Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 84 (2004) 619–629

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Direction of arrival (DOA) parameter estimation with the SAGE algorithm

Jo Verhaevert*, Emmanuel Van Lil, Antoine Van de Capelle

K.U.Leuven, Div. ESAT-TELEMIC, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

Received 18 July 2002; received in revised form 28 April 2003

Abstract

In this study, we investigate the estimation of the relative delay, the azimuth and elevation angle and the complex amplitude of a known received signal with the space alternating generalised expectation (SAGE) maximisation algorithm. This optimisation algorithm is used to replace the high-dimensional optimisation procedure, necessary to compute the joint maximum likelihood estimate of the parameters using several separate maximisation processes, which can be performed sequentially. Spherical arrays provide wide scan coverage with low grating lobe levels. Because a uniform antenna element distribution on this spherical surface gives rise to the smallest variations in received signal levels over the whole angular space, a uniform distribution of antennas on a sphere is used. For this completely general three-dimensional array, we have for the first time investigated the resolution of the SAGE algorithm, the convergence and the bit error rate as a function of the signal to noise ratio. We will show that this algorithm is a powerful tool that can be applied successfully for high-resolution parameter extraction.

Keywords: Smart antennas; DOA; Maximum likelihood; SAGE algorithm

1. Introduction

The spatial resolution has been the object of many studies based on incident direction of arrival (DOA) algorithms, where the different users are located by distance and angle. In this work, we propose a technique derived from the maximum likelihood (ML) principle, which allows for high-resolution determination of propagation delay, azimuth and elevation incidence angle and amplitude of the complex signal. The expectation maximisation (EM) algorithm [7] updates all parameters simultaneously, which might imply a slow convergence and a difficult maximisation. In [3], the space alternating generalised expectation (SAGE) maximisation method is described. This algorithm updates the parameters sequentially by replacing the high-dimensional optimisation process, necessary to compute the estimates of the parameters, by several separate, low-dimensional maximisation procedures, which are performed sequentially. These separate maximisation procedures are linked up in the SAGE algorithm [2,4,5]. In this paper, we present results of a detailed study of the performance of the SAGE algorithm, described in [6]. We discuss the resolution of the algorithm, the convergence speed

* Corresponding author. Tel.: +32-16-32-11-51; fax: +32-16-32-19-86.
E-mail address: jo.verhaevert@esat.kuleuven.ac.be (J. Verhaevert).

0165-1684/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2003.12.003

and the bit error rate (BER) as a function of the signal to noise ratio (SNR) in the general three-dimensional case. In previous work, the two-dimensional case was explained [9].

The ambiguity of detecting different users in a three-dimensional case with a two-dimensional antenna configuration, can be resolved by using, for instance, two parallel planes with a quarter of a wavelength distance. This solution suffers from the spatially limited radiation pattern of the array elements. By using spherical arrays, there will always be a number of elements radiating into the direction of the signals, hence increasing the SNR of the subsequent digital processing. To minimise the coupling and hence to increase the decorrelation between antenna elements, they should be spaced uniformly over the sphere. This configuration will also increase the resolution of the algorithm.

This paper is organised as follows. Section 2 handles the uniform spherical distribution. Section 3 presents the used model description and Section 4 the SAGE algorithm itself. The initialisation of the SAGE algorithm will be discussed in Section 5. In Section 6, the different performance results will be explained.

2. Uniform spherical distribution

Uniform spherical distributions can be calculated by plotting a regular polyhedron based on regular polygons in a sphere. In previous work [10], exact uniform spherical distributions were derived and calculated. Only five realisations are possible. The feasible regular polyhedrons are rather well known and shown in Table 1. The names of the regular polyhedrons (named after the number of lateral faces) are shown together with their number of the lateral faces and their number of angular points. These angular points, also points on the circumscribed sphere, determine the exact uniform distribution. One can notice that the maximum number of elements that can be placed uniformly on a spherical surface, is 20.

To focus the rest of the research, the SAGE algorithm is applied on a spherical antenna array with a radius $R = \lambda/4$. For a frequency of the carrier signal of 900 MHz, the radius of the circumscribed sphere is $R = 0.0833$ m. Based on this radius, the vertex for each of the configurations is calculated and shown in the last column of Table 1. Moreover, only the octahedral configuration is simulated with the SAGE algorithm. The $N_a = 6$ omnidirectional antennas are placed on the angular points. The co-ordinates of each antenna point are shown in Table 2 and are plotted in Fig. 1.

Table 1
All possible solutions of regular polyhedrons with their number of lateral faces (# LF), with their number of angular points (# AP) and with their vertex l

Name	# LF	# AP	l
Tetrahedron	4	4	0.1360
Octahedron	8	6	0.1178
Hexahedron	6	8	0.0962
Icosahedron	20	12	0.0876
Dodecahedron	12	20	0.0594

Table 2
Co-ordinates of octahedron with circumscribed radius $R = 0.0833$ m

	x	y	z
Point 1	0	0	0.0833
Point 2	0.0833	0	0
Point 3	0	0.0833	0
Point 4	−0.0833	0	0
Point 5	0	−0.0833	0
Point 6	0	0	−0.0833

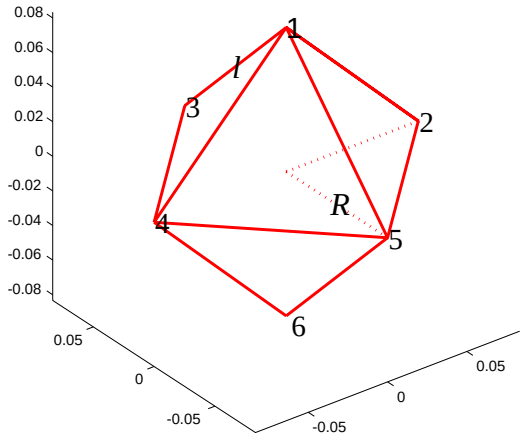

Fig. 1. The octahedral configuration of the antenna array (all distances in m).

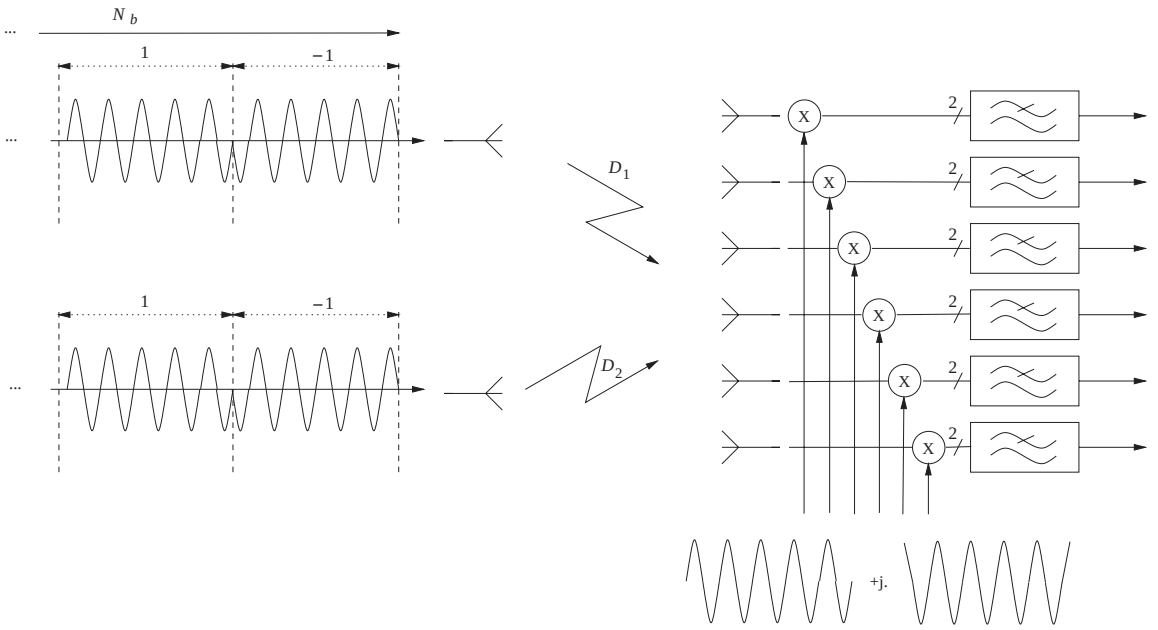

Fig. 2. The model definition.

There may be other types of element distributions that can be used advantageously for a spherical antenna array. But the regular polyhedrons circumscribed in a sphere, have been found to provide exact uniform distributions [8].

3. Model description

Each user sends a known training sequence containing N_b bits. One bit corresponds with 5 periods of the carrier frequency. The carrier signal of the binary phase shift keying (BPSK) modulated signal is 900 MHz and direct sequence-spread spectrum (DS-SS) is used. The N_u users are spread in a three-dimensional space. Because we consider no local scatterers, the number of users N_u is also the number of paths L. The resulting transmitted training sequence is a known combination of these bits. In the channel, the signal is delayed and attenuated, resulting in a relative delay and a complex amplitude at the receiver antennas. To reconstruct the original signal, the incident signals have to be downconverted by multiplying the signal with the signal of a local oscillator. After the downconversion, the signal is filtered in order to get a baseband result. A schematic view can be seen in Fig. 2.

Fig. 3 plots the signal before the processing in the time-domain and in the frequency-domain. The delay and the attenuation of the added signals can be seen in the time-domain signal (left-hand side). The carrier frequency of 900 MHz can easily be checked in the frequency-domain signal (right-hand side).

The signal after the downconversion and the low-pass filtering (as indicated in Fig. 2), is given in Fig. 4. On the left-hand side, the time-domain signal is plotted, the right-hand side gives the frequency-domain signal. The carrier frequency is removed by the downconversion and the frequencies above 900 MHz are cut away by a theoretical perfect low-pass filter.

Because we know the training sequence that has been sent at the receiver side, it is possible to calculate the correlation between the received signal (corrected for one user) and the calculated signal (depending on the (DOA) parameters. If we maximise this cost function, we can extract the correct parameters.

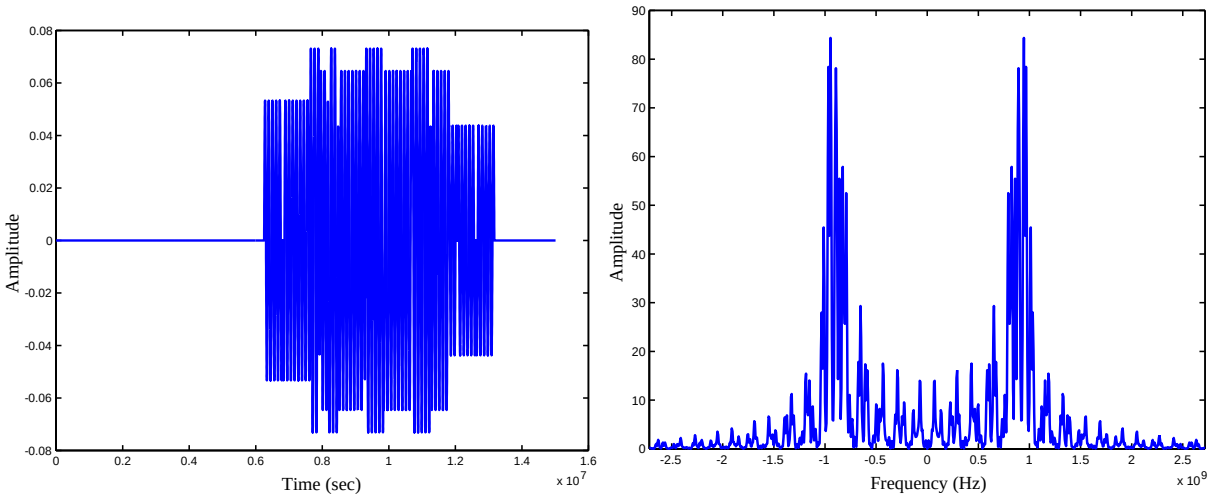

Fig. 3. The received signal on antenna 1 in the time-domain (left) and in the frequency-domain (right) before the processing.

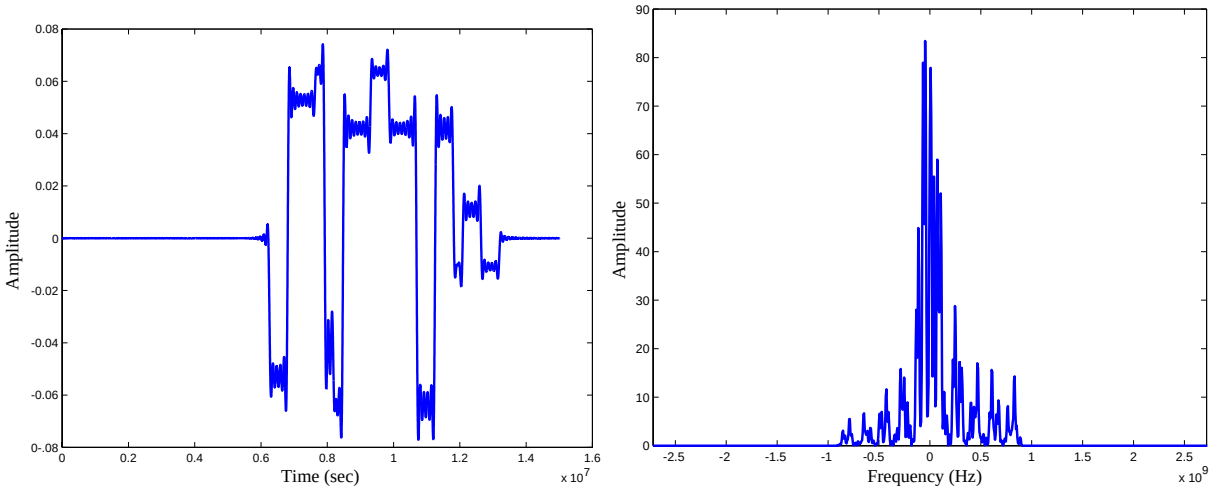

Fig. 4. The received signal on antenna 1 in the time-domain (left) and in the frequency-domain (right) after the processing.

4. The SAGE algorithm

The parameters to be estimated are the relative delay τ_l, the azimuth and elevation angle ϕ_l and θ_l of the incident wave, and the complex amplitude γ_l. The axial convention for the azimuth and elevation angle is plotted in Fig. 5. All these parameters are placed together in a vector ζ_l. The different incident waves form L different ζ_l, because the number of paths is L. These vectors are all stacked together in the matrix ζ. Each step of the SAGE algorithm is an estimate of a subset of the components of ζ, while keeping the estimates of the other components fixed. We define one iteration cycle of the SAGE algorithm as L consecutive iteration steps for updating the parameters of all waves one time.

The correlation function $z(\hat{\tau}, \hat{\phi}, \hat{\theta}, \hat{x}_l(t; \hat{\zeta}))$ between the calculated and the received signal is the cost function that has to be maximised as follows (the symbols with a hat indicate an estimation for

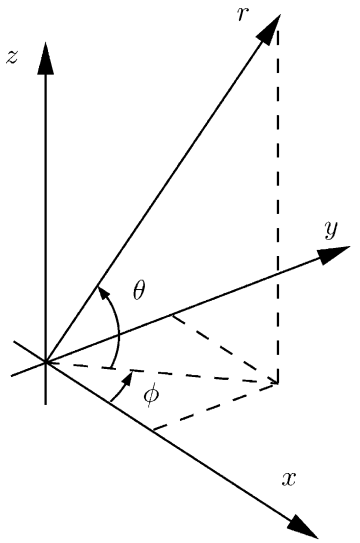

Fig. 5. The axial convention.

that variable):

$$\hat{\tau}'_l = \arg\max_{\tau}\{|z(\tau, \hat{\phi}_l, \hat{\theta}_l, \hat{x}_l(t;\hat{\zeta}))|^2\}, \quad (1)$$

$$\hat{\phi}'_l = \arg\max_{\phi}\{|z(\hat{\tau}'_l, \phi, \hat{\theta}_l, \hat{x}_l(t;\hat{\zeta}))|^2\}, \quad (2)$$

$$\hat{\theta}'_l = \arg\max_{\theta}\{|z(\hat{\tau}'_l, \hat{\phi}'_l, \theta, \hat{x}_l(t;\hat{\zeta}))|^2\}, \quad (3)$$

$$\hat{\gamma}'_l = \frac{1}{N_a T_a}|z(\hat{\tau}'_l, \hat{\phi}'_l, \hat{\theta}'_l, \hat{x}_l(t;\hat{\zeta}))|, \quad (4)$$

where the prime variables denote the new values after the iteration. After each step the estimation vector $\hat{\zeta}$ is updated with the last calculation. Please notify that N_a stands for the number of omnidirectional antennas, and T_a stands for the measured time interval. The correlation function $z(\hat{\tau}, \hat{\phi}, \hat{\theta}, \hat{x}_l(t;\hat{\zeta}))$ is defined as follows:

$$z(\hat{\tau}, \hat{\phi}, \hat{\theta}, \hat{x}_l) = [c(\hat{\phi}, \hat{\theta})]^H \int u^*(t' - \hat{\tau})\hat{x}_l(t';\hat{\zeta})\,dt' \quad (5)$$

with $u^*(t)$ the complex conjugate of the desired signal $u(t)$, which is based on the training sequence and corrected with an estimate of the relative delay $\hat{\tau}$. The ^{H} stands for the Hermitian operator, which is the complex conjugate transpose of that vector. The received signal $x_l(t)$ for one user is calculated based on the received global signal $y(t)$ and on the estimates of the signals $s(t;\hat{\zeta}'_l)$ of the other interfering sources (which might be due to the channel itself, for instance inter-symbol interference (ISI)):

$$x_l(t;\zeta) = y(t) - \sum_{\substack{l'=1 \\ l'\neq l}}^{L} s(t;\zeta_{l'}), \quad (6)$$

$$s(t;\zeta_l) = \gamma_l c(\phi_l, \theta_l)u(t - \tau_l). \quad (7)$$

The direction vector $e(\phi, \theta)$ and the steering vector $c(\phi, \theta)$ are defined as follows (with r the location of the different elements of the antenna array, λ the wave length, θ_l the elevation and ϕ_l the azimuth angle of the lth path):

$$e(\phi_l, \theta_l) = \begin{bmatrix} \cos(\phi_l)\cos(\theta_l) \\ \sin(\phi_l)\cos(\theta_l) \\ \sin(\theta_l) \end{bmatrix} \quad (8)$$

$$c(\phi_l, \theta_l) \triangleq e^{j\frac{2\pi}{\lambda}\langle e(\phi_l,\theta_l), r\rangle}. \quad (9)$$

In the previous formula, the $\langle .,. \rangle$, indicates the inner product of the two vectors.

5. Initialisations

To calculate the correlation between the received and calculated signals, we need a very good initialisation. The received signal is corrected by subtracting the other (interfering) signals. Together with a very good initialisation, a model of the signals has to be known. The convergence of the SAGE algorithm depends to a large extent on the initialisation of $\hat{\zeta}$. In Section 6.4, we will show the influence of different initialisations on the convergence speed. We start the initialisation procedure with $\hat{\zeta} = [0, 0, ...0]$. The further steps of our initialisation procedure of this parameter $\hat{\zeta}$ are described in [6].

In order to initialise the SAGE maximisation procedure, the results of another (e.g. MUSIC, ESPRIT, ...) DOA algorithm can be used. Despite the heavy computer calculations, this method can be very accurate, but is not investigated in this paper.

6. Evaluation of performance

For the performance evaluation of this algorithm, we use the exact uniform spherical antenna array, consisting of $N_a = 6$ omnidirectional antennas, calculated

Table 3
Correct values for the 2 mobile users

	User 1	User 2
τ (s)	6.254×10^{-8}	7.610×10^{-8}
ϕ (degrees)	-86.70	111.12
θ (degrees)	35.13	-52.60
γ (V)	$-0.026 + 0.006i$	$-0.002 + 0.022i$
$k_{x,n}$	0.0471	-0.2189
$k_{y,n}$	-0.8165	0.5666

in the first part of this paper. We will envisage a scenario with $N_u = 2$ mobile users. Because there are no reflections implemented, the number of multipaths $L = 2$. We receive the signals on the octahedron (a spherical equispaced array) and the SAGE algorithm calculates an estimate of the relative delay, the azimuth and the elevation angle of the incident waves and the complex amplitude of the different paths.

6.1. One-dimensional algorithm resolution

We can plot the resolution of the SAGE algorithm by calculating the cost function as a function of a parameter to be maximised, according to Eq. (5). Table 3 contains the correct values. The results are plotted in Figs. 6–8, where the correct values are indicated with dots. By maximising the cost function, we notice that the exact values can be obtained. The plots are essentially used to show the difference in resolution between time and angle.

Fig. 6 shows two narrow peaks around the correct values of the delay time. If there is a small difference between the exact and the calculated value of τ, the cost function, depending on the correlation between the calculated and the correct signal, varies a lot. The resolution, depending on the chip rate of the training sequence, can be computed by determing the distance between the two points -3 dB below the maximum: it is in both cases 1.67 ns or 0.5 m.

In contrast, in Fig. 7, we see two broad peaks around the correct values of ϕ. The steering vector influences these curves of the cost function: a small difference between the real and the calculated value of ϕ only influences the steering vector and *not* the correlation function in the cost function. The flatness of the steering vector is due to the small size of the detector array. The azimuthal resolution can be improved by increasing the radius of the antenna array. The -3 dB azimuthal resolution is 91.7° and 126.1°, respectively for user 1 and user 2.

The steering vector depends not only on the azimuth angle ϕ, but also on the elevation angle θ. Therefore,

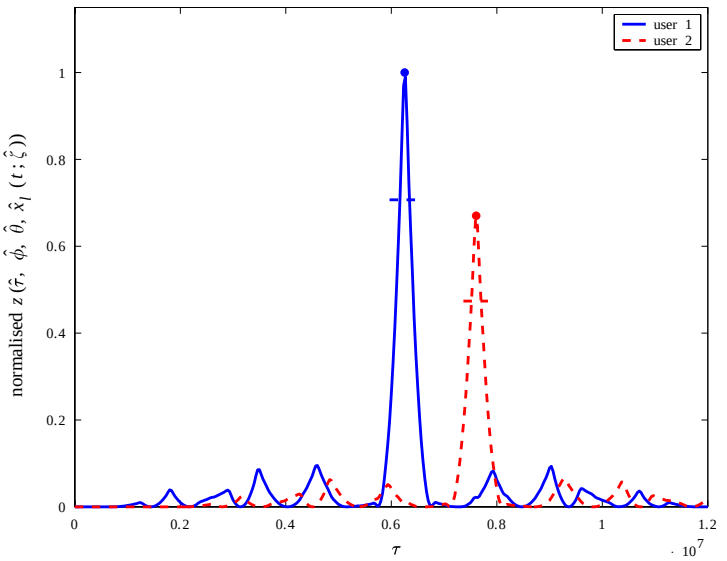

Fig. 6. Algorithm resolution: normalised cost function as a function of τ (in s).

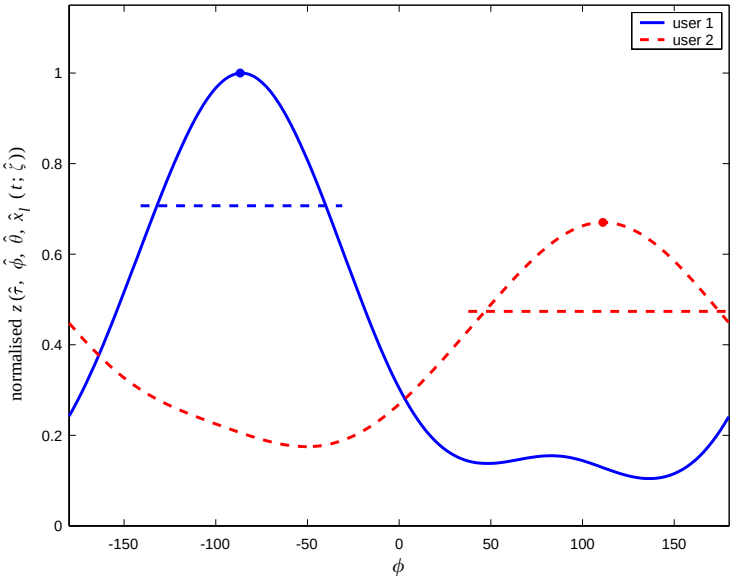

Fig. 7. Algorithm resolution: normalised cost function as a function of ϕ (in degrees).

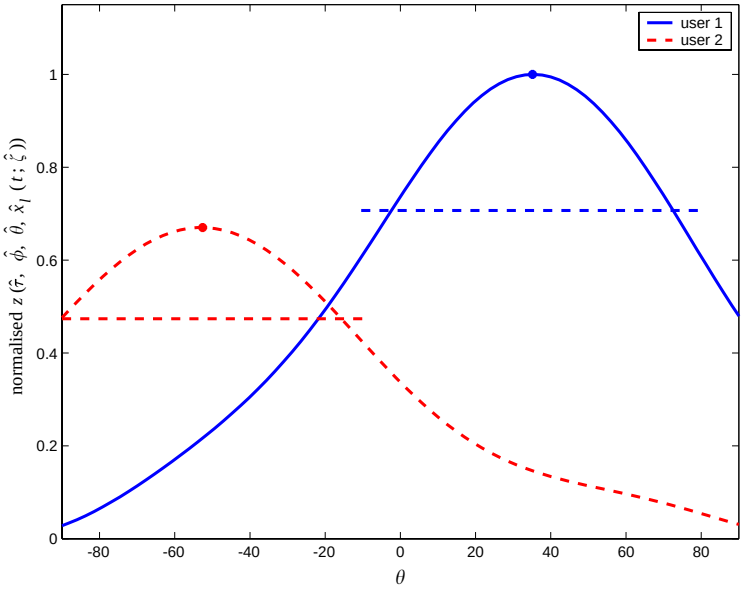

Fig. 8. Algorithm resolution: normalised cost function as a function of θ (in degrees).

a similar curve, as in Fig. 7, is obtained for the dependence of θ in the cost function in Fig. 8. Once again, a small difference between the real and the calculated value of θ only influences the steering vector and *not* the correlation function in the cost function. Also the elevation resolution can be improved by increasing the radius and so narrowing the steering vector. For user 1, the -3 dB elevation resolution is 73.3°

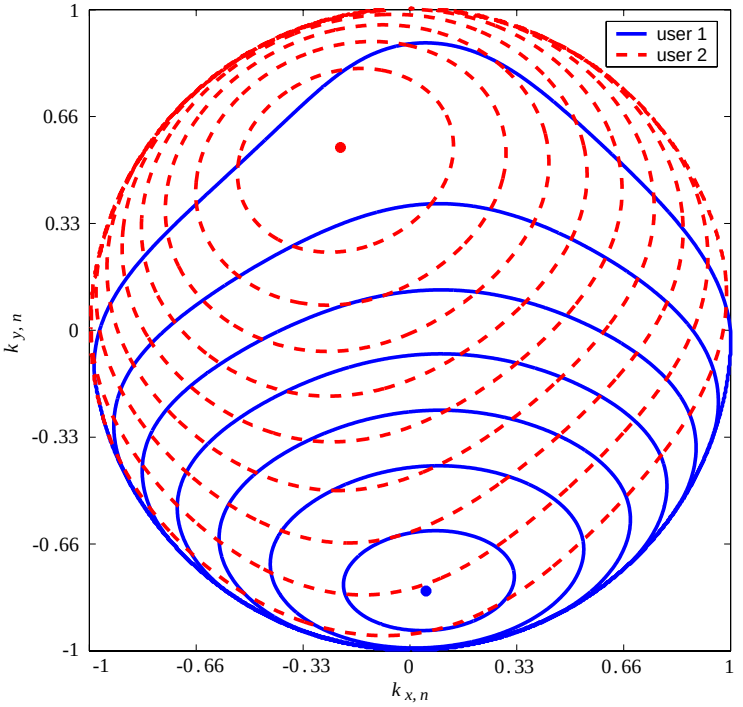

Fig. 9. Algorithm resolution: contours of cost function in the normalised wave space.

and for user 2 is 71.0°. Both can be approximated by the −3 dB beamwidth of the corresponding uniform array. A comparison of the resolution of these three parameters is made in Section 6.2, where the results from Figs. 6–8 are combined.

6.2. Three-dimensional algorithm resolution

The normalised wave numbers (based on the radial positions ϕ and θ) are defined as follows:

$$k_{x,n} = \cos(\phi)\cos(\theta), \tag{10}$$

$$k_{y,n} = \sin(\phi)\cos(\theta). \tag{11}$$

In Fig. 9, the cost function is visualised as a function of the earlier defined normalised wave numbers (depending on ϕ and θ). The correct values are indicated with dots and can be checked with the values in Table 3. Based on this figure, one can conclude that the angular resolution is comparable in both directions.

To represent this cost function depending on three parameters (delay τ, azimuth and elevation angle ϕ and θ) a hologram should be used, giving raise to curves that can be compared with for instance electron densities around the nucleus of an atom. Here, we have taken the values of the parameters $\tau = f(\phi, \theta)$ corresponding with 80% of the cost function at the maximum (in the comparison with electron density profiles, the maximum would occur at the nucleus of the atom). Normally a given direction (ϕ, θ) intersects the 80% cost contour with more than one point (normally two). The values chosen in Fig. 10 correspond with their maximum value $\tau_{\max}$ (indicated by the solid line). In order to illustrate the global resolution of the algorithm (both angular and radial), the same contour is plotted, but corresponding with $\tau = 1.005\tau_{\max}$ (dashed line in Fig. 10).

By comparing the two previous contours and by connecting corresponding points, the sensitivity and hence the resolution of the algorithm can be derived. In Fig. 10, the time delay is indicated with values of τ and the angular spread with normalised wave numbers

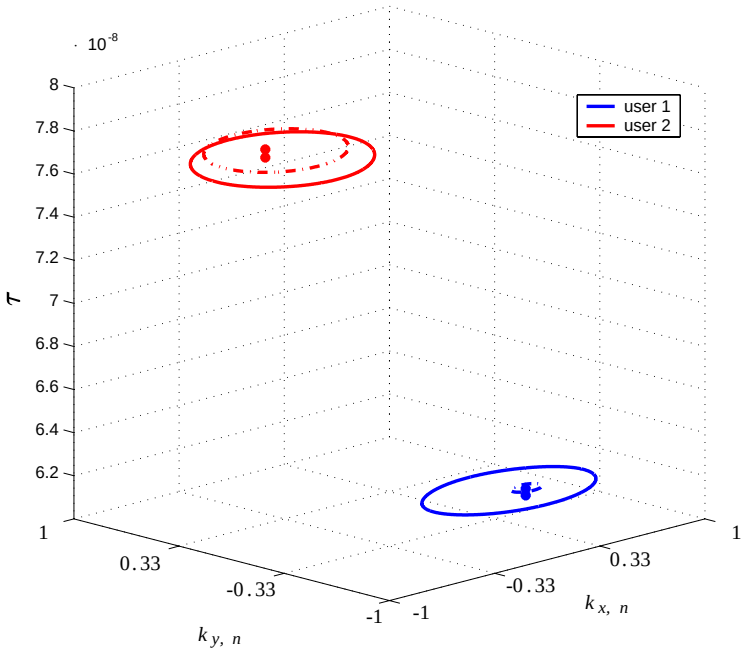

Fig. 10. Algorithm resolution: contours of cost function. The horizontal is the normalised wave space and the vertical dimension is the relative delay (in s).

($k_{x,n}$ and $k_{y,n}$). We can conclude that the estimate of the time delay is much more accurate than the angular estimate, because the contours are nearby the same height. A small excess (0.5%) in the radial position results in a big difference in contours. This confirms the two-dimensional simulation results in [9].

Based on Figs. 9 and 10 and based on the explanation above, we can conclude that estimating the relative time delay (or the distance) with the SAGE algorithm has a very high resolution. The algorithm will be able to discriminate more problematic multi-user scenarios.

6.3. Bit error probability

The probability of a bit error or BER for BPSK is given by [1] as a function of the SNR

$$\mathrm{BER} = \frac{1}{2}\,\mathrm{erfc}(\sqrt{SNR})$$

$$\text{with } \mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t. \qquad (12)$$

Two different BERs are calculated. The first one is obtained with the signal after the downconversion and with the SAGE processing. P_{cal} stands for the power included in the signal, reconstructed with the SAGE results, $P_{\mathrm{cal-won}}$ is the noise power, obtained by the subtraction of the reconstructed signal with the SAGE results, with the received and processed signal without noise, in order to obtain the noise signal, due to both the thermal noise and the 'noise' due to amongst other limited resolution of the data input to the SAGE processing (failure of extraction of interferences):

$$\mathrm{BER}_{\text{calculated values}} = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{P_{\mathrm{cal}}}{P_{\mathrm{cal-won}}}}\right). \qquad (13)$$

The other BER, the theoretical maximum, is calculated as follows: The P_{signal} is the power of the received and processed signal without noise, whereas P_{noise} stands for the power of the received thermal noise:

$$\mathrm{BER}_{\text{theoretical maximum}} = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}}\right). \qquad (14)$$

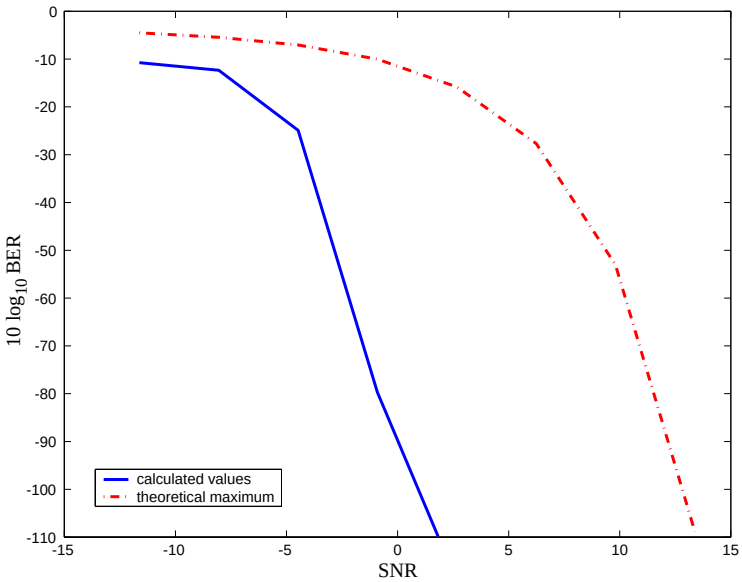

Fig. 11. The BER as a function of SNR: average of 50 runs.

In Fig. 11, both BERs that have been calculated in the two previous formulae, are plotted as a function of the SNR. The $BER_{theoretical\ maximum}$, as a reference scenario, is plotted in a dash-dotted line. It is the BER of the received signal with the processing (downconversion, ...) included. The $BER_{calculated\ values}$ gives therefore the performance of the SAGE algorithm and is plotted in a full line. Even for very low SNR, a BER of -20 dB is possible. In case of high SNR, the calculation accuracy determines the BER. A degradation of 10 dB for the same BER is hence achievable.

6.4. Convergence speed

In this paragraph, we investigate the influence of the initialisation on the convergence speed. We compare the results of the SAGE algorithm with different sets of starting values. The first are the calculated initialisation values of Section 5, based on the procedure described in [6], the second random initialisation values and the third set uses the correct values as starting values. There is no noise added.

First a comparison of the BER in the three different cases is made. The results are plotted in Fig. 12. The estimation started with exact values, converges almost immediately (dash-dotted line). The guess based on the initialisation procedure described in Section 5, converges within three steps of the SAGE algorithm (full line). They both converge to the same value, within the preset accuracy of the SAGE algorithm. In contrast, a complete random guess converges to a local maximum (dashed line).

Based on these two analyses, we can see that a good initialisation is very important.

7. Conclusions

In this paper we investigated the space alternating generalised expectation maximisation (SAGE) algorithm based on uniform spherical and hence for the first time fully three-dimensional distributions. We estimated the relative delay, the azimuth and the elevation angle and the complex amplitude of different signals on an antenna array. Based on contour plots of the three parameters (the relative time delay and the two-dimensional wave numbers), one can conclude that for estimating the relative time delay, the SAGE algorithm can be very accurate. We have proven that a good estimate for initial values reduces processing time significantly.

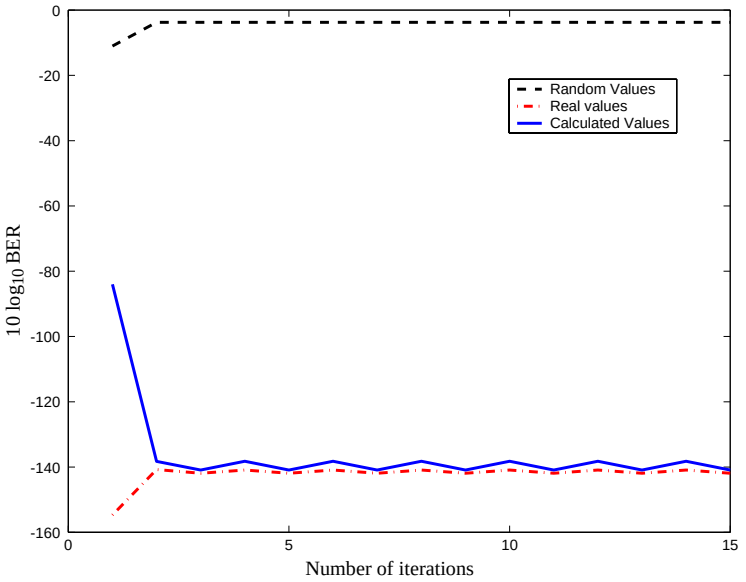

Fig. 12. Convergence speed of the SAGE algorithm: the BER as a function of the number of iterations.

References

[1] L.W. Couch II, Digital and Analog Communication Systems, 5th Edition, Prentice-Hall, Englewood Cliffs, NJ, 1997.

[2] D. Dahlhaus, A. Jarosch, B.H. Fleury, R. Heddergott, Joint demodulation in DS/CDMA systems exploiting the space and time diversity of the mobile radio channel, Eighth IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC 97), Helsinki, Finland, September 1997.

[3] J.A. Fessler, A.O. Hero, Space alternating generalised expectation maximisation algorithm, IEEE Trans. Signal Process. 42 (10) (October 1994) 2664–2677.

[4] B.H. Fleury, D. Dahlhaus, R. Heddergott, M. Tschudin, Wideband angle of arrival estimation using the SAGE Algorithm, IEEE Fourth International Symposium on Spread Spectrum Techniques and Applications, September 1996, pp. 77–85.

[5] B.H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, Channel parameter estimation in mobile radio environments using the SAGE algorithm, IEEE J. Selected Areas Comm. 17 (3) (March 1999) 434–450.

[6] R. Heddergott, Implementierung und untersuchung eines verfahrens zur hochauflösenden schätzung von parametern aus breitbandigen Funkkanalmessungen, Master's Thesis, Institut für Kommunikationstechnik, 1998/1999.

[7] L.B. Nelson, H.V. Poor, Iterative multiuser receivers for CDMA channels: an EM-based approach, IEEE Trans. Comm. 44 (12) (December 1996) 1700–1710.

[8] D.L. Sengupta, T.M. Smith, R.L. Larson, Radiation characteristics of a spherical array of circularly polarized elements, IEEE Trans. Antennas and Propagation 16 (1) (January 1968) 2–7.

[9] J. Verhaevert, E. Van Lil, A. Van de Capelle, Behaviour of the SAGE algorithm in a two-dimensional multi-user environment, Eleventh International Conference on Antennas and Propagation (ICAP 2001), Manchester, UK, April 2001.

[10] J. Verhaevert, E. Van Lil, A. Van de Capelle, Uniform spherical distributions for adaptive arrays, The IEEE Vehicular Technology Conference (VTC2001 Spring), Rhodes, Greece, May 2001.