A constant modulus factorization technique for smart antenna applications in mobile communications

Alle-Jan van der Veen and Arogyaswami Paulraj

Stanford University, Dept. Electrical Eng./ISL
Stanford, Palo Alto, CA 94305
email: allejan@isl.stanford.edu, paulraj@isl.stanford.edu

A fundamental problem in sensor array signal processing is to separate and retrieve all independent co-channel signals that arrive at the antenna array. Such problems arise in smart antenna applications for mobile wireless communication, such as interference reduction and in-cell frequency reuse. In a mobile environment, the presence of large delay multipath makes the array manifold poorly defined, and spatial model methods are not applicable. However, in case the signals have a constant modulus property (as in FDMA/FM systems like AMPS or TACS), iterative algorithms such as Godard and CMA have been used to retrieve the signals. Because of a non-convex optimization criterion, these algorithms suffer from local minima and random convergence behavior, with no satisfactory remedy known as yet. In this paper, we present an algorithm to compute the exact solution to the underlying constant modulus (CM) factorization problem. With this new approach, it is possible to detect the number of CM signals present at the array, and to retrieve all of them exactly, rejecting other, non-CM signals. Only a modest amount of samples are required. The algorithm is robust in the presence of noise, and is tested on real data, collected from an experimental set-up.

Keywords: Constant modulus algorithms, blind beamforming, simultaneous diagonalization.

1. INTRODUCTION

A problem in sensor array signal processing with important applications is concerned with the case where there are several unknown constant modulus (CM) signals impinging on the array, and the objective is to copy each of them without using knowledge of the array response vector. Mathematically, we are given a data matrix $X : m \times n$, with x_{ij} the j-th sample of the i-th antenna, and we have to find a factorization of X, if it exists, as

$$X = AS = \mathbf{a}_1\mathbf{s}_1 + \cdots + \mathbf{a}_d\mathbf{s}_d \,, \tag{1}$$
$$(A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_d] \in \mathbb{C}^{m\times d} \,,\ S = [\mathbf{s}_1^* \ \cdots \ \mathbf{s}_d^*]^* \in \mathbb{C}^{d\times n}) \,,$$

such that A, S are full rank d, and all $\mathbf{s}_k \in \mathbb{C}^n$ are CM signals. The CM factorization problem can be reformulated as finding all weight vectors $\mathbf{w}$ such that $\mathbf{w}X = \mathbf{s}$, for as many linearly independent CM signals $\mathbf{s}$ as possible. This formulation is more general: not all d signals present in X need to be CM signals, but only $\delta \le d$, say.

The CM factorization problem gained much interest in the area of communications, where many modulation or coding schemes produce signals that have the CM property, such as FM and phase-modulated signals. One envisioned application is in wireless telephony, where users broadcast CM signals to a central site. The current AMPS and TACS systems fall within this category. The capacity of these systems is limited by the number of frequency slots that are available. To increase the overall capacity, engineers are considering to reuse frequencies within a single communication cell. Due to co-channel interference of signals that have the same carrier frequency, the output of a receiving antenna will be some linear combination of the signals that have been broadcasted. With multiple antennas, however, it becomes possible to take linear combinations of the antenna outputs to cancel all crosstalk (*i.e.*, null-steering), and hence to retrieve each of the original signals. For this to be possible, it is necessary to have some more information, either on the structure of the array response matrix, or on the signals. In the wireless application, the antenna array is known in principle. However, there is often no direct line-of-sight, and the effect of multipath due to scattering off buildings causes a loss of structure of the actual array response matrix, and, as a result, standard direction-of-arrival estimation algorithms such as MVDR, MUSIC or ESPRIT are not applicable. The property that remains available is that the output signal of the beamformer, after successful null-steering, has a constant amplitude. With some simplifying assumptions, the CM factorization problem is an appropriate mathematical formulation of this 'blind null-steering' scenario: X is the received data matrix; its entries are n samples from m sensors; A is the array response matrix, and the rows $\mathbf{s}_k$ of S are the signals to be estimated.

SPIE 1994, "Advanced Signal Processing Algorithms, Architectures, and Implementations V", vol. 2296

For a long time, the CM factorization problem was considered to be too non-linear to admit a closed-form analytical solution, and only iterative, gradient-descent schemes have been developed, mostly based on pioneering work by Godard [1], Treichler and Agee [2], and Treichler and Larimore [4, 3]. The algorithms are known as CMAs. In its basic form, the algorithm consist of an LMS-type recursion,

$$\begin{array}{rcl} (\mathbf{s})_n & = & \mathbf{w}^{(n)}\mathbf{x}_n \\ \mathbf{w}^{(n+1)} & = & \mathbf{w}^{(n)} - \mu\{\,|(\mathbf{s})_n|^2 - 1\}(\mathbf{s})_n\mathbf{x}_n^*\,. \end{array} \tag{2}$$

($\mathbf{x}_n$ is the n-th column of X.) If $\mathbf{w}^{(n)}$ converges to a (non-zero) stationary value, then $\mathbf{w}X = \mathbf{s}$ is a CM signal. There are many variants of this scheme [4, 3, 5–9]. Despite many efforts, all CMAs up to date suffer from problems which seriously limit their practical and automated applicability, and which are basically a consequence of the non-convex optimization cost functions that all of them use. The most important deficiencies are (1) the cost functions have local minima, not corresponding to interference removing solutions and, depending on the initialization, algorithms do sometimes converge to these local minima [4, 3, 10–15]; (2) the speed of convergence is highly dependent on the initialization, but no suitable default initial points are known; (3) the algorithms search for a minimum and thus retrieve only one signal at a time; the other signals have to be found by starting from other initial points and hoping (or encouraging [7, 16, 17]) that they converge to something else. Weak signals are hard to retrieve in this way; (4) the only way to detect the number of CM signals is a posteriori, by counting the number of independent CM signals that have been obtained.

An analytic approach

The CM factorization $X = AS$ is not a standard linear algebra factorization, and it is not clear at first that solutions exist and are unique. Not much is known about this factorization. For large enough n, most matrices do not admit such a factorization: in general, solutions exist only if the matrix X was constructed from such a product in the first place. On the other hand, if n is too small, then the factorization is not well-determined: solutions S are possible which bear no relation to signals that generated X. Even if n is large enough, the solution, if it exists, can be unique only up to equivalence transformations $T : d \times d$ such that $X = AS = (AT^{-1})(TS)$ and TS is a CM matrix. Trivial examples of T, independent of A, S, are permutation matrices (reordering the signals) and unitary diagonal matrices (defining the initial phase of each signal). These trivial transformations cannot be avoided and are admissible. If X has a CM factorization and n is large enough ($n > d(d-1)$ is shown to be sufficient), then the factorization is usually unique, up to the above equivalences [18]. There are examples of matrices X for which the factorization is not unique, but this can occur only if there are specific phase relations between the signals. These relations are hard to analyze, but we may assume that they are generically not satisfied, especially if the signals are statistically independent or travel along independent paths.

We will assume, from now on, that $n > d^2$ is large enough so that there is a unique factorization, up to trivial transformations. The contributions of the paper are summarized as follows.

- It is possible to determine the number of CM signals present among other types of signals in X: δ equals the dimension of the kernel of a certain matrix constructed from X.
- W and S such that $WX = S$ is a CM matrix can be computed exactly, from a certain eigenvalue decomposition.
- With X distorted by additive noise, a generalization of the algorithm is robust in finding S, even when n is quite small. This is demonstrated with experimental data.

The algorithm is derived by setting up the equations for $\mathbf{w}$ to have $\mathbf{w}X \in \mathcal{CM}$ (section 2). These equations are quadratic in $\mathbf{w}$, but can be linearized when written in terms of the Kronecker product $\mathbf{w} \otimes \overline{\mathbf{w}}$, a vector with d^2 entries ($\overline{\mathbf{w}}$ is the complex conjugate of $\mathbf{w}$). If $n > d^2$, then the number of solution vectors to this linear system of equations indicates how many CM signals are present in X. In general, solutions to the linear system do not have the structure $\mathbf{w} \otimes \overline{\mathbf{w}}$: the system is underdetermined and there is an (affine) subspace of solutions. The core of the CM problem is to find those solutions that have a Kronecker structure. It is shown how a unitary transformation reduces this problem to a generalization of an eigenvalue problem: the simultaneous diagonalization of a number of matrices. Without noise, this problem has an essentially unique solution which can be found using standard linear algebra tools. With noise added to X, there is in general no exact diagonal solution, and we have to find an approximate simultaneous diagonalization. This is a challenging, non-standard linear algebra problem, for which we propose an algorithm that shows quadratic convergence (section 3). In section 4, the algorithm is tested on measured data sets.

2. EXACT SOLUTION TO THE CM PROBLEM

2.1. The Gerchberg-Saxton algorithm

Denote by $\mathcal{R}'(X)$ the subspace spanned by the rows of X (the co-range of X) and by $\mathcal{CM}$ the set of CM matrices. Assuming that there is a unique solution, the CM factorization problem is precisely equivalent to the following problem:

Problem P1. *Find all linearly independent signals* $\mathbf{s}$ *that satisfy*

$$\begin{array}{ll} (A) & \mathbf{s} \in \mathcal{CM}\,, \\ (B) & \mathbf{s} \in \mathcal{R}'(X)\,. \end{array}$$

From this formulation, it is straightforward to devise an algorithm based on alternating projections: start with a (random) choice of $\mathbf{s}$ in the row span of X, and alternatingly project it onto $\mathcal{CM}$ and back onto the row span of X. The set $\mathcal{CM}$ is not a linear subspace, so that the the projection onto $\mathcal{CM}$ is non-linear:

$$\mathbf{P}_{\mathcal{CM}}(\mathbf{y}) = \left[\frac{(\mathbf{y})_k}{|(\mathbf{y})_k|} \right]_{k=1}^{n} \,,$$

i.e., every entry of the vector is radially projected onto the complex unit circle. It is customary to estimate weight vectors $\mathbf{w}$ rather than signals, in which case the alternating projection algorithm is expressed as the iteration

$$\mathbf{w}^{(i+1)} = \left[\mathbf{P}_{\mathcal{CM}}(\mathbf{w}^{(i)}X)\right]X^{\dagger}\,. \tag{3}$$

(Note that $\mathbf{s}^{(i)} = \mathbf{w}^{(i)}X$, and $\cdot X^{\dagger}X = \mathbf{P}_{\mathcal{R}'(X)}(\,\cdot\,)$ represents a projection onto the row span of X.)

It turns out that this is a well-established algorithm in the field of optics for solving the phase-retrieval problem, where it is known as the Gerchberg-Saxton algorithm (GSA) [19]. The connection of the phase-retrieval problem with the CM problem was made only recently [18]. Essentially the same algorithm was derived from the CMA by Agee [6], and called the LSCMA. Simulation studies [6,18] comparing these methods to CMA and OCMA indicate a better performance, in the sense of a smaller residual distance of the resulting $\mathbf{s} \in \mathcal{R}'(X)$ to $\mathcal{CM}$. The other convergence properties are similar to the classical CMA schemes: converge is satisfactorily provided the initial choice for $\mathbf{w}$ is close to a true solution vector, it is possible that an initial guess converges extremely slowly, and not all stationary points correspond to CM signals. We will use this algorithm as a benchmark.

2.2. Kronecker product formulation

To derive an analytic solution to the CM factorization problem, we elaborate on problem P1: we are looking for all vectors $\mathbf{s}$ that are in the row span $\mathcal{R}'(X)$ of X and also have the CM property. In a series of steps, this problem is translated into an equivalent but more tractable form.

Let $X = U\Sigma V : U \in \mathbb{C}^{m\times m}, \Sigma \in \mathbb{R}^{m\times n}, V \in \mathbb{C}^{n\times n}$ be a singular value decomposition of X: U and V are unitary matrices, and Σ is a real diagonal matrix with non-negative entries. Suppose that $\mathrm{rank}(X) = d$. We can write

$$X = \hat{U}\hat{\Sigma}\hat{V}\,; \qquad \hat{U} \in \mathbb{C}^{m\times d},\ \hat{\Sigma} \in \mathbb{R}^{d\times d},\ \hat{V} \in \mathbb{C}^{d\times n}$$

where $\hat{U}$, $\hat{\Sigma}$, $\hat{V}$ are submatrices of U, Σ, V, respectively, corresponding to the non-zero singular values of X. In particular, $\mathcal{R}'(X) = \mathcal{R}'(\hat{V})$, and the rows of $\hat{V}$ form an orthonormal basis of the row span of X. Using $\hat{V}$, we can rewrite condition (B) in problem P1 as

$$(B): \quad \mathbf{s} \in \mathcal{R}'(X) \quad \Leftrightarrow \quad \mathbf{s} = \mathbf{w}\hat{V}\,, \qquad \hat{V} : d \times n\,.$$

Here, the weight vector $\mathbf{w}$ is not precisely the same as before: it is now acting on the orthogonal basis vectors of $\mathcal{R}'(X)$, rather than directly on X. This reduces the number of parameters to estimate from m to d, and ensures that linearly independent $\mathbf{w}$ result in linearly independent $\mathbf{s}$. To rewrite condition $(A): \mathbf{s} \in \mathcal{CM}$, put $\hat{V} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_n]$, where $\mathbf{v}_i \in \mathbb{C}^{d}$ is the i-th column in $\hat{V}$. Then

$$\begin{array}{rcl} (A): \quad \mathbf{s} = [\,(\mathbf{s})_1 \ \cdots \ (\mathbf{s})_n\,] \in \mathcal{CM} & \Leftrightarrow & [\,|(\mathbf{s})_1|^2 \ \cdots \ |(\mathbf{s})_n|^2\,] = [\,1 \ \cdots \ 1\,] \\ & \Leftrightarrow & \left\{ \begin{array}{rcl} \mathbf{w}\mathbf{v}_1\mathbf{v}_1^*\mathbf{w}^* & = & 1 \\ & \vdots & \\ \mathbf{w}\mathbf{v}_n\mathbf{v}_n^*\mathbf{w}^* & = & 1 \end{array} \right. \end{array}$$

Define $P_k = \mathbf{v}_k \mathbf{v}_k^* \in \mathbb{C}^{d\times d}$, for $k = 1,\cdots,n$. Then the CM problem is equivalent to finding all linearly independent vectors $\mathbf{w}$, such that

$$\mathbf{w} P_k \mathbf{w}^* = 1\,, \qquad k = 1,\cdots,n\,. \tag{4}$$

The n conditions on $\mathbf{w}$ are quadratic equations in the entries of $\mathbf{w}$, and are hard to solve directly. The approach is to expand these equations in the entries of $\mathbf{w}$, which gives rise to Kronecker products $\mathbf{w}\otimes\overline{\mathbf{w}}$. At this point, we introduce the notation, for $Y \in \mathbb{C}^{d\times d}$, $\mathbf{y} \in \mathbb{C}^{d^2}$,

$$\mathrm{vec}(Y) := \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{21} \\ \vdots \\ Y_{dd} \end{bmatrix}, \qquad \mathrm{vec}^{-1}(\mathbf{y}) := \begin{bmatrix} (\mathbf{y})_1 & (\mathbf{y})_2 & \cdots & (\mathbf{y})_d \\ (\mathbf{y})_{d+1} & (\mathbf{y})_{d+2} & \cdots & (\mathbf{y})_{2d} \\ \vdots & & \ddots & \vdots \\ (\mathbf{y})_{d^2-d+1} & & \cdots & (\mathbf{y})_{d^2} \end{bmatrix}.$$

With these definitions, the quadratic expression $\mathbf{w}P_k\mathbf{w}^*$ is 'linearized' as

$$\mathbf{w}P_k\mathbf{w}^* \;=\; \mathbf{p}_k\mathbf{y}\,, \quad \text{where} \quad \mathbf{y} = \mathrm{vec}(\mathbf{w}^*\mathbf{w}) = \mathbf{w}\otimes\overline{\mathbf{w}} \in \mathbb{C}^{d^2\times 1}, \quad \mathbf{p}_k = \mathrm{vec}(P_k)^T\,. \tag{5}$$

Collect the condition vectors $\mathbf{p}_k$ in one matrix P of size $n \times d^2$:

$$P = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_n \end{bmatrix}.$$

With the above definitions, the CM problem is now transformed into the following problem:*

Problem P2. *The CM problem is equivalent to finding all linearly independent vectors* $\mathbf{y}$ *satisfying*

$$P\mathbf{y} \;=\; \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{6}$$

$$\mathbf{y} \;=\; \mathbf{w}\otimes\overline{\mathbf{w}} \tag{7}$$

For each solution $\mathbf{w}$*, the corresponding CM signal is given by* $\mathbf{s} = \mathbf{w}\hat{V}$.

To prove that this problem is equivalent to solving equation (4), and hence problem P1, it remains to show that a set of solutions $\{\mathbf{y}\}_1^\delta$ is linearly independent if and only if the corresponding set $\{\mathbf{w}_k\}_1^\delta$ is linearly independent. This is straightforward.

The system of equations is overdetermined for large enough n, but still might have multiple solutions: any vector in the kernel of P can be added to any solution $\mathbf{y}$. In general, the solution to (6) can be written as an affine set of solutions. However, it is convenient at this point to do a linear transformation such that the solution set becomes linear. Let Q be any $n\times n$ unitary matrix such that

$$Q\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} n^{1/2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{8}$$

Simple choices for Q suffice; *e.g.*, Q could be a Householder transformation [21]. Apply Q to P:

$$QP =: \begin{bmatrix} \hat{\mathbf{p}}_1 \\ \hat{P} \end{bmatrix}, \qquad \begin{array}{ll} \hat{\mathbf{p}}_1: & 1\times d^2 \\ \hat{P}: & (n-1)\times d^2 \end{array}$$

With these definitions, we have

$$P\mathbf{y} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} (i) & \hat{\mathbf{p}}_1\mathbf{y} = n^{1/2} \\ (ii) & \hat{P}\mathbf{y} = 0 \end{cases}$$

*Similar equations have been derived by Jamali et al. [20]. They revert to an iterative method for solving this problem; we show here that there is an explicit solution.

One can show that $\hat{\mathbf{p}}_1\mathbf{y} = n^{1/2} \Leftrightarrow \mathrm{tr}(Y) = n$, where $Y = \mathrm{vec}^{-1}\mathbf{y}$ and $\mathrm{tr}(\,\cdot\,)$ is the trace operator. The proof of this is technical and uses the fact that P was generated from the isometry $\hat{V}$. When $\mathbf{y} = \mathbf{w} \otimes \overline{\mathbf{w}}$, then $\mathrm{tr}(Y) = \mathrm{tr}(\mathbf{w}^*\mathbf{w}) = \mathbf{w}\mathbf{w}^* = ||\mathbf{w}||^2$ so that $\hat{\mathbf{p}}_1\mathbf{y} = n^{1/2} \Leftrightarrow ||\mathbf{w}||^2 = n$. The CM problem is thus shown to be equivalent to the following problem.

Problem P3. *Let X be a given matrix. Construct $\hat{P} \in \mathbb{C}^{(n-1)\times d^2}$ from X. Find all linearly independent solutions $\mathbf{y}$ that satisfy*

$$\left\{\begin{array}{lccc} (i) & \hat{P}\mathbf{y} & = & 0 \\ (ii) & \mathbf{y} & = & \mathbf{w} \otimes \overline{\mathbf{w}} \\ (iii) & ||\mathbf{w}||^2 & = & n \end{array}\right.$$

For each solution $\mathbf{w}$, $\mathbf{s} = \mathbf{w}\hat{V}$ is a CM signal contained in X.

2.3. Detection of the number of CM signals

If the number of CM signals that are present in X is defined to be $\delta \leq d$, then there are δ linearly independent solutions $\mathbf{w}$ to the CM factorization problem, corresponding to δ linearly independent vectors $\mathbf{y} = \mathbf{w} \otimes \overline{\mathbf{w}}$ in the kernel of $\hat{P}$. Hence, it is necessary that $\dim\ker\hat{P} \geq \delta$. Since $\hat{P} : (n-1) \times d^2$, we also have $\dim\ker\hat{P} \geq d^2 - (n-1)$. To be able to detect δ from $\dim\ker\hat{P}$, we have to require that, at least, $n \geq d^2 + 1$.

Proposition 1. *Let $\delta \leq d$ be the number of CM signals in X, and let $n \geq d^2+1$. Then* $\dim\ker\hat{P} \geq \delta$. *Generically,* $\dim\ker\hat{P} = \delta$.

The dimension of the kernel will be larger than δ only if there are specific relations between the signals, valid for all points in time. It is not a trivial task to analyze these relations; however, in practical cases, signals are unrelated and signal paths are different, making it unlikely that the rank of the matrix $\hat{P}$ is any lower than necessary.

2.4. Computation of W

We will assume from now on that $\dim\ker\hat{P} = \delta$ is equal to the number of CM signals in X. Let $\mathbf{y}_1, \cdots, \mathbf{y}_\delta$ be δ non-zero and linearly independent solutions to $\hat{P}\mathbf{y} = 0$: $\{\mathbf{y}_k\}_1^\delta$ is a basis for the kernel of $\hat{P}$. It can be computed from an RQ factorization of $\hat{P}$, or, with more numerical accuracy, using an SVD. In the latter case, $\{\mathbf{y}_k\}_1^\delta$ are the right singular vectors corresponding to the δ singular values of $\hat{P}$ that are zero. Denote by $Y_k = \mathrm{vec}^{-1}(\mathbf{y}_k)$, $k = 1, \cdots, \delta$, the corresponding $d \times d$ matrices constructed from these vectors. We will use the property that $\mathbf{y}$ has a Kronecker structure, $\mathbf{y} = \mathbf{w} \otimes \overline{\mathbf{w}}$, if and only if $Y = \mathbf{w}^*\mathbf{w}$ is symmetric and of rank 1.

Because linear combinations of vectors in the kernel are also vectors in the kernel, we have in general that each each $\mathbf{y}_k$ does not have a Kronecker structure. But the existence of such vectors in the kernel implies that there are scalar parameters $\alpha_1, \cdots, \alpha_\delta$ such that

$$\alpha_1\mathbf{y}_1 + \cdots + \alpha_\delta\mathbf{y}_\delta = \mathbf{w} \otimes \overline{\mathbf{w}},$$

or equivalently, such that

$$\alpha_1 Y_1 + \cdots + \alpha_\delta Y_\delta = \mathbf{w}^*\mathbf{w}$$

is symmetric and of rank 1. Since there are δ independent solutions, there must be δ independent parameter vectors $[\alpha_1, \cdots, \alpha_\delta]$. The main step in the CMA algorithm is to find these parameter vectors.

It is hard to enforce a Kronecker product structure directly on the $\mathbf{y}_k$'s, but in terms of the Y_k's, this problem can be viewed as a kind of generalized eigenvalue problem. Indeed, if $d = \delta = 2$, then there are two matrices Y_1 and Y_2, each of size 2×2, and we have to find $\lambda = \alpha_2/\alpha_1$ such that $Y_1 + \lambda Y_2$ has its rank reduced by one (to become one). For larger δ, there are more than two matrices, and the rank should be reduced to one by taking linear combinations of all of them. This can be viewed as an extension of the generalized eigenvalue problem.

From the opposite perspective, suppose that the solutions of the CM problem are $\mathbf{w}_1, \cdots, \mathbf{w}_\delta$. We already said that $\mathbf{w}_1 \otimes \overline{\mathbf{w}}_1, \cdots, \mathbf{w}_\delta \otimes \overline{\mathbf{w}}_\delta$ is a set of linearly independent vectors; together they must span the kernel of $\hat{P}$. Moving to matrices, each of the matrices $Y_1, \cdots, Y_\delta$ is a different (independent) linear combination of $\mathbf{w}_1^*\mathbf{w}_1, \cdots, \mathbf{w}_\delta^*\mathbf{w}_\delta$, *i.e.*,

$$\begin{array}{rclcl} Y_1 & = & \lambda_{11}\mathbf{w}_1^*\mathbf{w}_1 + \lambda_{12}\mathbf{w}_2^*\mathbf{w}_2 + \cdots + \lambda_{1\delta}\mathbf{w}_\delta^*\mathbf{w}_\delta & = & W^*\Lambda_1 W \\ Y_2 & = & \lambda_{21}\mathbf{w}_1^*\mathbf{w}_1 + \lambda_{22}\mathbf{w}_2^*\mathbf{w}_2 + \cdots + \lambda_{2\delta}\mathbf{w}_\delta^*\mathbf{w}_\delta & = & W^*\Lambda_2 W \\ & \vdots & & & \\ Y_\delta & = & \lambda_{\delta 1}\mathbf{w}_1^*\mathbf{w}_1 + \lambda_{\delta 2}\mathbf{w}_2^*\mathbf{w}_2 + \cdots + \lambda_{\delta\delta}\mathbf{w}_\delta^*\mathbf{w}_\delta & = & W^*\Lambda_\delta W \end{array} \tag{9}$$

where

$$W = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_\delta \end{bmatrix}, \qquad \Lambda_k = \begin{bmatrix} \lambda_{k1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_{k\delta} \end{bmatrix} \qquad (k = 1, \cdots \delta).$$

Hence, by the existence of a solution to the CM problem, there must be a matrix W whose inverse simultaneously diagonalizes $Y_1, \cdots, Y_\delta$. The rows of W, scaled to have norm $n^{1/2}$, are the weight vectors that solve the CM problem.

Generically, Y_1 and Y_2 are $d \times d$ matrices of rank δ, and not less than δ. In this case, a generalized eigenvalue decomposition of just Y_1 and Y_2 will already determine W: there exist matrices M, N (invertible) such that

$$\begin{aligned} M^* Y_1 N &= \Lambda_1 \\ M^* Y_2 N &= \Lambda_2 \end{aligned}$$

where Λ_1, Λ_2 are diagonal matrices, size $d \times d$, each with δ non-zero entries. Reducing these matrices to $\delta \times \delta$ diagonal matrices with non-zero entries on the diagonal, and trimming M and N likewise to full rank $\delta \times d$ matrices, we obtain the decomposition

$$\begin{aligned} \hat{M}^* Y_1 \hat{N} &= \hat{\Lambda}_1 \\ \hat{M}^* Y_2 \hat{N} &= \hat{\Lambda}_2 . \end{aligned}$$

For the full rank case, $\hat{M}$ and $\hat{N}$ are unique up to equal permutations of their columns and (possibly different) right diagonal invertible factors. This uniqueness implies that, after a suitable diagonal scaling, we can arrange it such that $\hat{M} = \hat{N} = W^\dagger$, or $W = \hat{N}^\dagger$, with each row of W having norm $n^{1/2}$. For the case where Y_1 and Y_2 are not of rank δ, it is possible that they do not fully determine W, so that the other Y_k also have to be taken into account. It is obvious that it is possible to obtain W also in this case, but we omit the details of this more general procedure at this point. Numerically, it is better to take all Y_k into account in all cases. Such an algorithm is described in the next section.

Hence, we have shown that problem P3 can be solved analytically, via (9) and using eigenvalue decompositions. With W known, and properly scaled, the solution of the CM problem is $S = W\hat{V}$.

3. THE CM PROBLEM WITH ADDITIVE NOISE

3.1. Equivalent optimization problem

With noise added to the data, $X = AS + N$, an exact decomposition of X as $X = AS$ is no longer possible. To recover the CM signals, the problem can be posed as finding δ independent signals $\mathbf{s}$ that are minimizers of

$$\min\{\mathrm{dist}(\mathbf{s}, \mathcal{CM}) \mid \mathbf{s} \in \hat{\mathcal{R}}'(X)\}, \qquad \mathrm{dist}(\mathbf{s}, \mathcal{CM}) := \sum_{k=1}^{n} (\,|(\mathbf{s})_k|^2 - 1)^2 . \tag{10}$$

Here, $\hat{\mathcal{R}}'(X)$ is the estimated row span of S, which we will take to be the principal row span of X, as determined using an SVD. Thus let $X = U\Sigma V$ as before, and let d be the number of singular values of X that are significantly larger than zero. The rows of V corresponding to these singular values form an orthogonal basis of the principal row span of X, and are collected in the matrix $\hat{V}$. The matrix P and $\hat{P}$ are constructed from $\hat{V}$ as in section 2. We can express the cost function in equation (10) in terms of $\hat{P}$ and $\mathbf{w}$. Omitting the details, suffice it to say that the minimization problem is separable into the constrained minimization problem for $\mathbf{w}$,

$$\varepsilon^2 := \min \|\, \hat{P}(\mathbf{w} \otimes \overline{\mathbf{w}}) \,\|^2 \qquad \text{s.t. } \|\,\mathbf{w}\,\|^2 = n ,$$

which will provide the direction of $\mathbf{w}$, and the computation of a scalar c to minimize $\mathrm{dist}(c\mathbf{w}\hat{V}, \mathcal{CM})$. The corrective scaling can be expressed in term of ε, but it is close to 1 and of no importance in practice, as it will only scale the amplitude of the corresponding signal $\mathbf{s}$. Hence, with noise, we have to solve the following equivalent of problem P3:

Problem P4. *Find all independent minimizers* $\mathbf{y}$ *of*

$$(i) \quad \min_{\mathbf{y}} \|\, \hat{P}\mathbf{y} \,\|^2 , \tag{11}$$

such that (ii) $\mathbf{y} = \mathbf{w} \otimes \overline{\mathbf{w}}$, (iii) $\|\,\mathbf{w}\,\|^2 = n$. *Then the solution to the CM problem with noise is* $\mathbf{s} = \mathbf{w}\hat{V}$.

Minimizing (11) with the given conditions on $\mathbf{y}$ undoubtedly requires some iterative method, but the route set out by the solution of the noiseless case will provide excellent initial points for such a method. Thus, we compute a set of orthogonal vectors $\mathbf{y}_k$ that solve (11), construct the corresponding matrices Y_k, and subsequently impose the required structure on these matrices: linear combinations of the Y_k should result in matrices that are close to rank-1 symmetric matrices of the form $\mathbf{w}^*\mathbf{w}$. This modified problem is numerically feasible to solve, and for moderate noise levels, or large enough n, experiments show that the solution is quite close to the optimal solution of problem P4. We will now describe each of the steps in more detail.

For $n > d^2$, the number of CM signals δ can be estimated from the singular values of $\hat{P}$: it is equal to the number of singular values that are significantly smaller than the others. The corresponding right singular vectors form a basis $\{\mathbf{y}_k\}_1^\delta$ for the numerical kernel of $\hat{P}$. As before, the vectors $\mathbf{y}_1, \cdots, \mathbf{y}_\delta$ are reordered into matrices $Y_1, \cdots, Y_\delta$ in $\mathbb{C}^{d\times d}$. The next step is to decouple these matrices into rank-1 matrices: find independent parameter vectors $[\alpha_1 \cdots \alpha_\delta]$ such that

$$\alpha_1 Y_1 + \cdots + \alpha_\delta Y_\delta \simeq Y = \mathbf{w}^*\mathbf{w}, \tag{12}$$

i.e., a rank 1 symmetric matrix. With noise, there is no exact solution to this problem; we can only try to minimize the distance of this linear combination to the set of rank-1 symmetric matrices. Again, it is not clear how to solve such a problem exactly, but we will give a good approximate solution to it. In the noise-free case, the solution was given by a simultaneous diagonalization of $Y_1, \cdots, Y_\delta$. We will extend this approach to the noisy case.

3.2. Simultaneous diagonalization as a super-generalized Schur decomposition problem

Assume, for the moment, that there is no noise added to X. As we have seen, there exists a full rank matrix $W \in \mathbb{C}^{\delta\times d}$ such that

$$\begin{array}{rcl} Y_1 &=& W^*\Lambda_1 W \qquad\qquad (\Lambda_1, \cdots, \Lambda_\delta \in \mathbb{C}^{\delta\times\delta}, \text{ diagonal}) \\ Y_2 &=& W^*\Lambda_2 W \\ &\vdots& \\ Y_\delta &=& W^*\Lambda_\delta W\,. \end{array} \tag{13}$$

With noise, we can try to find $M = W^\dagger$ to simultaneously make $M^*Y_1M, \cdots, M^*Y_\delta M$ as much diagonal as possible. Because M is not unitary, the fact that it has to have full rank is hard to quantify, and it makes sense to rewrite this δ-generalized eigenvalue problem as a δ-generalized Schur decomposition. We first explain the procedure for the noise-free case. Bring in a QR factorization of W^* and an RQ decomposition of W: $W^* = Q^*R'$, $W = R''Z^*$, where Q, Z are unitary $d\times d$ matrices, and $R' \in \mathbb{C}^{d\times\delta}$, $R'' \in \mathbb{C}^{\delta\times d}$ are upper triangular. If $\delta < d$, then we can make sure that only the leading $\delta\times\delta$ blocks of R' and R'' are non-zero. Substitution into (13) leads to

$$\begin{array}{rcl} QY_1Z &=& R_1 \qquad\qquad (R_1, \cdots, R_\delta \in \mathbb{C}^{d\times d}, \text{ upper triangular}) \\ QY_2Z &=& R_2 \\ &\vdots& \\ QY_\delta Z &=& R_\delta \end{array} \tag{14}$$

with

$$R_1 = R'\Lambda_1 R'', \; R_2 = R'\Lambda_2 R'', \; \cdots, \; R_\delta = R'\Lambda_\delta R''. \tag{15}$$

Only the top-left $\delta\times\delta$ block of each R_k is non-zero. Hence, there exists Q, Z such that QY_kZ is upper triangular, for $k = 1, \cdots, \delta$, which is some kind of generalized Schur decomposition. With this decomposition, it is seen that a parameter vector $[\alpha_1 \cdots \alpha_\delta]$ satisfies (12) if and only if it satisfies

$$\alpha_1 R_1 + \cdots + \alpha_\delta R_\delta \quad \text{is rank 1.} \tag{16}$$

With the model of $R_1, \cdots, R_\delta$ in (15), we obtain

$$R'(\alpha_1\Lambda_1 + \cdots + \alpha_\delta\Lambda_\delta)R'' \quad \text{is rank 1.}$$

Since all the Λ_k are diagonal, the α_k are straightforward to compute: only one entry of the diagonal matrix $\alpha_1\Lambda_1 + \cdots + \alpha_\delta\Lambda_\delta$ can be non-zero. Setting this entry equal to one, all possible parameter vectors $[\alpha_1 \cdots \alpha_\delta]$ follow by constructing a matrix whose columns consist of the diagonal entries of the Λ_k,

$$\Lambda = \begin{bmatrix} (\Lambda_1)_{11} & \cdots & (\Lambda_1)_{\delta\delta} \\ \vdots & & \vdots \\ (\Lambda_\delta)_{11} & \cdots & (\Lambda_\delta)_{\delta\delta} \end{bmatrix}.$$

The rows of Λ^{-1} are the independent vectors $[\alpha_1 \ \cdots \ \alpha_\delta]$.

It is not necessary to compute the factorization (15): the parameters can be computed directly from the R_k in (14). A necessary condition for condition (16) to hold is that the resulting matrix has a main diagonal with at most one non-zero entry. In view of the existence of factorization (15) and the uniqueness of the CM factorization, this is sufficient, too.

Proposition 2. *For given $Y_1, \cdots, Y_\delta$, assume the decomposition (14). All independent parameter vectors $[\alpha_1 \ \cdots \ \alpha_\delta]$ such that $\alpha_1 Y_1 + \cdots + \alpha_\delta Y_\delta$ has rank 1 are given by the rows of A:*

$$A = R^{-1}, \qquad R = \begin{bmatrix} (R_1)_{11} & \cdots & (R_1)_{\delta\delta} \\ \vdots & & \vdots \\ (R_\delta)_{11} & \cdots & (R_\delta)_{\delta\delta} \end{bmatrix}.$$

Factoring each of the δ rank-1 matrices that is obtained in this way gives δ independent vectors $\mathbf{w}$, which form the rows of the matrix W that we were looking for in equation (13). Hence, in the noise-free case, the computation of a super-generalized Schur decomposition, *i.e.,* two unitary matrices Q, Z that satisfy (14), gives the solution to the simultaneous diagonalization problem. Although it seems at first sight that we have doubled the number of parameters to estimate (two matrices Q, Z, rather than one matrix W), this is not true: the fact that the matrices are unitary makes that the total number of parameters to estimate is precisely the same. However, the constraint that Q, Z be unitary is a desirable condition, whereas the fact that W must have full rank is difficult to handle.

We now return to the case where the data matrix X is distorted by noise. In this case, there is no Q, Z which simultaneously makes all matrices Y_k upper triangular. However, we can try to find Q, Z to make these matrices as much upper triangular as possible, by minimizing the Frobenius norm of the residual lower triangular entries. One approach for doing this is described in the next subsection. It is an extension to more than two matrices of the usual QZ iteration for computing the generalized Schur decomposition of two matrices. With Q, Z and hence $R_1, \cdots, R_\delta$ obtained this way, we can compute all independent parameter vectors $[\alpha_1 \cdots \alpha_\delta]$ as in proposition 2. The resulting matrices Y have approximately the form $Y = \mathbf{w}^* \mathbf{w}$, and each $\mathbf{w}$ can be estimated as the singular vector corresponding to the largest singular value of each Y. It remains to scale $\mathbf{w}$ to ensure that $||\mathbf{w}|| = n^{1/2}$.

The above scheme provides an approximate solution to problem P4. The algorithm is summarized in figure 1. It is not clear in what sense the solution approximates the optimal solution; however, it finds the exact solution if there is no noise, and yields perfect results for moderate noise levels or large enough n. For high noise levels, closely spaced signals, or small n, the vectors $\mathbf{w}$ that are obtained by the above procedure can be used as initial starting points in a Gerchberg-Saxton iteration, which effectively finds the minimum of P4. Since these starting points are accurate, a few iterations usually suffice, and independent signals are obtained for each $\mathbf{w}$. An example of the application of the algorithm to real data is given in section 4.

3.3. Super-generalized Schur decomposition

In this subsection, we describe a possible approach to the super-generalized Schur decomposition problem: for given matrices $Y_1, \cdots, Y_\delta$, find Q, Z (unitary) such that

$$QY_1Z = R_1, \quad \cdots, \quad QY_\delta Z = R_\delta,$$

where $R_1, \cdots, R_\delta$ are as much upper triangular as possible. As said in section 3.2, this problem is the core of the CM factorization problem: all the preceding steps just served to massage the problem into this form, and once the Schur decomposition is obtained, the solution of the CM problem follows directly from it.

For two matrices, upper triangular R_1, R_2 exist and are given by a standard generalized Schur decomposition. The extension to more than two matrices makes it a non-standard linear algebra problem and, as an extension to an eigenvalue problem, not easy to solve for general matrices Y_k. We rely on the fact that there *is* a good solution in the no-noise case (all matrices R_k upper triangular), and that the matrices Y_k do not abruptly change as the noise level is gradually increased. Our approach is to modify the standard QZ iteration method used for computing the Schur decomposition so that it works for more than two matrices. There is more than one way to do this. We will present a variant that has shown quadratic convergence in the no-noise case and treats all matrices $Y_1, \cdots Y_\delta$ equally.

The QZ iteration for computing the Schur decomposition of two matrices [21] starts with setting $Q^{(0)} = I$, $Z^{(0)} = I$. At the k-th iteration step, a unitary matrix $Q^{(k)}$ is computed such that $Q^{(k)}(Y_1 Z^{(k-1)})$ is upper triangular, and a unitary matrix $Z^{(k)}$ is computed to make $(Q^{(k)} Y_2) Z^{(k)}$ is upper triangular. As an extension to more than two matrices, we propose the following two step iteration.

Suppose $X = AS + N \in \mathbb{C}^{m\times n}$. An estimate of $S \in \mathcal{CM}$ is obtained as follows:
1. Compute SVD(X): $X =: U\Sigma V$
2. Estimate the number of signals d as the number of large entries in Σ
3. $\hat{V} =$ first d rows of V
4. Construct $\hat{P} : (n-1)\times d^2$ from $\hat{V}$:
   $\hat{V} =: [\mathbf{v}_1 \cdots \mathbf{v}_n]$
   $P = [\mathrm{vec}(\mathbf{v}_1\mathbf{v}_1^*) \ \cdots \ \mathrm{vec}(\mathbf{v}_n\mathbf{v}_n^*)]^T$
   $\begin{bmatrix} \hat{\mathbf{p}}_1 \\ \hat{P} \end{bmatrix} = QP$, with Q as in (8)
5. Compute SVD($\hat{P}$): $\hat{P} =: U_p\Sigma_p V_p^*$
6. Estimate the number of CM signals $\delta \le d$ as the number of small entries in Σ_p
7. $[\mathbf{y}_1 \ \cdots \ \mathbf{y}_\delta] :=$ last δ columns of V_p
8. $Y_1 = \mathrm{vec}^{-1}(\mathbf{y}_1), \cdots, Y_\delta = \mathrm{vec}^{-1}(\mathbf{y}_\delta)$
9. Find Q, Z to make $R_1 := QY_1Z, \cdots, R_\delta := QY_\delta Z$ approximately upper (section 3.3)
10. From $R_1, \cdots, R_\delta$, compute all vectors $[\alpha_{k1} \cdots \alpha_{k\delta}], k = 1, \cdots, \delta$
    s.t. $\hat{Y}_k := \alpha_{k1}Y_1 + \cdots + \alpha_{k\delta}Y_\delta$ is approximately rank 1 (proposition 2)
11. For each $\hat{Y}_k$:
    Compute $\mathbf{w}$ such that $\hat{Y}_k \simeq: \mathbf{w}_k^*\mathbf{w}_k$
    scale $\mathbf{w}_k$ such that $||\mathbf{w}_k|| = n^{1/2}$; (scale by c)
    $\mathbf{s}_k := \mathbf{w}_k\hat{V}$
    (perform a few Gerchberg-Saxton iterations, as in equation (3))

The vectors $\mathbf{s}_1, \cdots, \mathbf{s}_\delta$ are the rows of S.

Figure 1. CM factorization algorithm.

Denote by $||\cdot||_{LF}$ the Frobenius norm of the strictly lower triangular part of a matrix.

$$\begin{array}{ll} \text{Find } Q^{(k)} \text{ (unitary) to minimize} & ||\,Q^{(k)}(Y_1Z^{(k-1)})\,||_{LF}^2 + \cdots + ||\,Q^{(k)}(Y_\delta Z^{(k-1)})\,||_{LF}^2\,, \\ \text{find } Z^{(k)} \text{ (unitary) to minimize} & ||\,(Q^{(k)}Y_1)Z^{(k)}\,||_{LF}^2 + \cdots + ||\,(Q^{(k)}Y_\delta)Z^{(k)}\,||_{LF}^2\,. \end{array} \tag{17}$$

Each of these steps is a least squares problem with an exact solution. To describe this solution, suppose that, at the k-th stage, we have matrices $R_1 := Q^{(k-1)}Y_1Z^{(k-1)}, \cdots, R_\delta := Q^{(k-1)}Y_\delta Z^{(k-1)}$, and we have to find a unitary matrix Q that minimizes the below-diagonal norm of $QR_1, \cdots, QR_\delta$. Q is obtained as the product of $d-1$ more elementary unitary matrices

$$Q = \left[\begin{array}{c|c} I_{d-2} & 0 \\ \hline 0 & q_{d-1} \end{array}\right] \cdots \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & q_2 \end{array}\right] q_1\,.$$

The first factor, q_1, will simultaneously minimize the below-diagonal norms of only the first column of each of the matrices $R_1, \ldots, R_\delta$. Similarly, q_k is used to minimize the below-diagonal norm of the k-th columns of each of the matrices. Denote by $(R_1)_1$ the first column of R_1, and similarly for the other R_k's, then

$$q_1\,[(R_1)_1 \ \cdots \ (R_\delta)_1] = \left[\begin{array}{ccc} * & \cdots & * \\ \hline & E & \end{array}\right]$$

where $[*\cdots*]$ is the first row of the result, and E contains the remaining rows. The objective is to find q_1 such that $||\,E\,||_F$ is minimized. The solution is not unique, but a possible q_1 follows directly from an SVD:

$$[(R_1)_1 \ \cdots \ (R_\delta)_1] =: U\Sigma V^* \quad \Rightarrow \quad q_1 = U^*\,.$$

Indeed, for this choice of q_1, we have $||\,E\,||_F^2 = \sigma_2^2 + \cdots \sigma_\delta^2$, which is as small as we can hope for. After q_1 has been computed and applied to $R_1, \cdots R_\delta$, we have obtained new matrices $R_1', \cdots, R_\delta'$. The next factor, q_2, is used to minimize the below-diagonal norm of the second columns of these matrices. As q_2 is unitary and the first rows of $R_1', \cdots, R_\delta'$ stay unchanged, this will not change the

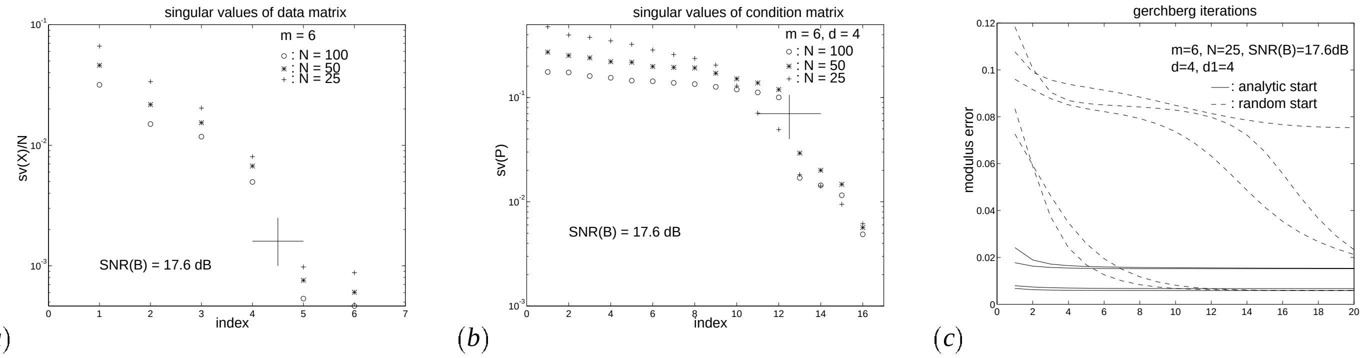

Figure 2. Experiment with $d = 4$ FM broadcasters and an array of $m = 6$ receiving antenna's. (a) svd(X), (b) svd$(\hat{P})$, (c) dist$(\mathbf{s}, \mathcal{CM})/n$ during Gerchberg iterations, with analytically computed and random initial starts.

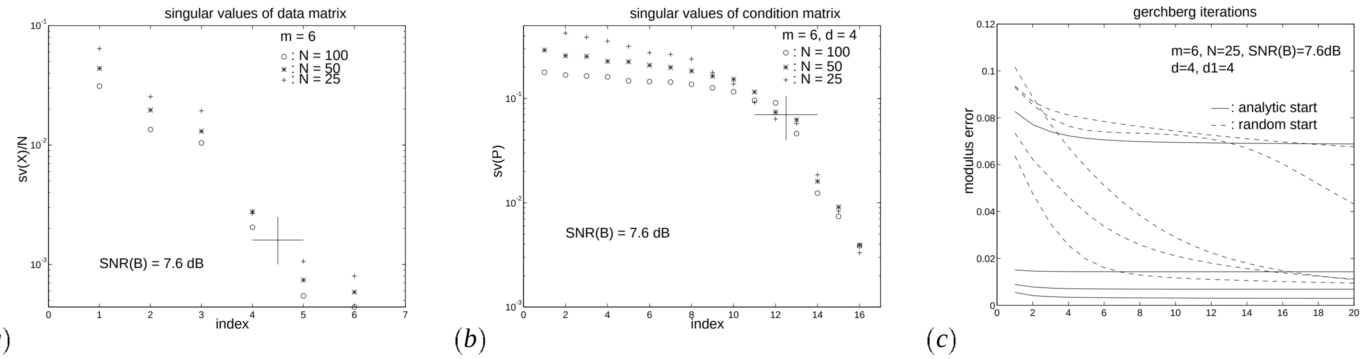

Figure 3. Same experiment as in figure 2, but now with SNR(B) lowered to 7.6 dB.

below-diagonal norm of the first columns. In fact, q_2 can be found in precisely the same way as q_1 by looking at the reduced problem where we act on $R'_1, \cdots, R'_\delta$ with their first rows and columns removed. The matrices $q_3, \cdots, q_{d-1}$ follow in turn.

The resulting QZ iteration (17) is observed to converge fast, usually quadratically in 3–5 iterations. Because the inner loop consist of SVDs, the scheme is only practical if d is small, which is certainly the case for the currently envisioned applications ($d \leq 6$, say).

4. EXPERIMENTAL EVALUATION

To assess the performance of the algorithm, we have applied it to a number of test matrices, based both on computer generated data and on real data collected from an experimental set-up. The results are quite convincing. For example, the algorithm could separate four CM signals using 4 sensors and only 17 data samples, sometimes even if each signal has a signal to background noise ratio of 5 dB. In this paper, we will report on only one such example, using measurement data collected from a rooftop antenna set-up at ArgoSystems, Inc (Sunnyvale, CA). In this experiment, there are $m = 6$ receiver antenna's, arranged in a certain nondescript pattern, and $d = 4$ FM transmitters, marked *A*–*D*, broadcasting music, speech, and modulated tones. The angles of the transmitters with respect to the array broadside were $-1.5^\circ, 0^\circ, 7.1^\circ, 42.6^\circ$, for A, B, C, D, respectively, and their signal-to-background noise levels were 19.1 dB, 17.6 dB, 17.9 dB, 16.7 dB. In a second experiment, the power of B was lowered to 7.6 dB.

In figure 2 (a) and (b), the singular values of X and $\hat{P}$ are shown. For $n = 100$ and $n = 50$, it is clear that there are 4 signals, and that in this example all of them have constant modulus. Figure 2(c) shows the Gerchberg iterations that result when we take $d = 4$ and $\delta = 4$, for $n = 25$ samples. The solid lines is the modulus error that result when the iterations are started from the $\delta = 4$ analytically computed weight vectors, the dashed lines is the error when we start with a number of randomly selected weight vectors. The Gerchberg iteration improve the analytically computed weight vectors only marginally: they are already close to optimal. For $n = 50$, straight lines occurred (not shown).

In a second experiment, the power of signal B was lowered to 7.6 dB. As the spacing of the B-antenna to the A-antenna is still only

$1.5°$, this is a challenging test of the algorithm. The results are depicted in 3. The detection of the other three signals from the singular values of $\hat{P}$ remained the same, but the fourth singular value (apparently corresponding to B) is raised and now somewhere in the middle of the gap between the large and small singular values. The detection that there are four independent signals from the singular values of X is also more difficult now, even if $n = 100$.

The main conclusion to draw from these and a large number of other experiments is that in all observed cases the algorithm obtains the optimum of the minimization problem if the number of samples n is sufficiently large. For four signals, $n = 50$–100 is typically large enough, even under severe conditions. For smaller n, the estimates move away from their optimal values, but usually, the algorithm still finds all CM signals if their number has been estimated correctly, and the optima can be obtained by adding a few iterations of the Gerchberg-Saxton algorithm as postprocessing. The effect of a smaller n is mostly felt in a closing of the gap between the larger and smaller singular values of $\hat{P}$, which limits the detection of δ. This is mitigated to some extent by the property that the algorithm is quite robust when d or δ are overestimated.

5. CONCLUDING REMARKS

In this paper, we have described an analytic method for solving the constant modulus factorization problem. The method condenses all conditions on the weight vectors $\mathbf{w}$ into a single matrix $\hat{P}$, and finds all independent vectors in the kernel of this matrix that have a Kronecker product structure. This problem, in turn, is shown to be a super-generalized eigenvalue problem, which may be formulated in terms of a super-generalized Schur decomposition: for given matrices $Y_1, \cdots, Y_\delta$, find Q, Z (unitary) such that

$$\begin{aligned} QY_1Z &= R_1 \\ &\vdots \\ QY_\delta Z &= R_\delta \end{aligned}$$

where $R_1, \cdots, R_\delta$ are as much upper triangular as possible. We have proposed a modified QZ iteration which treats all Y_k equally, converges to upper triangular matrices R_k in the absence of noise, and usually has quadratic convergence. Other interesting iterations might be devised as well.

The analytic algorithm is definitely more complex than the usual iterative approaches for blind beamforming and blind deconvolution of constant modulus signals. However, it gives fundamental solutions to a number of problems that have plagued iterative CM algorithms ever since their inception in the early 1980s. The most important advantages of the analytic approach are

1. It is less blind: the number of CM signals can be detected beforehand, from the close-to-zero singular values of $\hat{P}$. Not all signals have to be CM signals.
2. It is deterministic: it does not rely on lucky initial choices of $\mathbf{w}$. All CM signals are found. It does not lock on other signals (local minima). The only parameters that have to be set are the total number of signals, and the number of CM signals.

These two properties make the algorithm more reliable, so that it can operate with a lower number of samples and at a lower SNR. The number of samples is an important issue in applications where multipath causes fast fading, such as in wireless communications in cities. A lower SNR requirement translates into lower transmitting powers of the mobiles, *i.e.*, longer battery life.

The underlying reason that the analytic method can perform better seems to be the fact that it translates all conditions on the signals into conditions on the weight vectors $\mathbf{w}$. These are used in setting up a generalized eigenvalue problem, and as more samples are taken into account, the problem becomes better conditioned. The dimensionality of this problem stays constant as more samples are taken into account. In contrast, the iterative methods try to estimate the full signal $\mathbf{s}$, which sits in a high-dimensional space of increasing dimensionality, and the translation of the requirement that $\mathbf{s}$ should be a constant modulus signal into conditions on $\mathbf{w}$ is done only weakly. As a result, the convergence performance of these algorithms does not improve much if the number of samples is increased, other than that the averaged final error becomes smaller.

ACKNOWLEDGEMENTS

This research was supported in part by by ARPA, contract no. F49620-91-C-0086, monitored by the AFOSR. We are grateful to F. McCarthy of ArgoSystems, Inc, for providing us a collection of measurement data.

REFERENCES

[1] D.N. Godard, “Self-recovering equalization and carrier tracking in two-dimensional data communication systems,” *IEEE Tr. Communications*, vol. 28, pp. 1867–1875, Nov. 1980.

[2] J.R. Treichler and B.G. Agee, “A new approach to multipath correction of constant modulus signals,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 459–471, Apr. 1983.

[3] J.R. Treichler and M.G. Larimore, “New processing techniques based on constant modulus adaptive algorithm,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, Apr. 1985.

[4] M.G. Larimore and J.R. Treichler, “Convergence behavior of the constant modulus algorithm,” in *ICASSP-83*, pp. 13–16 vol.1, IEEE, 1983.

[5] R. Gooch and J. Lundell, “Convergence rate limitations in certain frequency-domain adaptive filters,” in *ICASSP-86*, (Tokyo), pp. 2523–2526, IEEE, 1986.

[6] B.G. Agee, “The least-squares CMA: A new technique for rapid correction of constant modulus signals,” in *ICASSP-86*, (Tokyo), pp. 953–956, IEEE, 1986.

[7] B.G. Agee, “Fast adaptive polarization control using the least-squares constant modulus algorithm,” in *Asilomar Conf. on Signals, Systems and Computers*, pp. 590–595, 1987.

[8] K. Hilal and P. Duhamel, “A convergence study of the constant modulus algorithm leading to a normalized-CMA and a block-normalized-CMA,” in *Signal Processing VI. Proc. EUSIPCO-92* (J. Vandewalle e.a., ed.), pp. 135–138 vol.1, Elsevier, 1992.

[9] R. Pickholtz and K. Elbarbary, “The recursive constant modulus algorithm: A new approach for real-time array processing,” in *Asilomar Conf. on Signals, Systems and Computers*, 1993.

[10] C.R. Johnson, Jr., S. Dasgupta, and W.A. Sethares, “Averaging analysis of local stability of a real constant modulus algorithm adaptive filter,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, pp. 900–910, Apr. 1988.

[11] C.K. Chan and J.J. Shynk, “Stationary points of the constant modulus algorithm for real Gaussian signals,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 38, pp. 2176–2181, Dec. 1990.

[12] J.R. Treichler, V. Wolff, and C.R. Johnson, Jr., “Observed misconvergence in the constant modulus adaptive algorithm,” in *Asilomar Conf. on Signals, Systems and Computers*, pp. 663–667 vol.2, 1991.

[13] Z. Ding, R.A. Kennedy, B.D.O. Anderson, and C.R. Johnson, Jr., “Ill-convergence of Godard blind equalizers in data communication systems,” *IEEE Trans. Comm.*, vol. 39, pp. 1313–1327, Sept. 1991.

[14] Z. Ding and C.R. Johnson, Jr., “On the nonvanishing stability of undesirable equilibria for FIR Godard blind equalizers,” *IEEE Trans. Signal Processing*, vol. 41, pp. 1940–1944, May 1993.

[15] J.J. Shynk and C.K. Chan, “Performance surfaces of the constant modulus algorithm based on a conditional Gaussian model,” *IEEE Trans. Signal Processing*, vol. 41, pp. 1965–1969, May 1993.

[16] B.G. Agee, “Blind separation and capture of communication signals using a multitarget constant modulus beamformer,” in *MILCOM 89*, pp. 340–346 vol.2, IEEE, 1989.

[17] F. McCarthy, “Multiple signal direction-finding and interference reduction techniques,” in *Wescon/93 Conf. Record*, (San Francisco, CA), pp. 354–361, Sept. 1993.

[18] Y. Wang, Y.C. Pati, Y.M. Cho, A. Paulraj, and T. Kailath, “A matrix factorization approach to signal copy of constant modulus signals arriving at an antenna array,” in *Proc. CISS*, (Princeton), 1994.

[19] R.W. Gerchberg and W.O. Saxton, “A practical algorithm for the determination of phase from image and diffraction plane pictures,” *Optik*, vol. 35, pp. 237–246, 1972.

[20] H. Jamali and S.L. Wood, “Error surface analysis for the complex constant modulus adaptive algorithm,” in *Asilomar Conf. on Signals, Systems and Computers*, pp. 248–252 vol.1, 1990.

[21] G. Golub and C.F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1989.