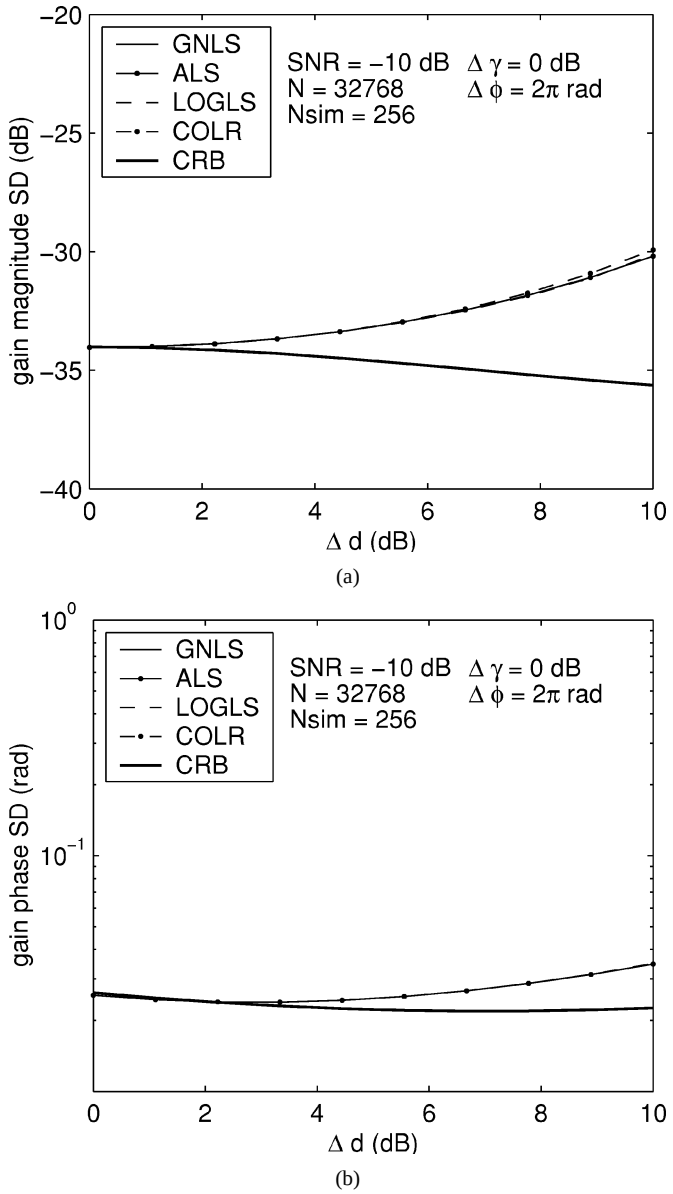


Fig. 8. Nonweighted gain estimation standard deviation versus dispersion of the system noise. (a) Gain magnitude SD. (b) Gain phase SD.

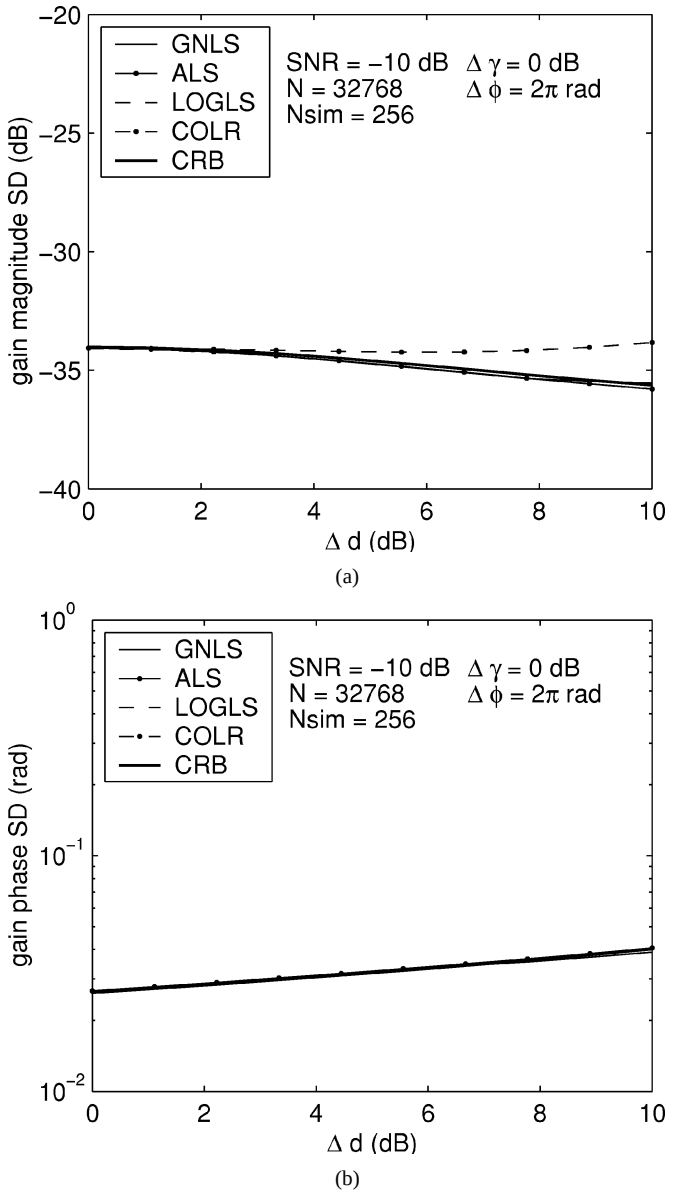


Fig. 9. Weighted gain estimation standard deviation versus dispersion of the system noise. (a) Gain magnitude SD. (b) Gain phase SD.

We split the data into 32 frequency bins, each with a bandwidth of 39 kHz, which fits the narrowband assumption reasonably well (propagation delays across the array).

The data model in the experiment is $\mathbf{R} = \mathbf{g}\sigma_s^2\mathbf{g}^H + \mathbf{D}$, where $\sigma_s$ is the source flux (known from tables). The outcome of the algorithms forms estimates for $\mathbf{g}$ and $\mathbf{D}$. To compare this to known telescope parameters, we first note that the gain vector $\mathbf{g}$ and the noise vector $\mathbf{n}$ contain a common unknown electronic amplification factor $\mathbf{\Gamma}_{el}$ (a diagonal matrix), which is frequency dependent and cannot be obtained separately. Thus, define

$$\mathbf{g} = \mathbf{\Gamma}_{el}\,\underline{\mathbf{g}}, \qquad \mathbf{D} = \underline{\mathbf{D}}\mathbf{\Gamma}_{el}^2$$

where $\underline{\mathbf{g}}$ and $\underline{\mathbf{D}}$ are the telescope gain vector and telescope noise covariance matrix at a location in the system prior to amplification. Note that the ratio $\underline{\mathbf{D}}^{-(1/2)}\underline{\mathbf{g}}$ is independent from the electronic gain. Hence, $\sigma_s\mathbf{D}^{-(1/2)}\mathbf{g} = \sigma_s\underline{\mathbf{D}}^{-(1/2)}\underline{\mathbf{g}}$. The right-hand side is the SNR at the input of the low-noise amplifiers. Its nominal value depends on the construction of the telescopes (collecting area), the telescope system noise, and the source flux and is known from calibration tables. Thus, we can compare the estimate for $\sigma_s\mathbf{D}^{-(1/2)}\mathbf{g}$ to the literature. In our experiment, the 3C48 source emits spectrally continuum radiowaves at 1420.4 MHz with a source power, according to calibration tables, which lies 13 dB below the WSRT system noise or $\sigma_s^2|g_i|^2/d_i \approx 0.05$.

### *B. Experimental Results*

The weighted LOGLS method was applied to a WSRT telescope data set consisting of $p = 8$ telescopes and $N = 131\,072$ samples in 32 frequency bins. The estimates of $\sigma_s^2|g_i|^2$, $d_i$, and $\phi_i$ for $i = 1, \ldots, 8$ as function of frequency are shown in Fig. 10.

Fig. 10(a) confirms that the received SNR is about $-13$ dB for each antenna. Note that there is a bump in the noise power

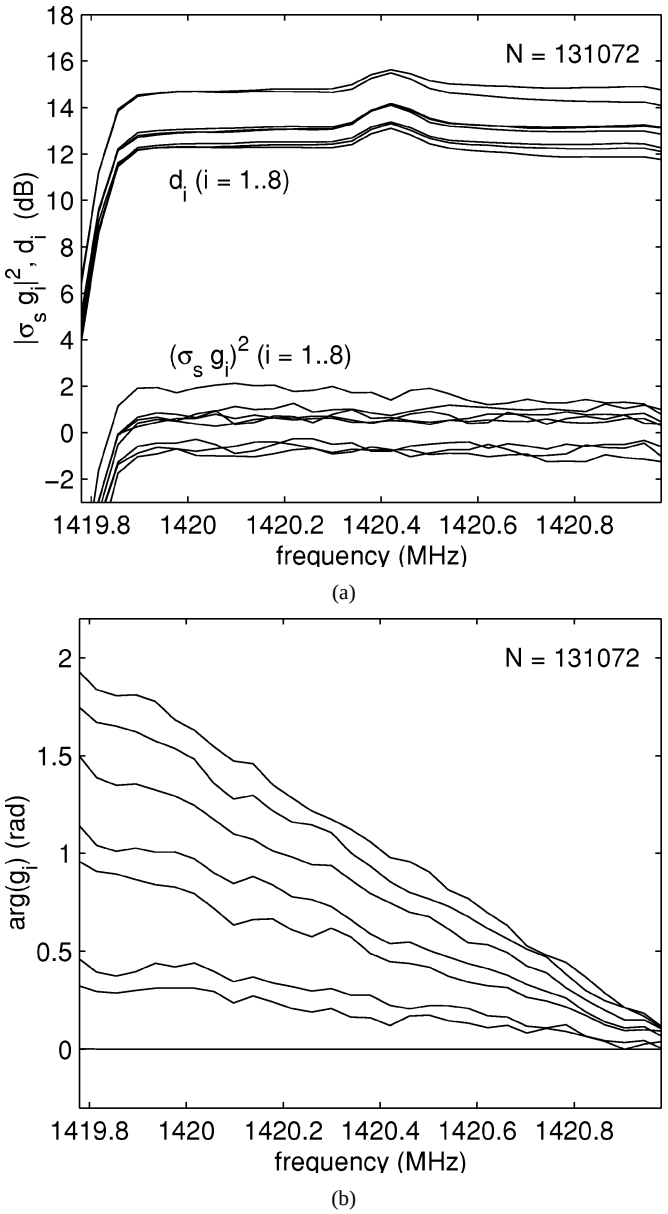


Fig. 10. Gain magnitude, (a) noise power estimates, and (b) gain phase estimates, as a function of frequency. Estimates are obtained using the weighted LOGLS method, based on an observation of the astronomical source 3C48.

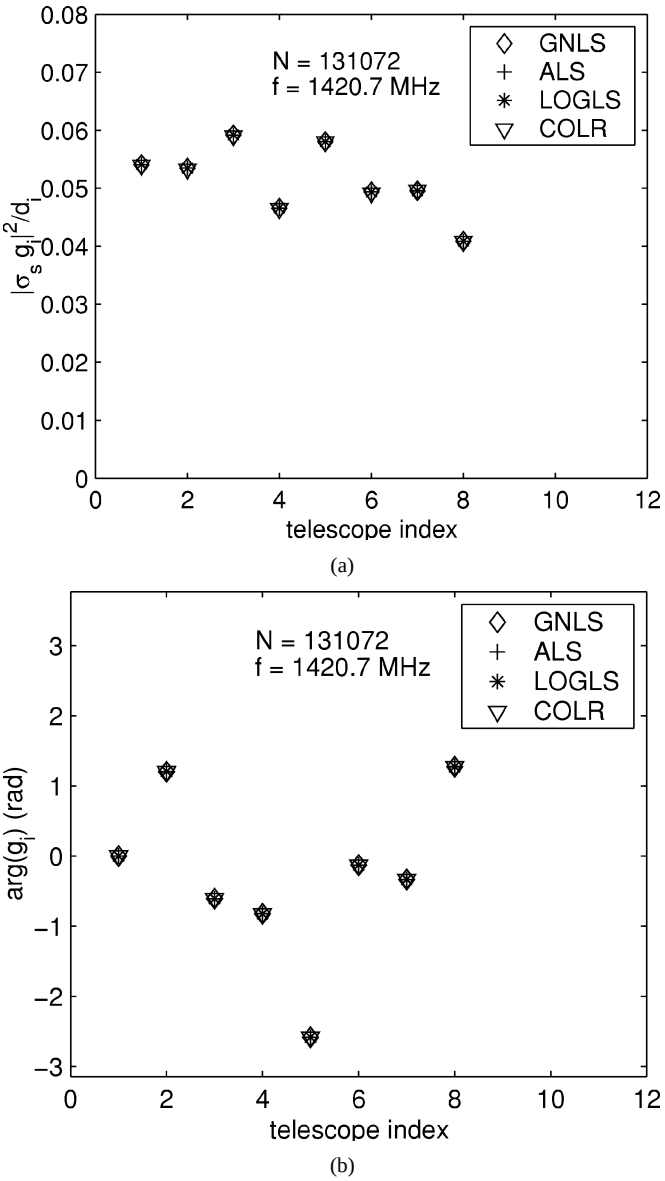


Fig. 11. (a) Observed SNR and (b) estimated gain phase for the astronomical source 3C48 at $f = 1420.7$ MHz and channel bandwidth $\Delta f = 39$ kHz.

curves at 1420.4 MHz. This corresponds to the spectral line of neutral hydrogen and is caused by the galactic emission of our Milky Way. As the Milky Way is a spatially wide source of radiowaves, it is not resolved by the WSRT interferometers and is therefore visible only in the noise estimates. Fig. 10(b) shows the estimated phases of the telescopes, which are frequency dependent because of the geometric delay of the incident source across the array. The horizontal line corresponds to the first telescope: the steepest line to the farthest. The phase slope over a frequency band $\Delta f$ is given by $\Delta\phi_i = 2\pi(\Delta f)L_i/c$, where $L_i$ is the geometric delay of telescope $i$. For the longest telescope distance, the calculated phase slope over the passband is 1.9 rad, which matches the observed value.

Next, all four weighted gain estimation methods are applied to the same dataset but for one frequency bin only. The ratio of the gains and noise vector components $\sigma_s^2|g_i|^2/d_i$ and the observed phases are plotted in Fig. 11. All four estimators yield the same ratios of about 0.05, which is the expected value. Note that although the variation among the telescopes of gain magnitudes and noise power in Fig. 10 is about 3 dB, the fluctuation in the gain-noise *ratios* (Fig. 11) is only 25% around the average. This is because the electronic gain variation is quite large, and this is factored out by the division.

Finally, we investigate the effect of the data sample size $N$ on the estimates. The data set was split into subsets increasing in size from $N = 512$ to 65 536 in steps of a factor two. The number of subsets decreased correspondingly from $N_{sim} = 256$ down to 2, as the total number of available samples is constant (131 072). Both the estimation standard deviation and the CRB are shown in Fig. 12, where the CRB was derived from the "true" values estimated from the complete data set. There is a fair match between the observed standard deviation and the theoretical bound, except at the edges where either $N$ or $N_{sim}$ are too small to obtain reliable statistics.

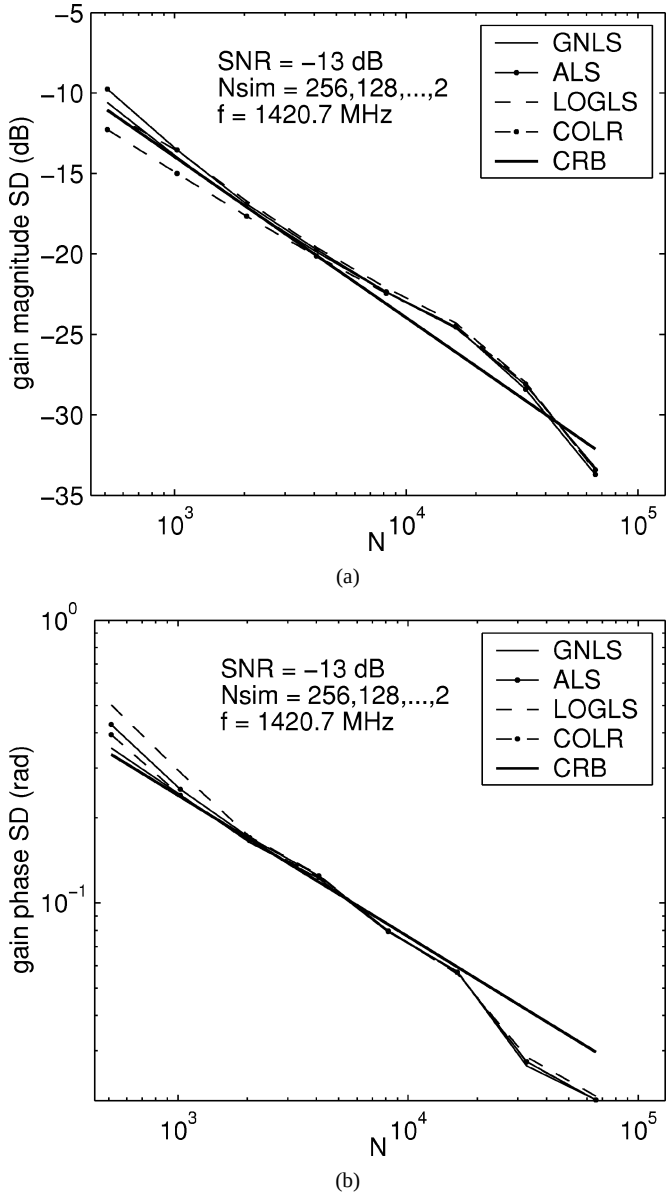


Fig. 12. Gain estimation standard deviation as function of the integration length ($f = 1420.7$ MHz, $\Delta f = 39$ kHz). (a) Gain magnitude SD. (b) Gain phase SD.

## VI. Conclusions

In this paper, we have derived several algorithms for estimating gain and noise parameters of a phased array from an observed covariance matrix. We have shown asymptotically efficient weighted least square estimators and have derived several closed form estimators that, under certain conditions, are also asymptotically efficient. For low SNRs, in the regime where all algorithms under consideration are biased, the LS estimators perform slightly better than the closed-form *ad hoc* algorithms. Nonetheless, the closed-form algorithms provide the essential starting points for the iterative methods (alternating LS and Gauss–Newton LS optimization).

For uniform gain and noise vectors, the performance of the unweighted algorithms is close to the CRB, and weighting does not improve much. However, for parameters with a large spread (>3 dB), the weighting dramatically improves the performance, making all algorithms (except COLR) asymptotically efficient, and improving the speed of convergence for the ALS algorithm by a factor two.

An advantage of the LOGLS method is its low computational expenditure, proportional to $p^2$ rather than $p^3$, as for the other algorithms. Unfortunately, this method is not easily generalized to the estimation of multiple gain vectors, which are needed, for example, in the calibration of dual-polarized telescope arrays.

## Appendix
## Derivation of Fisher Information Matrix Components

In this Appendix, we elaborate on the form of the Fisher information matrix in (4) and (5). $\mathbf{F}_\gamma$ can be written as

$$\mathbf{F}_\gamma = \frac{\partial(\overline{\mathbf{g}} \otimes \mathbf{g})}{\partial \boldsymbol{\gamma}^t} = \frac{\partial \overline{\mathbf{g}}}{\partial \boldsymbol{\gamma}^t} \otimes \mathbf{g} + \overline{\mathbf{g}} \otimes \frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}^t}$$

where $(\partial\overline{\mathbf{g}}/\partial\boldsymbol{\gamma}^t) \otimes \mathbf{g} = (\mathbf{I} \circ \mathbf{G})\overline{\boldsymbol{\Phi}}$, and where $\overline{\mathbf{g}} \otimes (\partial \mathbf{g}/\partial \boldsymbol{\gamma}^t) = (\overline{\mathbf{G}} \circ \mathbf{I})\boldsymbol{\Phi}$. It follows readily that

$$\mathbf{F}_\gamma = \left(\overline{\mathbf{G}} \circ \mathbf{I}\right) \boldsymbol{\Phi} + (\mathbf{I} \circ \mathbf{G})\overline{\boldsymbol{\Phi}} \tag{27}$$

$$\mathbf{F}_\phi = \imath \left( \left(\overline{\mathbf{G}} \circ \mathbf{I}\right) \boldsymbol{\Gamma}\boldsymbol{\Phi} - (\mathbf{I} \circ \mathbf{G})\boldsymbol{\Gamma}\overline{\boldsymbol{\Phi}} \right) \mathbf{I}_s$$

$$\mathbf{F}_d = \mathbf{I} \circ \mathbf{I} \tag{28}$$

where $\mathbf{I}_s$ is a selection matrix that is equal to the identity matrix with its first column removed so that the derivative to $\phi_1$ is omitted.

The FIM can be partitioned as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{\gamma\gamma} & \mathbf{M}_{\gamma\phi} & \mathbf{M}_{\gamma d} \\ \mathbf{M}_{\phi\gamma} & \mathbf{M}_{\phi\phi} & \mathbf{M}_{\phi d} \\ \mathbf{M}_{d\gamma} & \mathbf{M}_{d\phi} & \mathbf{M}_{dd} \end{bmatrix}. \tag{29}$$

We now show that

$$\mathbf{M}_{\gamma\phi} = \mathbf{F}_\gamma^H \left(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}\right) \mathbf{F}_\phi = 0 \tag{30}$$

$$\mathbf{M}_{d\phi} = \mathbf{F}_d^H \left(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}\right) \mathbf{F}_\phi = 0 \tag{31}$$

so that[2]

$$\mathbf{M}^{-1} = \begin{bmatrix} (\mathbf{M}_{\gamma\gamma} - \mathbf{M}_{\gamma d}\mathbf{M}_{dd}^{-1}\mathbf{M}_{d\gamma})^{-1} & 0 & * \\ 0 & \mathbf{M}_{\phi\phi}^{-1} & 0 \\ * & 0 & (\mathbf{M}_{dd} - \mathbf{M}_{d\gamma}\mathbf{M}_{\gamma\gamma}^{-1}\mathbf{M}_{\gamma d})^{-1} \end{bmatrix}$$

which indicates that the phase parameters are decoupled from the gain magnitude and noise parameters.

To show (30), we start with $\mathbf{M}_{\gamma\phi} = \mathbf{F}_\gamma^H(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1})\mathbf{F}_\phi$, where inserting the Jacobians (27) and (28) produces

$$\mathbf{M}_{\gamma\phi} = \left( \left(\overline{\mathbf{G}} \circ \mathbf{I}\right) \boldsymbol{\Phi} + (\mathbf{I} \circ \mathbf{G})\overline{\boldsymbol{\Phi}} \right)^H \left( \overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1} \right) \cdot \imath \left( \left(\overline{\mathbf{G}} \circ \mathbf{I}\right) \boldsymbol{\Gamma}\boldsymbol{\Phi} - (\mathbf{I} \circ \mathbf{G})\boldsymbol{\Gamma}\overline{\boldsymbol{\Phi}} \right) \mathbf{I}_s.$$

[2]Here, $*$ denotes certain sections of the matrix that are not of interest.

Factoring out this equation produces four terms

$$\begin{aligned}\mathbf{M}_{\gamma\phi} = &\imath\overline{\mathbf{\Phi}}\left(\left(\overline{\mathbf{G}}^{H}\overline{\mathbf{R}}^{-1}\overline{\mathbf{G}}\right)\odot\mathbf{R}^{-1}\right)\mathbf{\Gamma}\mathbf{\Phi}\mathbf{I}_s\\ &-\imath\overline{\mathbf{\Phi}}\left(\left(\overline{\mathbf{G}}^{H}\overline{\mathbf{R}}^{-1}\right)\odot(\mathbf{R}^{-1}\mathbf{G})\right)\mathbf{\Gamma}\overline{\mathbf{\Phi}}\mathbf{I}_s\\ &+\imath\mathbf{\Phi}\left(\left(\overline{\mathbf{R}}^{-1}\overline{\mathbf{G}}\right)\odot\left(\mathbf{G}^{H}\mathbf{R}^{-1}\right)\right)\mathbf{\Gamma}\mathbf{\Phi}\mathbf{I}_s\\ &-\imath\mathbf{\Phi}\left(\overline{\mathbf{R}}^{-1}\odot\left(\mathbf{G}^{H}\mathbf{R}^{-1}\mathbf{G}\right)\right)\mathbf{\Gamma}\overline{\mathbf{\Phi}}\mathbf{I}_s.\end{aligned}$$

This can be expanded using $\mathbf{G}=\mathbf{g}\mathbf{1}^t=\mathbf{\Phi}\boldsymbol{\gamma}\mathbf{1}^t$ and by defining $\mathbf{R}=\mathbf{\Phi}\mathbf{R}_a\overline{\mathbf{\Phi}}$, where $\mathbf{R}_a$ is a matrix containing the absolute values of the elements of the matrix $\mathbf{R}$. The result is

$$\begin{aligned}\mathbf{M}_{\gamma\phi} = &\imath\overline{\mathbf{\Phi}}\left(\left(\mathbf{1}\boldsymbol{\gamma}^t\mathbf{\Phi}\overline{\mathbf{\Phi}}\mathbf{R}_a^{-1}\mathbf{\Phi}\overline{\mathbf{\Phi}}\boldsymbol{\gamma}\mathbf{1}^t\right)\odot\left(\mathbf{\Phi}\mathbf{R}_a^{-1}\overline{\mathbf{\Phi}}\right)\right)\mathbf{\Gamma}\mathbf{\Phi}\mathbf{I}_s\\ &-\imath\overline{\mathbf{\Phi}}\left(\left(\mathbf{1}\boldsymbol{\gamma}^t\mathbf{\Phi}\overline{\mathbf{\Phi}}\mathbf{R}_a^{-1}\mathbf{\Phi}\right)\odot\left(\mathbf{\Phi}\mathbf{R}_a^{-1}\overline{\mathbf{\Phi}}\mathbf{\Phi}\boldsymbol{\gamma}\mathbf{1}^t\right)\right)\mathbf{\Gamma}\overline{\mathbf{\Phi}}\mathbf{I}_s\\ &+\imath\mathbf{\Phi}\left(\left(\overline{\mathbf{\Phi}}\mathbf{R}_a^{-1}\mathbf{\Phi}\overline{\mathbf{\Phi}}\boldsymbol{\gamma}\mathbf{1}^t\right)\odot\left(\mathbf{1}\boldsymbol{\gamma}^t\overline{\mathbf{\Phi}}\mathbf{\Phi}\mathbf{R}_a^{-1}\overline{\mathbf{\Phi}}\right)\right)\mathbf{\Gamma}\mathbf{\Phi}\mathbf{I}_s\\ &-\imath\mathbf{\Phi}\left(\left(\overline{\mathbf{\Phi}}\mathbf{R}_a^{-1}\mathbf{\Phi}\right)\odot\left(\mathbf{1}\boldsymbol{\gamma}^t\overline{\mathbf{\Phi}}\mathbf{\Phi}\mathbf{R}_a^{-1}\overline{\mathbf{\Phi}}\mathbf{\Phi}\boldsymbol{\gamma}\mathbf{1}^t\right)\right)\mathbf{\Gamma}\overline{\mathbf{\Phi}}\mathbf{I}_s.\end{aligned}$$

By noting that $\mathbf{\Phi}\overline{\mathbf{\Phi}}=\mathbf{I}$, it follows that all factors $\mathbf{\Phi}$ and $\overline{\mathbf{\Phi}}$ cancel each other. Since the second and the third term of the equation cancel each other, and so do the first and fourth term, $\mathbf{M}_{\gamma\phi}=0$. In the same way, it can be shown that $\mathbf{M}_{\phi d}=0$. The nonzero FIM components can be derived following the same procedure. The results are, using the definition $\alpha=\boldsymbol{\gamma}^t\mathbf{R}_a^{-1}\boldsymbol{\gamma}$, and using the fact that $\mathbf{M}$ is Hermitian

$$\begin{aligned}\mathbf{M}_{\gamma\gamma} &= 2\left(\alpha\mathbf{R}_a^{-1}+\mathbf{R}_a^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^t\mathbf{R}_a^{-1}\right)\\ \mathbf{M}_{\gamma\phi} &= \mathbf{M}_{\phi\gamma}=0\\ \mathbf{M}_{\gamma d} &= \mathbf{M}_{d\gamma}=2\left(\mathbf{1}\boldsymbol{\gamma}^t\mathbf{R}_a^{-1}\odot\mathbf{R}_a^{-1}\right)\\ \mathbf{M}_{\phi\phi} &= 2\mathbf{I}_s^t\mathbf{\Gamma}\left(\alpha\mathbf{R}_a^{-1}-\mathbf{R}_a^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^t\mathbf{R}_a^{-1}\right)\mathbf{\Gamma}\mathbf{I}_s\\ \mathbf{M}_{\phi d} &= \mathbf{M}_{d\phi}=0\\ \mathbf{M}_{dd} &= \overline{\mathbf{R}}^{-1}\odot\mathbf{R}^{-1}=\mathbf{R}_a^{-1}\odot\mathbf{R}_a^{-1}.\end{aligned}$$

The FIM components are real and do not depend on $\boldsymbol{\phi}$. This proves that the estimation accuracy bounds are independent of the particular values of the phases.

## REFERENCES


[1] R. P. F. Schwab and A. H. Bridle, "Synthesis imaging in radio astronomy," in *Proc. Astron. Soc. Pacific Conf. Series*, vol. 6, 1994.

[2] A. Thompson, J. Moran, and G. Swenson, *Interferometry and Synthesis in Radio Astronomy*, first ed. New York: Wiley, 1986.

[3] H. van Someren-Greve, "Logarithmic least square gain decomposition algorithm for the WSRT," IWOS software documentation, Dwingeloo, The Netherlands, ASTRON internal document, 1980.

[4] T. Cornwell and P. N. Wilkinson, "A new method for making maps with unstable radio interferometers," *Proc. Mon. Not. R. Astron. Soc.*, vol. 196, pp. 1067–1086, 1981.

[5] M. Wieringa, "An investigation of the telescope based calibration methods 'redudancy' and 'self-cal'," *Experiment. Astron.*, vol. 2, pp. 203–225, 1992.

[6] A. B. Smolders and M. van Haarlem, Eds., *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*. Dwingeloo, The Netherlands: ASTRON, Apr. 1999.

[7] A. Leshem, A. van der Veen, and A. Boonstra, "Multichannel interference mitigation techniques in radio astronomy," *Astrophys. J. Supplements*, vol. 131, pp. 355–374, Nov. 2000.

[8] M. Pesavento and A. Gershman, "Array processing in the presence of unknown nonuniform sensor noise: A maximum likelihood direction finding algorithm and Cramer–Rao bounds," in *Proc. IEEE Workshop Statist. Signal Process.*, 2000.

[9] P. Stoica and M. Cedervall, "Detection tests for array processing in unknown correlated noise fields," *IEEE Trans. Signal Processing*, vol. 45, pp. 2351–2362, Sept. 1997.

[10] B. Friedlander and A. Weiss, "Direction finding using noise covariance modeling," *IEEE Trans. Signal Processing*, vol. 43, pp. 1557–1567, July 1995.

[11] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Process. Rev. J.*, vol. 8, pp. 185–210, 1998.

[12] J. B. K. V. Mardia and J. T. Kent, *Multivariate Analysis*. London, U.K.: Academic, 1979, Probability and Mathematical Statistics.

[13] D. Lawley and A. Maxwell, *Factor Analysis as a Statistical Method*. London, U.K.: Butterworth, 1971.

[14] S. Haykin, J. Litva, and T. S., Eds., *Radar Array Processing*. Berlin, Germany: Springer-Verlag, 1993, vol. 25.

[15] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, vol. 1.

[16] P. Stoica, B. Ottersten, M. Viberg, and R. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Trans. Signal Processing*, vol. 44, pp. 96–105, Jan. 1996.

[17] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. London, U.K.: Academic, 1995.

[18] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[19] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.

[20] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. Chichester, U.K.: Ellis Horwood, 1981.


**Albert-Jan Boonstra** was born in The Netherlands in 1961. He received the B.Sc. and M.Sc. degrees in applied physics from Groningen University, Groningen, the Netherlands, in 1984 and 1987, respectively. Since 1999, he has also been with the Delft University of Technology, Delft, The Netherlands, where he is pursuing the Ph.D. degree.

He was with the Laboratory for Space Research, Groningen, from 1987 to 1991, where he was involved in developing the short wavelength spectrometer (SWS) for the infrared space observatory satellite (ISO). In 1992, he joined ASTRON, the Netherlands Foundation for Research in Astronomy, initially at the Radio Observatory Westerbork, Westerbork, The Netherlands. He is currently with the ASTRON R&D Department, Dwingeloo, The Netherlands, where he is developing radio frequency interference (RFI) mitigation techniques for radio astronomy. His research interests lie in the area of signal processing, specifically RFI mitigation by digital filtering.

**Alle-Jan van der Veen** (S'87–M'94–SM'02) was born in The Netherlands in 1966. He graduated (cum laude) from the Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, in 1988 and received the Ph.D. degree (cum laude) from the same institute in 1993.

Throughout 1994, he was a postdoctoral scholar at Stanford University, Stanford, CA, in the Scientific Computing/Computational Mathematics group and in the Information Systems Laboratory. At present, he is a Full Professor with the Signal Processing Group of DIMES, Delft University of Technology. His research interests are in the general area of system theory applied to signal processing and, in particular, algebraic methods for array signal processing.

Dr. van der Veen was the recipient of the 1994 and 1997 IEEE SPS Young Author paper awards and was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1998 to 2001. He is currently Chairman of the IEEE SPS SPCOM Technical Committee and Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS.