

Domains of Attraction of Shalvi–Weinstein Receivers

Ming Gu, *Member, IEEE*, and Lang Tong, *Member, IEEE*

Abstract—Domains of attraction (DoA) of Shalvi–Weinstein (SW) receivers are analyzed. It is shown that there is a one-to-one correspondence between DoA in the receiver parameter space and those in the global (or combined channel-receiver) parameter space. For general noiseless channels, DoA of SW receivers in the global response space are the minimum distance decision regions on a unit sphere. In the presence of noise and for the class of orthogonal channels, DoA of SW receivers for independent and identically distributed (i.i.d.) input signals are the minimum distance decision regions on an ellipsoid determined by the channel coefficients and the noise variance. The DoA in the receiver parameter space are also characterized for the general nonuniformly distributed sources. The size of the DoA is shown to be affected by the signal power, the signal constellation, the noise level, and the channel condition. It is also demonstrated that although the optima of the Shalvi–Weinstein algorithm and those of the constant modulus algorithm are one-to-one correspondent, their DoA are different in general.

Index Terms—Adaptive filters, blind equalization and deconvolution.

I. INTRODUCTION

THE goal of blind signal estimation is to estimate input signals without knowing the channel. Perhaps the most popular blind equalizer design criteria are the constant modulus (CM) cost [9], [18] that measures the dispersion of the estimate around a constant, and the Shalvi–Weinstein (SW) cost [15] that evaluates the fourth-order cumulant under a power constraint. It has been shown [13], [17] that, for sub-Gaussian sources,¹ the minima of the CM criterion and the maxima of the SW criterion are equivalent in the sense that they differ only by scaling factors. In other words, both criteria lead to receivers with the same unbiased mean square error.

The optimization of CM and SW criteria can be accomplished by the gradient method that guarantees the convergence to a local optimum. The implementation of the gradient maximization algorithm has the form

$$\mathbf{f}_{k+1} = \mathbf{f}_k + \mu \mathbf{D}_k \nabla J(\mathbf{f}_k), \quad k = 0, 1, 2, \dots \quad (1)$$

where

μ positive step-size;

Manuscript received June 21, 1999; revised March 26, 2001. This work was supported in part by the National Science Foundation under Contract CCR-9804019 and the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dr. Ir. Bart L. R. De Moor.

M. Gu is with Voyan Technology, Santa Clara, CA 95054 USA (e-mail: mgu@ieee.org).

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ee.cornell.edu).

Publisher Item Identifier S 1053-587X(01)05182-0.

¹Sub-Gaussian sources have the negative fourth-order cumulants.

\mathbf{D}_k positive definite symmetric matrix ($\mathbf{D}_k = \mathbf{I}$ resulting in the steepest ascent method);

J continuously differentiable function.

Given a particular initialization \mathbf{f}_0 , for arbitrarily small step-size μ , the above update approximates the search of optima on the surface of the cost function, and \mathbf{f}_k converges to a local optimum \mathbf{f}^* . The domain of attraction (DoA)² associated with \mathbf{f}^* is a neighborhood of \mathbf{f}^* such that the gradient algorithm initialized with any element in the neighborhood leads to the convergence to \mathbf{f}^* . In this paper, we are concerned with the largest such neighborhood.

Although considerable progress has been made in characterizing the locations of CM and SW receivers defined as local optima of their corresponding cost functions [7], [10], [13], [20], the following questions remain unanswered: i) Given an initialization of the algorithm, where will it converge? ii) How do we initialize the algorithm so that a receiver will converge to the desired setting? The lack of definitive answers to these questions is directly related to the lack of characterization of the domains of attraction for CM and SW receivers.

The purpose of this paper is to analyze domains of attraction for the gradient ascent implementation of the SW cost function. In particular, we are interested in the trajectories of the ordinary differential equation associated with the gradient of the SW (CM) objective function. Although the analysis is performed with respect to the statistical average of the SW cost, using stochastic approximation [3], it allows us to predict the behavior of the stochastic gradient algorithm implemented in practice.

The main results of this paper are as follows. Using the signal space property, we first show that under the assumptions given in Section II-A, the analysis of DoA of the receiver parameter can be pursued in the global response space without loss of generality. Analytical results for two special cases are thus obtained next. In the absence of noise, we show that DoA are the minimum distance decision regions on a unit ball, and the relative size of the domain of attraction for each receiver remains unchanged as long as signal constellations are fixed. In the presence of noise, we show that for the class of orthogonal channels and when the input signals have the same constellation, DoA of SW receivers are the minimum distance decision regions on an ellipsoid determined by the channel, the noise, and the signal power. This result is also extended to the nonuniformly distributed sources case. DoA in the receiver parameter space are defined accordingly. For the orthogonal channels, we further investigate how the relative size of DoA is influenced by the channel condition, noise variance, signal power, and signal constellation. This analysis is of great practical significance for the blind signal separation problem when there are

²We use DoA as an abbreviation for both “domain of attraction” and “domains of attraction.” Which one it represents can be determined from the context.

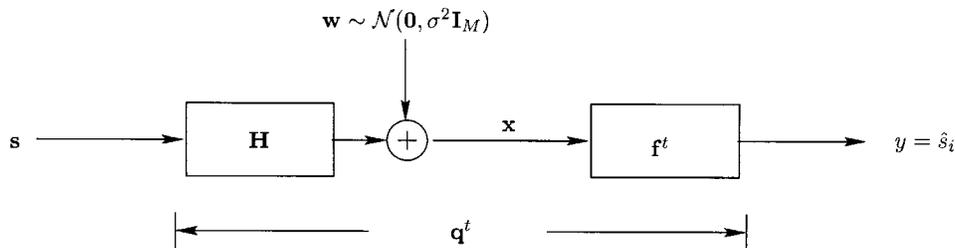


Fig. 1. Linear estimation model.

multiple users with different types of signals. Finally, we give an example of DoA for the two-dimensional (2-D) global response. In this case, the characterization of SW receivers is complete. It is interesting to note that in spite of the one-to-one correspondent relationship existing between SW and CM receivers, their DoA are different in general.

The two scenarios—general channels without noise and orthogonal channels with arbitrary noise—that DoA can be completely characterized are somewhat restrictive. However, they do include many practical applications in communications. Because the SW cost is smooth, the characterization of DoA obtained in this paper approximates well for general channels at high signal-to-noise ratio (SNR) and for channels that are approximately orthogonal. The latter may be the result of the transmission of code division multiple access (CDMA) signals using orthogonal codes. In the presence of channel dispersion, the signal waveforms from different users are no longer orthogonal. It has been shown that using SW receivers for such a case leads to minimum mean square error (MMSE)-like performance [17]. The analysis presented here offers an approximate description of DoA for mildly distorted channels. Another example that involves orthogonal channels is space-time coding. Tarokh *et al.* [16] showed that by constructing space-time block codes with columns being orthogonal, decoupled maximum likelihood decoding is achieved with remarkable performance at the expense of almost no extra processing. Other examples include orthogonal frequency division multiplexing (OFDM) used in wireless communications [19] and the discrete multitone (DMT) implemented for digital subscriber line (DSL) technology [11]. A review on applications of orthogonal transmultiplexers in communications was presented in [2].

There are few results available on the characterization of attraction domains for SW or CM receivers. The most relevant is a paper by Benveniste *et al.* [1], where they considered a class of objective functions that include the Sato algorithm as a special case and characterized DoA in the absence of noise for the blind equalization problem. Our results differ from theirs in two ways. First, their analysis does not apply directly to smooth cost functions such as the SW criterion. Second, we have included several parameters in the model ignored in [1]. Specifically, we have included noise and considered sources with different power and statistical properties. It is interesting, however, to note that some of our results coincide with theirs, even when the cost functions considered are different. A recent study of the DoA for fractionally spaced CM equalizers was made by Chung and Johnson [6]. They showed that under the noise-free assumption, the domain of attraction of a local CM minimum in the global re-

sponse space is an exponential hypercone that, when the power constraint is applied, agrees with our result of the minimum distance decision region on the unit ball. More general results, including the situation in the presence of noise, were also considered in Chung's thesis [5].

In the following section, the linear equalization model is described. The signal space property and the equivalent cost functions in the global response space are presented in Section II as well. Main results with regard to domains of attraction of SW receivers and their properties for both the noiseless case and the orthogonal channels are given in Section III. Conclusion is drawn in Section IV.

Throughout the paper, upper and lower case bold letters denote matrices and vectors, respectively. Key symbols are listed as follows.

$(\cdot)^t$	transpose;
$(\cdot)^\dagger$	pseudo-inverse;
$E\{\cdot\}$	expectation operator;
$\ \mathbf{x}\ _p$	l_p -norm of \mathbf{x} defined by $(\sum_i x_i ^p)^{1/p}$;
$\ \mathbf{x}\ _{\mathbf{A}}$	vector norm of \mathbf{x} on matrix \mathbf{A} defined by $\sqrt{\mathbf{x}^t \mathbf{A} \mathbf{x}}$;
$\mathcal{C}_{\mathbf{A}}$	range space of $\mathbf{A} \mathbf{A}^\dagger$ [12];
$\mathcal{C}_{\mathbf{A}^\perp}$	range of $\mathbf{I} - \mathbf{A} \mathbf{A}^\dagger$;
\mathbf{I}	identity matrix;
\mathbf{e}_i	unit column vector with 1 at the i th entry and zero elsewhere;
$ \mathcal{S} $	size of the set \mathcal{S} .

II. PROBLEM FORMULATION

A. Model

We consider the linear estimation model illustrated in Fig. 1 and described by

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w}, \quad (2)$$

$$y = \mathbf{f}^t \mathbf{x} = \mathbf{q}^t \mathbf{s} + \mathbf{f}^t \mathbf{w} \quad (3)$$

where the input signal $\mathbf{s} \in \mathcal{R}^N$ is transmitted through the channel $\mathbf{H} \in \mathcal{R}^{M \times N}$ (the impulse response matrix) and corrupted by the additive white Gaussian noise $\mathbf{w} \in \mathcal{R}^M$.

The linear estimator $\mathbf{f} \in \mathcal{R}^M$ is designed to estimate one of the components in \mathbf{s} , say s_i . The estimation is blind in the sense that only the marginal statistics of \mathbf{x} and \mathbf{s} are used, and the channel matrix \mathbf{H} is assumed to be unknown. Vector $\mathbf{q} \triangleq \mathbf{H}^t \mathbf{f}$ represents the global response between the signal to be estimated (s_i) and the estimator output y .

In our analysis, we assume the following.

- A1) Entries of \mathbf{s} are independent sub-Gaussian random variables.

- A2) Entries of \mathbf{w} are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance σ^2 .
- A3) \mathbf{s} and \mathbf{w} are independent.
- A4) $\mathbf{H} \in \mathcal{R}^{M \times N}$ has full column rank.

Because \mathbf{H} is unknown, it is without loss of generality to assume that $E\{\mathbf{ss}^t\} = \mathbf{I}$. If source signals are unequally powered, the input vector can be expressed as $\mathbf{E}\mathbf{s}$, where $\mathbf{E} = \text{diag}(\sqrt{\mathcal{E}_1}, \dots, \sqrt{\mathcal{E}_N})$ with \mathcal{E}_i being the power of the i th user. Since both \mathbf{H} and \mathbf{E} are unknown, we can absorb \mathbf{E} into \mathbf{H} , and thus, $E\{\mathbf{ss}^t\} = \mathbf{I}$ still holds true. The effect of signal power will be discussed later.

B. Cost Functions and Implementations

SW Cost and SW Algorithm: In order to estimate the input signal without knowing the channel, Shalvi and Weinstein proposed a criterion [15] to maximize the cost function

$$\bar{J}_s(\mathbf{f}) \triangleq |K_4(y)| \quad (4)$$

subject to the unit power constraint $E\{|y|^2\} = 1$, where $K_4(y)$ is the fourth-order cumulant of y . A local maximum of this constrained optimization is referred to as an SW receiver. To obtain an explicit form of $\bar{J}_s(\mathbf{f})$, by substituting (3) into (4), we have

$$\begin{aligned} \bar{J}_s(\mathbf{f}) &= |K_4(\mathbf{f}^t \mathbf{H} \mathbf{s} + \mathbf{f}^t \mathbf{w})| = |K_4(\mathbf{f}^t \mathbf{H} \mathbf{s}) + K_4(\mathbf{f}^t \mathbf{w})| \\ &= |K_4(\mathbf{f}^t \mathbf{H} \mathbf{s})| \\ &= \left| K_4 \left(\sum_{i=1}^N \left(\sum_{j=1}^M f_j h_{ji} \right) s_i \right) \right| \\ &= \left| \sum_{i=1}^N \left(\sum_{j=1}^M f_j h_{ji} \right)^4 K_4(s_i) \right| \\ &= \sum_{i=1}^N \left(\sum_{j=1}^M f_j h_{ji} \right)^4 \pi_i = \|\mathbf{\Pi}^{1/4} \mathbf{H}^t \mathbf{f}\|_4^4 \end{aligned} \quad (5)$$

where³

$$\mathbf{\Pi} = \text{diag}(\pi_1, \dots, \pi_N), \quad \pi_i = -K_4(s_i). \quad (6)$$

The above derivation makes use of some properties of cumulant [14] and assumptions A1)-A3) listed in the previous subsection. As for the constraint, applying $E\{\mathbf{ss}^t\} = \mathbf{I}$, $E\{\mathbf{ww}^t\} = \sigma^2 \mathbf{I}$ and $E\{\mathbf{sw}^t\} = \mathbf{0}$ leads to

$$\begin{aligned} E\{|y|^2\} &= E\{\mathbf{f}^t \mathbf{H} \mathbf{s} \mathbf{s}^t \mathbf{H}^t \mathbf{f}\} + E\{\mathbf{f}^t \mathbf{w} \mathbf{w}^t \mathbf{f}\} \\ &= \mathbf{f}^t (\mathbf{H} \mathbf{H}^t + \sigma^2 \mathbf{I}) \mathbf{f} = \|\mathbf{f}\|_{\mathbf{R}}^2 \end{aligned} \quad (7)$$

where

$$\mathbf{R} \triangleq E\{\mathbf{xx}^t\} = \mathbf{H} \mathbf{H}^t + \sigma^2 \mathbf{I}. \quad (8)$$

Therefore, the SW receiver can be obtained from

$$\mathbf{f}_s = \arg \max_{\mathbf{f}} \|\mathbf{\Pi}^{1/4} \mathbf{H}^t \mathbf{f}\|_4^4 \text{ subject to } \|\mathbf{f}\|_{\mathbf{R}}^2 = 1. \quad (9)$$

³For a sub-Gaussian signal s , $K_4(s) < 0$. The negative sign in this definition is meant to make π_i positive.

Two stochastic gradient algorithms were proposed in [15] for practical implementation.

CM Cost and CM Algorithm: Proposed earlier than the SW algorithm (SWA), the constant modulus criterion [9] aims to minimize an unconstrained objective function

$$\bar{J}_c(\mathbf{f}) \triangleq E\{(|y|^2 - r)^2\} \quad (10)$$

where $r \triangleq (E\{|s_i|^4\}) / (E\{|s_i|^2\})$ is the dispersion constant. Local minima of (10) constitute CM receivers. The explicit expression of the CM cost is obtained as

$$\begin{aligned} \bar{J}_c(\mathbf{f}) &= E\{y^4\} - 2rE\{y^2\} + r^2 \\ &= K_4(y) + 3E^2\{y^2\} - 2rE\{y^2\} + r^2 \\ &= - \underbrace{\|\mathbf{\Pi}^{1/4} \mathbf{H}^t \mathbf{f}\|_4^4 + 3\|\mathbf{f}\|_{\mathbf{R}}^4 - 2r\|\mathbf{f}\|_{\mathbf{R}}^2}_{\bar{J}_s(\mathbf{f})} + r^2. \end{aligned} \quad (11)$$

The connection of CM and SW receivers is obvious if we impose a unit power constraint $\|\mathbf{f}\|_{\mathbf{R}}^2 = 1$ on (11).

The constant modulus algorithm (CMA) is the stochastic gradient method based on the cost function (10) and given by

$$\mathbf{f}_{k+1} = \mathbf{f}_k - \mu \mathbf{x}_k y_k (|y_k|^2 - r), \quad \mu > 0. \quad (12)$$

Compared with the (constrained) SWA, this algorithm is easier to implement.

Power-Constrained CMA: The analysis of DoA for SW receivers can be more conveniently carried out by considering the power-constrained constant modulus (PC-CM) algorithm with receivers defined as

$$\mathbf{f}_p \triangleq \arg \min_{E\{|y|^2\}=1} E\{(|y|^2 - r)^2\} = \arg \min_{\|\mathbf{f}\|_{\mathbf{R}}^2=1} \bar{J}_c(\mathbf{f}). \quad (13)$$

Examining (5) and (11), it can be seen that under the constraint $\|\mathbf{f}\|_{\mathbf{R}}^2 = 1$, $\bar{J}_s(\mathbf{f})$ and $\bar{J}_c(\mathbf{f})$ have exactly the same form except for a sign difference, which implies that the minimization of the power-constrained CM cost is identical to the maximization of the SW cost. Therefore, the domain of attraction of an SW receiver is the same as that of the corresponding receiver obtained by a gradient search algorithm for the criterion described in (13).

In optimizing (13), we consider a two-step stochastic approximation procedure

$$\bar{\mathbf{f}}_{k+1} = \mathbf{f}_k - \mu \mathbf{x}_k y_k (|y_k|^2 - r), \quad \mathbf{f}_{k+1} = \frac{\bar{\mathbf{f}}_{k+1}}{\|\bar{\mathbf{f}}_{k+1}\|_{\mathbf{R}}} \quad (14)$$

where $\mu > 0$ is a step size. Obviously, the first step aims to attain a CM minimum which is co-linear with and a one-to-one correspondent to the SW maximum, whereas the projecting operation serves the purpose of satisfying the power constraint.

C. Signal Space Property

Direct analysis of convergence properties of SW receivers using the power-constrained constant modulus algorithm is a formidable task. One way to attack this problem is via transforming the optimization of the receiver coefficients \mathbf{f} to the optimization with respect to the global system response \mathbf{q} . However, these two optimizations are not equivalent in general. For convenience, we denote \mathcal{F} as the receiver parameter space, a collection of any linear estimator \mathbf{f} , and $\mathcal{Q} \triangleq \{\mathbf{q} \mid \mathbf{q} = \mathbf{H}^t \mathbf{f}, \mathbf{f} \in$

\mathcal{F} as the space for the global response. To make the analysis in \mathcal{Q} justifiable, we resort to an important concept—the signal space property [17], [20].

The signal space $\mathcal{C}_{\mathbf{H}}$ is defined as the column space of the channel matrix \mathbf{H} . A receiver \mathbf{f}^* is said to have the signal space property if $\mathbf{f}^* \in \mathcal{C}_{\mathbf{H}}$, i.e., $\mathbf{f}^* = \mathbf{H}\mathbf{v}$ for some \mathbf{v} . This implies that the receiver is made of a linear combination of filters matched to the columns of the channel matrix \mathbf{H} , or equivalently, the receiver can be viewed as a cascade of a matched filter and a linear combiner, and it filters out any disturbance not in the direction of the signal. Ericson [8] showed that an optimal receiver designed from any reasonable criterion includes a matched filter as its front end.

The signal space property of SW receivers was proved in [17] for general heterogeneous sources. For the homogeneous sub-Gaussian sources considered in this paper, the result still holds as follows: Assume that $\mathcal{C}_{\mathbf{H}\perp}$ is not empty (i.e., $M > N$). Let \mathbf{f}_s be an SW receiver defined as

$$\mathbf{f}_s \triangleq \arg \max_{\|\mathbf{f}\|_{\mathbb{R}}^2=1} \bar{J}_s(\mathbf{f}). \quad (15)$$

Then

$$\mathbf{f}_s = \arg \max_{\mathbf{f} \in \mathcal{C}_{\mathbf{H}}, \|\mathbf{f}\|_{\mathbb{R}}^2=1} \bar{J}_s(\mathbf{f}). \quad (16)$$

In general, introducing constraints alters locations of local optima and perhaps creates or removes local optima. The above statement shows that restricting an SW receiver to the signal space has no effect on the receiver itself. A similar result applies to CM receivers [17], [20]. It is this property that allows us to analyze SW receivers and their DoA in the global response space.

D. Equivalent Cost Functions

Because of the signal space properties of SW and CM receivers, there exists a one-to-one correspondence between receivers in \mathcal{F} and in \mathcal{Q} . Given a CM or SW receiver \mathbf{f}^* , since $\mathbf{f}^* \in \mathcal{C}_{\mathbf{H}}$, we have

$$\mathbf{f}^* = (\mathbf{H}^t)^{\dagger} \mathbf{q}^*. \quad (17)$$

Therefore, the optimization with respect to the receiver coefficient \mathbf{f} is equivalent to that with respect to the system parameter \mathbf{q} . To obtain the equivalent objective functions in \mathcal{Q} for the previous three algorithms, we need only to substitute (17) into corresponding costs and express the output power $E\{y^2\} = \|\mathbf{f}\|_{\mathbb{R}}^2$ in terms of \mathbf{q}

$$\begin{aligned} E\{y^2\} &= E\{(\mathbf{q}^t \mathbf{s} + \mathbf{f}^t \mathbf{w})^2\} = E\{\mathbf{q}^t \mathbf{s} \mathbf{s}^t \mathbf{q}\} + E\{\mathbf{f}^t \mathbf{w} \mathbf{w}^t \mathbf{f}\} \\ &= \mathbf{q}^t \mathbf{q} + \sigma^2 \mathbf{f}^t \mathbf{f} = \mathbf{q}^t [\mathbf{I} + \sigma^2 (\mathbf{H}^t \mathbf{H})^{-1}] \mathbf{q} = \|\mathbf{q}\|_{\Phi}^2 \end{aligned} \quad (18)$$

where

$$\Phi = \mathbf{I} + \sigma^2 (\mathbf{H}^t \mathbf{H})^{-1}. \quad (19)$$

Hence, corresponding to (5), (11), and (13) obtained in the receiver parameter domain, we have the following equivalent counterparts in the global response space:

$$\begin{aligned} \text{SW: } J_s(\mathbf{q}) &\triangleq \bar{J}_s((\mathbf{H}^t)^{\dagger} \mathbf{q}) \\ &= \|\mathbf{\Pi}^{1/4} \mathbf{q}\|_{\mathbb{R}}^4, \text{ subject to } \|\mathbf{q}\|_{\Phi}^2 = 1 \end{aligned} \quad (20)$$

$$\begin{aligned} \text{CM: } J_c(\mathbf{q}) &\triangleq \bar{J}_c((\mathbf{H}^t)^{\dagger} \mathbf{q}) \\ &= 3\|\mathbf{q}\|_{\Phi}^4 - 2r\|\mathbf{q}\|_{\Phi}^2 - \|\mathbf{\Pi}^{1/4} \mathbf{q}\|_{\mathbb{R}}^4 + r^2 \end{aligned} \quad (21)$$

$$\text{PC-CM: } \mathbf{q}_p \triangleq \arg \min_{\|\mathbf{q}\|_{\Phi}^2=1} J_c(\mathbf{q}). \quad (22)$$

Note that the above one-to-one correspondence exists only between receivers in \mathcal{F} and in \mathcal{Q} . When we consider the convergence region of a certain receiver, we will deal with a sequence $\{\mathbf{f}_k\}$, and each \mathbf{f}_k may not be in $\mathcal{C}_{\mathbf{H}}$, i.e., there does not necessarily exist a one-to-one correspondent relationship between two sequences $\{\mathbf{f}_k\}$ and $\{\mathbf{q}_k | \mathbf{q}_k = \mathbf{H}^t \mathbf{f}_k\}$. This problem will be discussed next.

III. DOMAINS OF ATTRACTION

We first give a definition of domains of attraction. The existence of DoA is guaranteed by the capture theorem [4].

Definition 1: Let g be continuously differentiable and \mathbf{x}^* the only local maximum of g within some open set. A domain of attraction of \mathbf{x}^* is defined as an open set $\mathcal{D}(\mathbf{x})$ containing \mathbf{x}^* such that for any initial point $\mathbf{x}_0 \in \mathcal{D}(\mathbf{x})$, the sequence \mathbf{x}_k generated by the gradient algorithm according to (1) with an arbitrarily small step-size $\mu > 0$ and satisfying $g(\mathbf{x}_{k+1}) \geq g(\mathbf{x}_k)$ for all k : i) remains in $\mathcal{D}(\mathbf{x})$ and ii) converges to \mathbf{x}^* .

To obtain an SW receiver \mathbf{f}^* adaptively, consider updating \mathbf{f}_k using a gradient algorithm with an arbitrarily small step-size on the ellipsoid $\|\mathbf{f}\|_{\mathbb{R}}^2 = 1$. At issue is what are the largest possible DoA of SW receivers. In order to answer this question, we need to know some convergence properties of the SW algorithm. Two scenarios will be considered in this section: the noiseless case and the class of orthogonal channels in the presence of noise. For the latter, the number and the locations of SW receivers have been completely characterized recently [10].

A. Relationship Between DoA in \mathcal{F} and in \mathcal{Q}

Exploring convergence properties in \mathcal{F} is difficult. While it has been shown that there is a one-to-one correspondence between local optima of the SW criterion in \mathcal{F} and those in \mathcal{Q} , it is not clear that DoA in the two spaces are one-to-one correspondent. In fact, when \mathbf{H} has full column rank with $M > N$, a point \mathbf{q}_k may correspond to infinitely many points in \mathcal{F} domain, but two different points in \mathcal{Q} space will not have the same inverse image. Suppose that there are ν receivers in \mathcal{F} , which are denoted as $\mathbf{f}^{(i)}$ ($i = 1, \dots, \nu$) with corresponding DoA $\mathcal{D}_i(\mathbf{f})$, and accordingly, there also exist ν receivers $\mathbf{q}^{(i)}$ ($i = 1, \dots, \nu$) in \mathcal{Q} with DoA $\mathcal{D}_i(\mathbf{q})$. The question is, Are $\mathcal{D}_i(\mathbf{f})$ and $\mathcal{D}_i(\mathbf{q})$ one-to-one correspondent? Here, we establish this connection that allows us to pursue the analysis of DoA in the global response space \mathcal{Q} .

Lemma 1: Let \mathbf{q}^* be a local maximum of (20) and $\mathcal{D}(\mathbf{q})$ be its DoA. Then, under assumptions A1)–A4), $\mathbf{f}^* = (\mathbf{H}^t)^\dagger \mathbf{q}^*$ is a local maximum of (9) with DoA

$$\mathcal{D}(\mathbf{f}) = \{\mathbf{f} \mid \mathbf{H}^t \mathbf{f} \in \mathcal{D}(\mathbf{q})\}. \quad (23)$$

Conversely, if \mathbf{f}^* is a local maximum of (9) and $\mathcal{D}(\mathbf{f})$ its DoA, then $\mathbf{q}^* = \mathbf{H}^t \mathbf{f}^*$ is a local maximum of (20) with DoA

$$\mathcal{D}(\mathbf{q}) = \{\mathbf{q} \mid \mathbf{q} = \mathbf{H}^t \mathbf{f}, \mathbf{f} \in \mathcal{D}(\mathbf{f})\}. \quad (24)$$

Proof: See the Appendix.

Therefore, DoA in \mathcal{Q} and those in \mathcal{F} are also one-to-one correspondent in the sense that for any receiver $\mathbf{q}^{(i)}$ with $\mathcal{D}_i(\mathbf{q})$, there exists a unique $\mathbf{f}^{(i)}$ with $\mathcal{D}_i(\mathbf{f})$, and vice versa, although different points in $\mathcal{D}_i(\mathbf{f})$ may have the same image in $\mathcal{D}_i(\mathbf{q})$ under $\mathbf{q} = \mathbf{H}^t \mathbf{f}$. Furthermore, if a sequence $\{\mathbf{q}_k\}$ in \mathcal{Q} converges to $\mathbf{q}^{(i)}$, any corresponding sequence $\{\mathbf{f}_k\}$ in \mathcal{F} will converge to $\mathbf{f}^{(i)}$. $\mathcal{D}_i(\mathbf{q})$ is simply a projection of $\mathcal{D}_i(\mathbf{f})$ in lower dimensions. From this point of view, we can analyze domains of attraction in the global response space.

B. Domains of Attraction for Noiseless or Orthogonal Channels

It was shown in [10] that each SW receiver $\mathbf{q}^{(i)}$ is co-linear with \mathbf{e}_i if and only if either there is no noise or the columns of \mathbf{H} are orthogonal. In both cases, matrix Φ defined in (19) is diagonal. Our goal in this subsection is to characterize domains of attraction for these two cases. Since the analysis for the noiseless channel ($\Phi = \mathbf{I}$) is a special case of that for the orthogonal channels ($\Phi = \mathbf{I} + \sigma^2 \Lambda$, Λ is diagonal), the characterization of DoA will be focused mainly on the general orthogonal channel case. We begin our discussion in the global response space and then in the receiver parameter domain.

Suppose that the input sequence has the same probability distribution but possibly unequal power with power matrix \mathbf{E} described in the system model. Then, SW receivers can be obtained from the following criterion:

$$\max \|\mathbf{q}\|_4^4, \text{ subject to } \|\mathbf{q}\|_\Phi^2 = 1 \quad (25)$$

where $\mathbf{q} = (\mathbf{H}\mathbf{E})^t \mathbf{f}$, and $\Phi = \mathbf{I} + \sigma^2 \mathbf{E}^{-1} (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{E}^{-1}$. Let $\mathbf{H}^t \mathbf{H} = \text{diag}(d_1, \dots, d_N)$. This leads to

$$\begin{aligned} \Phi &= \text{diag} \left(1 + \frac{\sigma^2}{d_1 \mathcal{E}_1}, \dots, 1 + \frac{\sigma^2}{d_N \mathcal{E}_N} \right) \\ &= \text{diag} \left(\frac{1}{\phi_1^2}, \dots, \frac{1}{\phi_N^2} \right), \quad \phi_i = \frac{1}{\sqrt{1 + \frac{\sigma^2}{d_i \mathcal{E}_i}}}. \end{aligned} \quad (26)$$

The diagonality of Φ enables us to find solutions to (25) and the corresponding DoA.

Lemma 2: Let the columns of $\mathbf{H}_{M \times N}$ be orthogonal and the source signals be i.i.d. Then, SW receivers are given by⁴

$$\mathbf{q}^{(i)} = \phi_i \mathbf{e}_i, \quad i = 1, 2, \dots, N \quad (27)$$

⁴For each $\mathbf{q}^{(i)} = \phi_i \mathbf{e}_i$, there always exists a symmetric receiver $-\mathbf{q}^{(i)}$. Hence, we need only to deal with the positive one.

with the (maximum) domains of attraction⁵

$$\begin{aligned} \mathcal{D}_i(\mathbf{q}) &= \{\mathbf{q} \mid \|\mathbf{q}\|_\Phi^2 = 1, \|\Phi^{-(1/2)} \mathbf{q} - \mathbf{e}_i\| \\ &< \|\Phi^{-(1/2)} \mathbf{q} - \mathbf{e}_j\|, \forall j \neq i, j = 1, \dots, N\} \end{aligned} \quad (28)$$

where Φ is defined in (26).

Proof: See the Appendix.

To illustrate Lemma 2 geometrically, let us consider the special orthonormal channels with unit power input signals in three dimensions as shown in Fig. 2. In this situation, $\Phi = (1 + \sigma^2) \mathbf{I}$ and $\|\mathbf{q}\|_\Phi^2 = 1$ defines a perfect ball. SW receivers are given by $\mathbf{q}^{(i)} = 1/(\sqrt{1 + \sigma^2}) \mathbf{e}_i$ ($i = 1, 2, 3$), and the corresponding $\mathcal{D}_i(\mathbf{q})$ is bounded by $\{\bar{\mathbf{q}} \mid \|\bar{\mathbf{q}}\|_\Phi^2 = 1, \bar{q}_i = \bar{q}_j, \bar{q}_i \geq \bar{q}_k, j \neq i, k \neq j\}$. Therefore, the DoA is specified by the intersection of the minimum distance decision region for \mathbf{e}_i and the sphere determined by the noise. As the SNR increases, the ball will expand, but the above relationship still holds true. This implies that the relative size of DoA remains the same for any SNR. In other words, we always have $|\mathcal{D}_1(\mathbf{q})| = |\mathcal{D}_2(\mathbf{q})| = |\mathcal{D}_3(\mathbf{q})|$ at an arbitrary noise level. More general results regarding the size of DoA will be presented in the next subsection. When $\sigma = 0$, $\|\mathbf{q}\|_\Phi^2 = 1$ turns out to be a unit ball. This noiseless case is the scenario considered in [1] but for different cost functions. The conclusion is also the same as in [6] when CM receivers are power constrained.

As the channel condition changes so that the columns of \mathbf{H} are no longer orthonormal (but still orthogonal), or input signals have distinct power, $\|\mathbf{q}\|_\Phi^2 = 1$ becomes an ellipsoid. The boundary points $\bar{\mathbf{q}}$ of $\mathcal{D}_i(\mathbf{q})$ satisfy $\phi_i \bar{q}_i = \phi_j \bar{q}_j$ ($\forall j \neq i$) with $\phi_i \bar{q}_i \geq \phi_k \bar{q}_k$ ($k \neq j, k \neq i$). Again, we see that DoA are the minimum distance decision regions but on an ellipsoid.

Now, let us extend the above discussion to a general case. What if input signals have different constellations? According to (20), SW receivers are obtained as $\mathbf{q}^* = \arg \max_{\|\mathbf{q}\|_\Phi^2=1} \|\mathbf{\Pi}^{1/4} \mathbf{q}\|_4^4$. To make use of the result in Lemma 2, we apply the transformation $\mathbf{q} = \mathbf{\Pi}^{-(1/4)} \mathbf{p}$, which leads to the following equivalent optimization:

$$\max \|\mathbf{p}\|_4^4, \text{ subject to } \|\mathbf{p}\|_\Gamma^2 = 1 \quad (29)$$

where $\Gamma = \mathbf{\Pi}^{-(1/4)} \Phi \mathbf{\Pi}^{-(1/4)}$ is diagonal. Substituting Φ with (26) yields

$$\begin{aligned} \Gamma &= \text{diag} \left(\frac{1 + \frac{\sigma^2}{d_1 \mathcal{E}_1}}{\sqrt{\pi_1}}, \dots, \frac{1 + \frac{\sigma^2}{d_N \mathcal{E}_N}}{\sqrt{\pi_N}} \right) \\ &= \text{diag} \left(\frac{1}{\gamma_1^2}, \dots, \frac{1}{\gamma_N^2} \right), \quad \gamma_i = \frac{\sqrt[4]{\pi_i}}{\sqrt{1 + \frac{\sigma^2}{d_i \mathcal{E}_i}}}. \end{aligned} \quad (30)$$

Because (29) and (25) are of exactly the same form, a similar conclusion can be reached as in (27) and (28).

Thus far, we have discussed domains of attraction of SW receivers in the space of the global system response. Due to the one-to-one correspondence existing between DoA in \mathcal{F} and in

⁵More accurately, the DoA of receiver $\mathbf{q}^{(i)}$ should satisfy $\|\Phi^{-(1/2)} \mathbf{q} \pm \mathbf{e}_i\| < \|\Phi^{-(1/2)} \mathbf{q} \pm \mathbf{e}_j\|$. However, since Φ is diagonal, $\|\mathbf{q}\|_\Phi^2 = 1$ is symmetric about all the axes, and $\mathcal{D}_i(\mathbf{q})$ is symmetric about \mathbf{e}_i . Therefore, we can simply consider the positive coordinates of DoA.

\mathcal{Q} , DoA in the receiver parameter space can be obtained accordingly.

Theorem 1: Let the columns of $\mathbf{H}_{M \times N}$ be orthogonal. SW receivers are given by

$$\mathbf{f}^{(i)} = \frac{\phi_i}{\sqrt{\mathcal{E}_i}} (\mathbf{H}^t)^\dagger \mathbf{e}_i, \quad i = 1, \dots, N \quad (31)$$

with the (maximum) domains of attraction

$$\mathcal{D}_i(\mathbf{f}) = \left\{ \mathbf{f} \mid \|\mathbf{f}\|_{\mathbf{R}}^2 = 1, \|\mathbf{A}\mathbf{f} - \mathbf{e}_i\| < \|\mathbf{A}\mathbf{f} - \mathbf{e}_j\| \right. \\ \left. \forall j \neq i, j = 1, \dots, N \right\} \quad (32)$$

where $\mathbf{A} = \mathbf{\Pi}^{(1/2)} \mathbf{\Phi}^{- (1/2)} \mathbf{E} \mathbf{H}^t$, and $\mathbf{\Phi}$ is related to \mathbf{R} by $\mathbf{R} = \mathbf{H} \mathbf{E} \mathbf{\Phi} \mathbf{E} \mathbf{H}^t$.

Proof: See the Appendix.

Theorem 1 enables us to analyze convergence regions of SW receivers for multiple users with unequal power and distinct constellations when columns of \mathbf{H} are orthogonal. As we mentioned before, orthogonal channels play an important part in communications, and the above results are of practical significance. One special application is found in synchronous CDMA systems. Let $\mathbf{C}_{M \times N}$ correspond to the codeword matrix with spreading gain M and N users and $\mathbf{c}_i \triangleq \mathbf{C} \mathbf{e}_i$ ($i = 1, \dots, N$) be the codeword for the i th user. Assume that the source signals are i.i.d. with unit power. If under certain circumstances [e.g., ideal channel without intersymbol interference (ISI)] we have $\mathbf{H} = \mathbf{C}$ with $\mathbf{C}^t \mathbf{C} = \mathbf{C} \mathbf{C}^t = \mathbf{I}$, then for any initialization $\mathbf{f}_0 \in \mathcal{D}_i(\mathbf{f})$, $\mathbf{f}_k \rightarrow \mathbf{c}_i$ as $k \rightarrow \infty$, where

$$\mathcal{D}_i(\mathbf{f}) = \left\{ \mathbf{f} \mid \|\mathbf{f}\| = \frac{1}{\sqrt{1 + \sigma^2}}, \|\mathbf{f} - \mathbf{c}_i\| < \|\mathbf{f} - \mathbf{c}_j\| \right. \\ \left. \forall j \neq i, j = 1, \dots, N \right\}. \quad (33)$$

It is interesting to note that every equalizer will converge to the corresponding user codeword (up to a scalar) at an arbitrary noise level. For each code \mathbf{c}_i , its DoA is defined by the intersection of the minimum distance decision region for \mathbf{c}_i and the sphere determined by the noise level. The geometrical illustration is exactly the same as in Fig. 2, except that it is now in \mathcal{F} domain and \mathbf{c}_i is in stead of \mathbf{e}_i .

Theorem 1 delineates DoA for SW receivers obtained by the gradient algorithm in the receiver parameter domain where practical adaptation is performed, whereas in the global response space, although implementation is impracticable, theoretical analysis is easier to carry out. Since both receivers and their convergence regions are one-to-one correspondent in these two spaces and DoA obtained in \mathcal{Q} have rather straightforward geometrical interpretation (especially for the orthogonal channel case), the following analysis will be concentrated on the global response space.

C. What Affect the Size of DOA?

In this subsection, we focus on the problem how the size of DoA is affected by different elements. Suppose that we are given receivers $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ for two different sources with DoA $\mathcal{D}_1(\mathbf{q})$ and $\mathcal{D}_2(\mathbf{q})$, respectively. Let $\mathbf{q}^{(i)}$ and $\mathcal{D}_i'(\mathbf{q})$ ($i =$

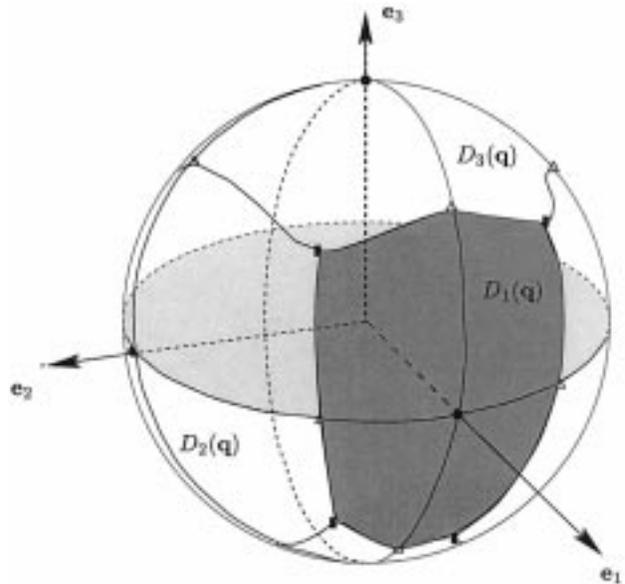


Fig. 2. Domains of attraction for orthonormal channels. Dot: maximum; square: minimum; triangle: saddle.

1, 2) be the corresponding receivers and DoA resulting from the change of some system parameter (e.g., SNR). Now the question is, What is the relationship between $|\mathcal{D}_1'(\mathbf{q})|$ and $|\mathcal{D}_2'(\mathbf{q})|$ given the relationship between $\mathcal{D}_1(\mathbf{q})$ and $\mathcal{D}_2(\mathbf{q})$, say, $|\mathcal{D}_1(\mathbf{q})| = |\mathcal{D}_2(\mathbf{q})|$? Specifically, is it easier to converge to receiver $\mathbf{q}^{(1)}$ than to receiver $\mathbf{q}^{(2)}$ ($|\mathcal{D}_1'(\mathbf{q})| > |\mathcal{D}_2'(\mathbf{q})|$), or are the two captured with equal probability ($|\mathcal{D}_1'(\mathbf{q})| = |\mathcal{D}_2'(\mathbf{q})|$)? Finding the answer to this question will allow us to determine whether and how we can shrink or enlarge the DoA of a receiver of interest by changing certain parameter(s). For example, it has been shown that for the i.i.d. equal power sources and orthonormal channels, $\mathcal{D}_1'(\mathbf{q})$ and $\mathcal{D}_2'(\mathbf{q})$ always have the same size, regardless of the noise level. Clearly, in this case, it is not possible to expand certain DoA by simply increasing or reducing the noise variance, but what happens to the general orthogonal channels and non-identically distributed sources?

Formulated in (20), SW receivers are obtained by maximizing $\|\mathbf{\Pi}^{1/4} \mathbf{q}\|_{\mathbf{\Phi}}^4$ subject to the power constraint $\|\mathbf{q}\|_{\mathbf{\Phi}}^2 = 1$. Since $\mathbf{\Pi}$ is diagonal, $\|\mathbf{\Pi}^{1/4} \mathbf{q}\|_{\mathbf{\Phi}}^4$ is nothing but a weighted 4-norm, and its maxima are always attained at $\phi_i \mathbf{e}_i$ whether $\mathbf{\Pi} = \mathbf{I}$ or not. However, its minima and saddles, of which the boundary of DoA consists, are usually not the same for different $\mathbf{\Pi}$. Therefore, signal constellations will affect the size of DoA. Generally speaking, as the shape of the ellipsoid and the boundary of DoA are determined by $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ [as defined in (26) and (30)], the size of DoA depends on the signal power, the signal constellation, the noise level, and the channel condition. In order to describe the relative size variation of DoA, we need to examine boundary points $\bar{\mathbf{q}}$, especially those coordinates satisfying the equality relationship. By the transformation $\mathbf{q} = \mathbf{\Pi}^{- (1/4)} \mathbf{p}$ introduced in the previous subsection, it is easy to obtain these coordinates with respect to \mathbf{p} : $\gamma_i \bar{p}_i = \gamma_j \bar{p}_j, j \neq i$, where $\gamma_i = \sqrt[4]{\pi_i} \phi_i$. Thus, a boundary point $\bar{\mathbf{q}}$ of $\mathcal{D}_i(\mathbf{q})$ satisfies

$$\sqrt{\pi_i} \phi_i \bar{q}_i = \sqrt{\pi_j} \phi_j \bar{q}_j, \quad j \neq i. \quad (34)$$

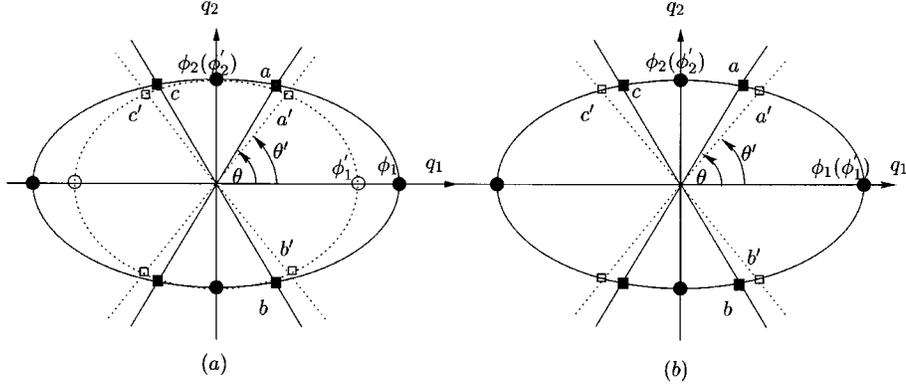


Fig. 3. Changes of DoA at a fixed noise level. (a) \mathcal{E}_1 or d_1 decreases. (b) π_1 decreases. Circle/dot: maximum; square: minimum.

When $\pi_i = \pi \forall i$, we have $\phi_i \bar{q}_i = \phi_j \bar{q}_j$, which defines the minimum distance decision region as obtained for the i.i.d. source case.

Now, we are able to show the effect of the above-mentioned four parameters on the size of DoA. We first discuss how π_i , d_i , and \mathcal{E}_i influence domains of attraction of SW receivers.

Property 1: For fixed noise variance ($\sigma \neq 0$), a decrease in the i th input signal power or the absolute value of its cumulant, or the i th diagonal element of $\mathbf{H}^t \mathbf{H}$ will lead to the decrease in $|\mathcal{D}_i(\mathbf{q})|$.

Proof: First, assume that the input has the same constellation, i.e., $\pi_i = \pi, i = 1, \dots, N$. Consider the maximum $\mathbf{q}^{(1)} = \phi_1 \mathbf{e}_1$. Suppose that \mathcal{E}_1 decreases, whereas other $\mathcal{E}_k (k \neq 1)$ remain unchanged. This results in the decrease of ϕ_1 , and the ellipsoid $\sum_{i=1}^N (q_i / \phi_i)^2 = 1$ shrinks at $\mathbf{q}^{(1)}$. Further, any point $\bar{\mathbf{q}}$ on the boundary of $\mathcal{D}_1(\mathbf{q})$ satisfies $\phi_1 \bar{q}_1 = \phi_k \bar{q}_k, k \neq 1$. Since the length of ϕ_1 is reduced, the ratio (\bar{q}_1 / \bar{q}_k) will increase, which implies that all the boundary points will move toward $\mathbf{q}^{(1)}$, causing the area of $\mathcal{D}_1(\mathbf{q})$ to become smaller. A decrease in d_1 produces the same result. Note that when $\sigma = 0$, all the DoA are of the same size on the unit sphere, and neither the channel condition nor the signal power will take effect.

Next, we show that this conclusion also holds for signals with different constellations. Assume that π_1 decreases with other $\pi_k (k \neq 1)$ being fixed. Since π_1 has nothing to do with ϕ_1 , the ellipsoid $\sum_{i=1}^N (q_i / \phi_i)^2 = 1$ remains as before. However, from (34), we have $(\bar{q}_1 / \bar{q}_k) = (\sqrt{\pi_k} \phi_k / \sqrt{\pi_1} \phi_1)$, which will increase as π_1 decreases, leading to the shrinkage of $\mathcal{D}_1(\mathbf{q})$. \square

The interpretation of this property in two dimensions is illustrated in Fig. 3. Suppose that $\phi_1 > \phi_2$ and $|\mathcal{D}_1(\mathbf{q})| > |\mathcal{D}_2(\mathbf{q})|$. In Fig. 3(a), as ϕ_1 is reduced to ϕ'_1 due to the decrease in \mathcal{E}_1 or d_1 , the ellipsoid shrinks from the solid one to the dotted one. Meanwhile, the angle θ defining the boundary of $\mathcal{D}_1(\mathbf{q})$ turns smaller (given $\pi_1 = \pi_2$)

$$\tan \theta' = \frac{\bar{q}'_2}{\bar{q}'_1} = \frac{\phi'_1}{\phi'_2} < \frac{\phi_1}{\phi_2} = \frac{\bar{q}_2}{\bar{q}_1} = \tan \theta. \quad (35)$$

Hence, $(|\mathcal{D}'_1(\mathbf{q})| / |\mathcal{D}'_2(\mathbf{q})|) = \left(\widehat{a'b'} / \widehat{a'c'} \right) < \left(\widehat{ab} / \widehat{ac} \right) = (|\mathcal{D}_1(\mathbf{q})| / |\mathcal{D}_2(\mathbf{q})|)$. Particularly, if we make ϕ'_1 equal to ϕ'_2 , then $|\mathcal{D}'_1(\mathbf{q})| = |\mathcal{D}'_2(\mathbf{q})|$. Further reducing ϕ_1 results in $|\mathcal{D}'_1(\mathbf{q})| < |\mathcal{D}'_2(\mathbf{q})|$. Here, we see an example that a global

receiver can be transformed into a local receiver (and vice versa) by varying a system parameter. In Fig. 3(b), the ellipsoid remains unchanged. However, as π_1 decreases

$$\tan \theta' = \frac{\bar{q}'_2}{\bar{q}'_1} = \frac{\sqrt{\pi'_1} \phi'_1}{\sqrt{\pi'_2} \phi'_2} < \frac{\sqrt{\pi_1} \phi_1}{\sqrt{\pi_2} \phi_2} = \frac{\bar{q}_2}{\bar{q}_1} = \tan \theta \quad (36)$$

causing the decrease in $|\mathcal{D}_1(\mathbf{q})|$. If, as an extreme example, π_1 is further reduced to zero, which corresponds to a Gaussian source, then $|\mathcal{D}_1(\mathbf{q})| = 0$, i.e., the DoA of $\mathbf{q}^{(1)}$ disappears. This agrees with the fact that the SWA cannot detect Gaussian random variables.

For the global SW receiver $\mathbf{q}^{(l)} = \phi_l \mathbf{e}_l, l = \arg \max_{1 \leq i \leq N} \phi_i$, the boundary points of $\mathcal{D}_l(\mathbf{q})$ satisfy $(\bar{q}_l / \bar{q}_k) = (\sqrt{\pi_k} \phi_k / \sqrt{\pi_l} \phi_l) (k \neq l, k = 1, \dots, N)$. Thus, if $\pi_1 = \dots = \pi_N$, then $\bar{q}_l < \bar{q}_k \forall k \neq l$, which leads to the following conclusion.

Corollary 1: If the source signals have the same constellation, the global SW receiver has the largest domain of attraction.

When we have nonuniform source distributions, a nonglobal receiver may have the largest DoA, provided that the corresponding source has a negative enough cumulant, i.e., the more sub-Gaussian a source, the larger the convergence region. Because the larger the DoA, the easier for the optimum to be captured, the above result indicates that, in practice, we can decrease the cumulant of a (sub-Gaussian) signal of interest to guarantee the convergence of the corresponding equalizer. Usually, signals with denser constellations have larger cumulants, e.g., the cumulant of 8PAM is -1.238 , whereas that of BPSK is -2 . Hence, for a communication system with nonidentically distributed sources, there exists some tradeoff between bandwidth efficiency and receiver convergence capability.

We next examine the influence of the noise variance on the size of DoA. Different from the previous case, as σ varies, all ϕ_i will change accordingly. Assume that σ increases with other parameters being fixed. First, we note that the ellipsoid will shrink at every optimum point due to

$$\frac{\phi'_i}{\phi_i} = \frac{1}{\sqrt{1 + \frac{\sigma^2}{a_i^2 \varepsilon_i}}} < 1. \quad (37)$$

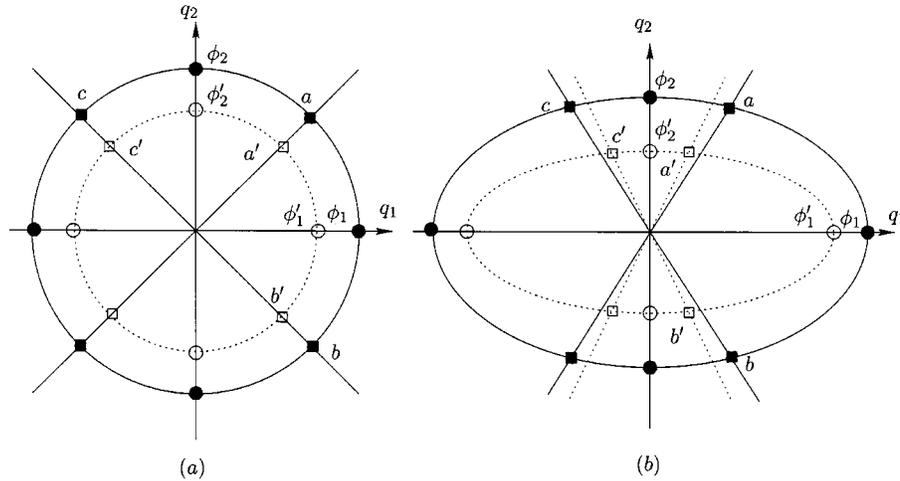


Fig. 4. Changes of DoA when a increases. (a) $d_1 \mathcal{E}_1 = d_2 \mathcal{E}_2$. (b) $d_1 \mathcal{E}_1 > d_2 \mathcal{E}_2$. Circle/dot: maximum; square: minimum.

Second, we need to figure out to what extent it shrinks at each optimum by comparing the ratio

$$\frac{\phi'_i/\phi_i}{\phi'_j/\phi_j} = \frac{\phi'_i \phi_j}{\phi_j \phi_i} = \sqrt{\frac{\left(1 + \frac{\sigma'^2}{d_j \mathcal{E}_j}\right) \left(1 + \frac{\sigma^2}{d_i \mathcal{E}_i}\right)}{\left(1 + \frac{\sigma'^2}{d_i \mathcal{E}_i}\right) \left(1 + \frac{\sigma^2}{d_j \mathcal{E}_j}\right)}} \quad (38)$$

which requires evaluating

$$\begin{aligned} \Delta &= \left(\frac{\sigma'^2}{d_j \mathcal{E}_j} + \frac{\sigma^2}{d_i \mathcal{E}_i} \right) - \left(\frac{\sigma'^2}{d_i \mathcal{E}_i} + \frac{\sigma^2}{d_j \mathcal{E}_j} \right) \\ &= (\sigma'^2 - \sigma^2) \left(\frac{1}{d_j \mathcal{E}_j} - \frac{1}{d_i \mathcal{E}_i} \right). \end{aligned} \quad (39)$$

Finally, the change in the relative size of DoA can be determined by checking the variation of the boundary points

$$\frac{\bar{q}'_i/\bar{q}'_j}{\bar{q}_i/\bar{q}_j} = \frac{\bar{q}'_i \bar{q}_j}{\bar{q}_i \bar{q}'_j} = \frac{\sqrt{\pi_j} \phi'_j \sqrt{\pi_i} \phi_i}{\sqrt{\pi_i} \phi'_i \sqrt{\pi_j} \phi_j} = \frac{\phi'_j \phi_i}{\phi'_i \phi_j}. \quad (40)$$

Consider the following two cases.

- i) See Fig. 4(a). If for any i and j , $d_i \mathcal{E}_i = d_j \mathcal{E}_j$, or $d_i = d_j$ and $\mathcal{E}_i = \mathcal{E}_j$, then $\Delta = 0$ and $((\phi'_i/\phi_i)/(\phi'_j/\phi_j)) = 1$. In this situation, $\|\mathbf{q}\|_{\Phi}^2 = 1$ defines a perfect ball ($\phi_i = \phi_j$), and the increase in σ will cause the whole ball to shrink proportionally with $\phi'_i = \phi'_j$. Hence, $((\bar{q}'_i/\bar{q}'_j)/(\bar{q}_i/\bar{q}_j)) = 1$, i.e., the relative size of the DoA for each receiver remains unchanged. If $\pi_1 = \dots = \pi_N$, then $|\mathcal{D}'_1(\mathbf{q})| = \dots = |\mathcal{D}'_N(\mathbf{q})|$.
- ii) See Fig. 4(b). Suppose $\pi_1 = \dots = \pi_N$ and $d_i \mathcal{E}_i > d_j \mathcal{E}_j$. This leads to $\Delta > 0$ and $\phi_i > \phi_j$. Then, $((\phi'_i/\phi_i)/(\phi'_j/\phi_j)) > 1$, which implies that the ellipsoid shrinks less at the point with larger ϕ_i , or larger DoA. Furthermore, $((\bar{q}'_i/\bar{q}'_j)/(\bar{q}_i/\bar{q}_j)) < 1$ indicates that the larger DoA becomes even larger. If we have nonuniform source statistics, although the signal constellation does have an effect on the boundary [see (34)], it neither exerts any influence on the size and the shape of the ellipsoid (which depend on Φ only), nor affects the relative variation of the boundary [see (40)]. Therefore, if $d_i \mathcal{E}_i > d_j \mathcal{E}_j$, we always have

$(|\mathcal{D}'_i(\mathbf{q})|/|\mathcal{D}'_j(\mathbf{q})|) > (|\mathcal{D}_i(\mathbf{q})|/|\mathcal{D}_j(\mathbf{q})|)$, regardless of values of signal cumulants. This result shows that even if the global receiver (with the largest ϕ_i) has a smaller domain of attraction (due to its larger cumulant), its convergence region will grow as σ^2 increases. That is to say, a worse SNR situation favors the global receiver. However, this growth has an end when the noise variance is sufficiently large because

$$\lim_{\sigma'^2 \rightarrow \infty} \frac{\phi'_i}{\phi'_j} = \lim_{\sigma'^2 \rightarrow \infty} \frac{\sqrt{1 + \frac{\sigma'^2}{d_j \mathcal{E}_j}}}{\sqrt{1 + \frac{\sigma'^2}{d_i \mathcal{E}_i}}} = \sqrt{\frac{d_i \mathcal{E}_i}{d_j \mathcal{E}_j}}. \quad (41)$$

In all, we have the following result.

Property 2: Given the signal constellation and power and the channel condition, as the noise variance increases, we have the following.

- i) For the equal power source and orthonormal channel case, the relative size of all the DoA remains unchanged.
- ii) Generally, the DoA corresponding to a larger ϕ_i tends to increase, or mathematically, if $\phi_i > \phi_j$, then $(|\mathcal{D}'_i(\mathbf{q})|/|\mathcal{D}'_j(\mathbf{q})|) > (|\mathcal{D}_i(\mathbf{q})|/|\mathcal{D}_j(\mathbf{q})|)$.

Since, in a low SNR scenario, the global receiver tends to achieve the largest possible domain of attraction, this property explains why a weak signal may never be detected under noisy circumstance. In order for all the signals to be captured, one way is to design i.i.d. signals and channels satisfying the conditions in i), or alternatively, and more practically, we can decrease the cumulant of a signal with low power to increase its detection probability at low SNR.

D. Two-Dimensional Case

To gain an insight into the previous analysis, in this subsection, we provide an illustrative example in 2-D global response space for i.i.d. sources. Due to the close relationship between Shalvi-Weinstein and constant modulus receivers [13], DoA of the latter are considered as well. We note that although CM and SW receivers are one-to-one correspondent, such a relationship does not hold for other stationary points. Since the boundary of DoA is determined by these stationary points, their domains of

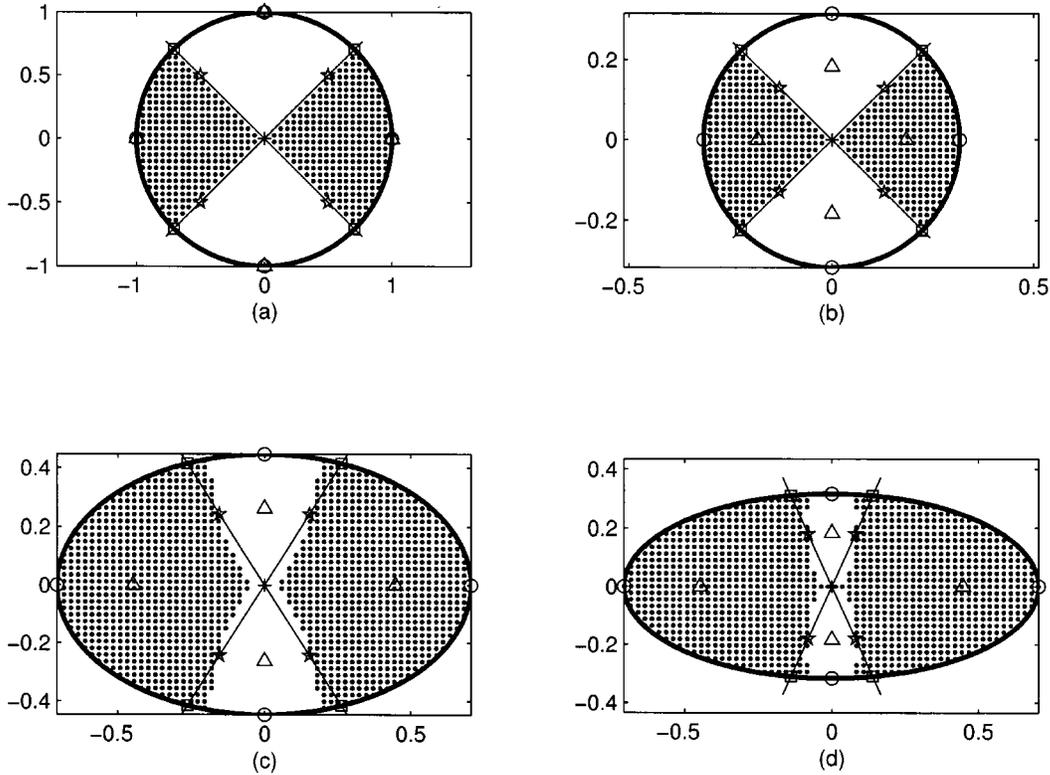


Fig. 5. Comparison between DoA of SW and CM receivers. Circle: maximum of SWA; square: minimum of SWA; triangle: minimum of CMA; plus: maximum of CMA; pentagram: saddle of CMA. (a) $\Phi = \text{diag}(1, 1)$. (b) $\Phi = \text{diag}(10, 10)$. (c) $\Phi = \text{diag}(2, 5)$. (d) $\Phi = \text{diag}(2, 10)$.

attraction are different in general. This key difference is demonstrated in Fig. 5. The ellipsoid drawn in thick solid line corresponds to $\|\mathbf{q}\|_{\Phi}^2 = 1$. Two solid cross lines demarcate the different DoA of SW receivers on the ellipsoid. For CMA, we only consider convergence regions defined by $\|\mathbf{q}\|_{\Phi}^2 \leq 1$ due to the output power constraint of any CM receiver proved in [20]. Dotted areas are DoA of global CM receivers, whereas blank regions in between correspond to DoA of local CM receivers. When there is no noise (see Fig. 5(a)), SW and CM receivers coincide and have the same DoA on the unit circle. In the presence of noise, they still can have the same convergence regions (on the ellipsoid), provided that the channel is orthonormal and the input signals have the same power as shown in Fig. 5(b). Note that there is a scalar difference between these two receivers. This scenario corresponds to the situation illustrated in Fig. 4(a). For general orthogonal channels, however, CM and SW receivers have distinct domains of attraction [see Fig. 5(c) and (d)], resulting from their different transient behavior in convergence. Further, DoA of both SW and CM local receivers shrink as ϕ_2 decreases (e.g., from $1/\sqrt{5}$ to $1/\sqrt{10}$), which implies that it is easier to capture the global optimum (associated with the stronger signal). This observation agrees with the theoretical result presented in the previous subsection.

As an extension to the orthogonal channel case, we consider the general 2-D global response. It was proved in [10] that there are either two or four SW receivers, depending on the noise level and the channel condition. Further, the numbers of local minima and maxima of the SW cost are the same, and they alternate on the ellipsoid defined by $\|\mathbf{q}\|_{\Phi}^2 = 1$. Hence, domains of attraction of SW receivers are bounded by the minima of the SW cost

(see Fig. 6). Because the location of each SW receiver cannot be obtained analytically, it is not easy to characterize the size of DoA in general. Nevertheless, from the simulation, it is not surprising to see that the global receiver always has a larger DoA than the local one.

IV. CONCLUSION

In this paper, we analyzed domains of attraction for the gradient ascent implementation of the SW criterion. DoA of SW receivers were characterized for the noiseless and orthogonal channel cases. When the columns of \mathbf{H} are orthogonal, we concluded that in the global response space, DoA of SW receivers are the minimum distance decision regions on an ellipsoid determined by the channel condition, the signal power, and the noise variance. The result obtained in the practical parameter space enables us to analyze convergence behavior of SW receivers for input signals with nonuniform statistics and possibly unequal power in a multiuser system. We further investigated the effect of system parameters on the relative size variation of domains of attraction, which provides some insight into the signal design principles, e.g., at a fixed noise level, in order for a signal to be easily detected, we can either increase its power or decrease its cumulant to make it more sub-Gaussian; if all users are of same importance, it is necessary to assign a smaller cumulant to a signal without high enough power so that even as SNR deteriorates it still can be captured. While we have shown the one-to-one correspondence between DoA in the global response space and those in the receiver parameter domain based on the signal space property, how DoA change in \mathcal{F} is still not clear

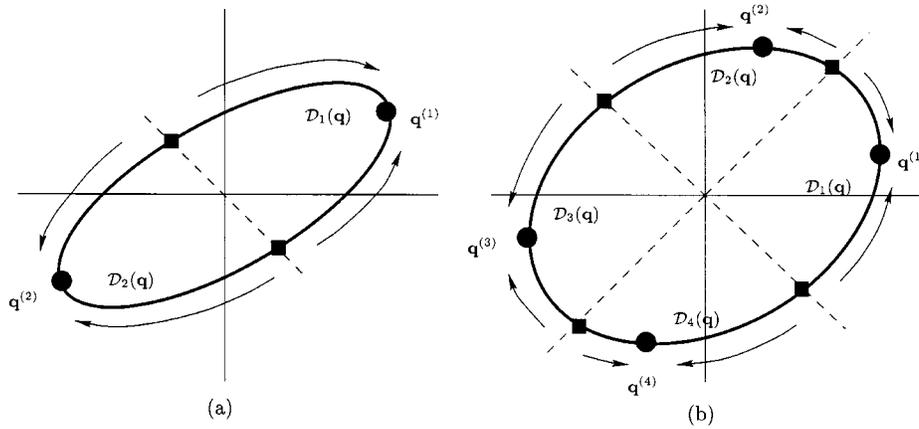


Fig. 6. DoA of SW receivers with (a) two maxima and (b) four maxima. Dot: maximum; square: minimum.

because the mapping from \mathcal{Q} to \mathcal{F} is possibly one-to-infinity. Further, DoA of SW receivers in the general high dimensions remain undefined due to the fact that the convergence properties are largely unknown in this case.

APPENDIX A PROOF OF LEMMA 1

The one-to-one correspondence between receivers $\mathbf{f}^{(i)}$ and $\mathbf{q}^{(i)}$ is guaranteed by the signal space property of SW receivers. We now prove that such relationship also exists for DoA in two spaces.

First, we claim that if $\bar{\mathbf{f}}$ is a stationary point in \mathcal{F} , then $\bar{\mathbf{q}} = \mathbf{H}^t \bar{\mathbf{f}}$ is also a stationary point in \mathcal{Q} ; conversely, if $\bar{\mathbf{q}}$ is a stationary point in \mathcal{Q} , then $\{\bar{\mathbf{f}} | \mathbf{H}^t \bar{\mathbf{f}} = \bar{\mathbf{q}}\}$ are also stationary points in \mathcal{F} . Without loss of generality, assume that $\mathbf{\Pi} = \mathbf{I}$. For constrained optimization, we can form the Lagrange function and calculate its gradient as

$$L(\mathbf{f}) = \sum_i (\mathbf{f}^t \mathbf{H} \mathbf{e}_i)^4 + \lambda \mathbf{f}^t \mathbf{R} \mathbf{f} \quad (42)$$

$$\nabla L(\mathbf{f}) = 4 \sum_i (\mathbf{f}^t \mathbf{H} \mathbf{e}_i)^3 \mathbf{H} \mathbf{e}_i + 2\lambda \mathbf{R} \mathbf{f}, \quad \mathbf{R} = \mathbf{H} \Phi \mathbf{H}^t \quad (43)$$

or in \mathcal{Q}

$$L(\mathbf{q}) = \sum_i (\mathbf{q}^t \mathbf{e}_i)^4 + \lambda \mathbf{q}^t \Phi \mathbf{q} \quad (44)$$

$$\nabla L(\mathbf{q}) = 4 \sum_i (\mathbf{q}^t \mathbf{e}_i)^3 \mathbf{e}_i + 2\lambda \Phi \mathbf{q}. \quad (45)$$

If $\bar{\mathbf{f}}$ is a stationary point, then $\nabla L(\bar{\mathbf{f}}) = \mathbf{0}$, i.e.,

$$4 \sum_i (\bar{\mathbf{q}}^t \mathbf{e}_i)^3 \mathbf{H} \mathbf{e}_i + 2\lambda \mathbf{H} \Phi \bar{\mathbf{q}} = \mathbf{0} \quad (46)$$

$$\Rightarrow \mathbf{H} \left[4 \sum_i (\bar{\mathbf{q}}^t \mathbf{e}_i)^3 \mathbf{e}_i + 2\lambda \Phi \bar{\mathbf{q}} \right] = \mathbf{0} \quad (47)$$

$$\Rightarrow \mathbf{H} \nabla L(\bar{\mathbf{q}}) = \mathbf{0}. \quad (48)$$

Since \mathbf{H} has full column rank, we have $\nabla L(\bar{\mathbf{q}}) = \mathbf{0}$. On the other hand, if $\nabla L(\bar{\mathbf{q}}) = \mathbf{0}$, then

$$4 \sum_i (\bar{\mathbf{f}}^t \mathbf{H} \mathbf{e}_i)^3 \mathbf{e}_i + 2\lambda \Phi \mathbf{H}^t \bar{\mathbf{f}} = \mathbf{0} \quad (49)$$

$$\Rightarrow 4 \sum_i (\bar{\mathbf{f}}^t \mathbf{H} \mathbf{e}_i)^3 \mathbf{H} \mathbf{e}_i + 2\lambda \mathbf{H} \Phi \mathbf{H}^t \bar{\mathbf{f}} = \mathbf{0} \quad (50)$$

$$\Rightarrow \nabla L(\bar{\mathbf{f}}) = \mathbf{0}. \quad (51)$$

As $M > N$, $\bar{\mathbf{q}}$ may correspond to $\{\bar{\mathbf{f}}^l\}$ ($l = 1, 2, \dots$). Due to the signal space property, only those nonoptimum stationary points that constitute the boundary of DoA can have infinitely many inverse images.

Second, suppose that $\{\mathbf{f}_k\}$ converges to $\mathbf{f}^{(i)}$, i.e., for any $\epsilon > 0$, there exists a neighborhood of $\mathbf{f}^{(i)}$ and an integer K such that $\|\mathbf{f}_k - \mathbf{f}^{(i)}\| < \epsilon/\alpha$ ($\alpha = \|\mathbf{H}\|$) for $k > K$. Then

$$\begin{aligned} \|\mathbf{q}_k - \mathbf{q}^{(i)}\| &= \|\mathbf{H}^t (\mathbf{f}_k - \mathbf{f}^{(i)})\| \\ &\leq \|\mathbf{H}^t\| \|\mathbf{f}_k - \mathbf{f}^{(i)}\| < \alpha \frac{\epsilon}{\alpha} = \epsilon, \quad k > K \end{aligned} \quad (52)$$

i.e., $\{\mathbf{q}_k\}$ converges to $\mathbf{q}^{(i)}$. This implies that if $\mathbf{f}_k \in \mathcal{D}_i(\mathbf{f})$, then $\mathbf{q}_k = \mathbf{H}^t \mathbf{f}_k \in \mathcal{D}_i(\mathbf{q})$. Conversely, assume that $\mathbf{q}_k \in \mathcal{D}_i(\mathbf{q})$ and that \mathbf{q}_k corresponds to $\{\mathbf{f}_k^l\}$. If for any l , $\mathbf{f}_k^l \in \mathcal{D}_j(\mathbf{f})$ and $j \neq i$, then $\mathbf{H}^t \mathbf{f}_k^l \in \mathcal{D}_j(\mathbf{q})$, i.e., $\mathbf{q}_k \in \mathcal{D}_j(\mathbf{q})$, which contradicts the assumption. Hence, if $\mathbf{q}_k \in \mathcal{D}_i(\mathbf{q})$, then $\{\mathbf{f}_k^l | \mathbf{H}^t \mathbf{f}_k^l = \mathbf{q}_k, l = 1, 2, \dots\} \subset \mathcal{D}_i(\mathbf{f})$.

Third, consider a sequence $\{\mathbf{q}_k\}$ generated by a gradient ascent algorithm with $J_s(\mathbf{q}_{k+1}) \geq J_s(\mathbf{q}_k) \forall k$. Suppose that \mathbf{q}_k corresponds to $\{\mathbf{f}_k^{l_0}\}$ ($l_0 = 1, 2, \dots$) and \mathbf{q}_{k+1} is associated with $\{\mathbf{f}_{k+1}^{l_1}\}$ ($l_1 = 1, 2, \dots$). Since

$$\bar{J}_s(\mathbf{f}_{k+1}^{l_1}) = J_s(\mathbf{q}_{k+1}) \geq J_s(\mathbf{q}_k) = \bar{J}_s(\mathbf{f}_k^{l_0}), \quad \forall l_0, l_1 \quad (53)$$

any sequence $\{\mathbf{f}_k | \mathbf{H}^t \mathbf{f}_k = \mathbf{q}_k\}$ (there could be infinitely many such sequences) satisfies $\bar{J}_s(\mathbf{f}_{k+1}) \geq \bar{J}_s(\mathbf{f}_k) \forall k$. Further, if $\mathbf{q}_k \in \mathcal{D}_i(\mathbf{q})$ and an update $\mathbf{q}_{k+1} \in \mathcal{D}_i(\mathbf{q})$, then $\{\mathbf{f}_k^{l_0}\} \subset \mathcal{D}_i(\mathbf{f})$, and $\{\mathbf{f}_{k+1}^{l_1}\} \subset \mathcal{D}_i(\mathbf{f})$. Similarly, it can be shown that a sequence $\{\mathbf{f}_k\}$ with $J_s(\mathbf{f}_{k+1}) \geq J_s(\mathbf{f}_k) (\forall k)$ has a corresponding sequence $\{\mathbf{q}_k | \mathbf{q}_k = \mathbf{H}^t \mathbf{f}_k\}$ with $J_s(\mathbf{q}_{k+1}) \geq J_s(\mathbf{q}_k)$ and if $\{\mathbf{f}_k\} \subset \mathcal{D}_i(\mathbf{f})$, then $\{\mathbf{q}_k\} \subset \mathcal{D}_i(\mathbf{q})$. This completes the proof. \square

APPENDIX B
 PROOF OF LEMMA 2

Applying Lagrange multiplier theory to the constrained maximization in (25), we obtain the maxima as shown in (27) [10].

To prove the DoA part, consider the gradient implementation of the power-constrained constant modulus criterion. Let $L(\mathbf{q}) = 1/4\|\mathbf{q}\|_4^4$ and $\mathbf{q}^* = \phi_i \mathbf{e}_i$ ($\phi_i > 0$) be a maximum. Since Φ is diagonal, $\|\Phi^{-(1/2)}\mathbf{q} - \mathbf{e}_i\| < \|\Phi^{-(1/2)}\mathbf{q} - \mathbf{e}_j\|$ is equivalent to $\phi_i q_i > \phi_j q_j$. For any k th iteration $\mathbf{q}_k \in \mathcal{D}_i(\mathbf{q})$, i.e., $\phi_i q_{i,k} > \phi_j q_{j,k}$ ($\forall j \neq i$), we can update it by first maximizing the cost function

$$\begin{aligned} \bar{\mathbf{q}}_{k+1} &= \mathbf{q}_k + \mu \Phi^{-1} \nabla L(\mathbf{q}_k) \\ &= \mathbf{q}_k + \mu \begin{pmatrix} q_{1,k}^3 \phi_1^2 \\ \vdots \\ q_{N,k}^3 \phi_N^2 \end{pmatrix}, \quad \mu > 0 \end{aligned} \quad (54)$$

and then by projecting $\bar{\mathbf{q}}_{k+1}$ back to the ellipsoid

$$\mathbf{q}_{k+1} = \frac{\bar{\mathbf{q}}_{k+1}}{\sqrt{\alpha(\mu)}}, \quad \alpha(\mu) = \sum_{l=1}^N \left(\frac{q_{l,k}}{\phi_l} + \mu q_{l,k}^3 \phi_l \right)^2. \quad (55)$$

We note that i) $\mathbf{q}_{k+1} \in \mathcal{D}_i(\mathbf{q})$. This is because

$$\begin{aligned} \|\mathbf{q}_{k+1}\|_{\Phi}^2 &= \frac{1}{\alpha(\mu)} \bar{\mathbf{q}}_{k+1}^t \Phi \bar{\mathbf{q}}_{k+1} \\ &= \frac{1}{\alpha(\mu)} \sum_{l=1}^N \left(\frac{q_{l,k} + \mu q_{l,k}^3 \phi_l^2}{\phi_l} \right)^2 = 1 \end{aligned} \quad (56)$$

and

$$\begin{aligned} \phi_i q_{i,k+1} &= \frac{1}{\sqrt{\alpha(\mu)}} (\phi_i q_{i,k} + \mu \phi_i^3 q_{i,k}^3) \\ &> \frac{1}{\sqrt{\alpha(\mu)}} (\phi_j q_{j,k} + \mu \phi_j^3 q_{j,k}^3) \\ &= \phi_j q_{j,k+1} \end{aligned} \quad (57)$$

if $\phi_i q_{i,k} > \phi_j q_{j,k}$. ii) $\|\mathbf{q}_{k+1}\|_4^4 > \|\mathbf{q}_k\|_4^4$. Let

$$\begin{aligned} g(\mu) &= \|\mathbf{q}_{k+1}\|_4^4 - \|\mathbf{q}_k\|_4^4 \\ &= \frac{1}{\alpha^2(\mu)} \sum_l (q_{l,k} + \mu q_{l,k}^3 \phi_l^2)^4 - \sum_l q_{l,k}^4. \end{aligned} \quad (58)$$

We will drop the subscript k in the following calculation for convenience. Since $\alpha(0) = 1$, $g(0) = 0$. We thus have (59), shown at the bottom of the page. Let $(q_l/\phi_l)^2 = p_l$. Then, $\sum_l p_l = 1$ due to $\|\mathbf{q}\|_{\Phi}^2 = 1$. Let $t \in \{t_1, \dots, t_N\}$ and

$t_l = p_l \phi_l^4$ with probability p_l , $l = 1, \dots, N$. By Jensen's inequality $E\{f(t)\} > f(E\{t\})$, if $f(t)$ is strictly convex, we have

$$\begin{aligned} \sum_l q_l^6 \phi_l^2 &= \sum_l \left(\frac{q_l}{\phi_l} \right)^4 \left(\frac{q_l}{\phi_l} \right)^2 \phi_l^8 \\ &= \sum_l (p_l^2 \phi_l^8) p_l = E\{t^2\} > (E\{t\})^2 \\ &= \left(\sum_l p_l \phi_l^4 p_l \right)^2 = \left(\sum_l q_l^4 \right)^2 \end{aligned} \quad (60)$$

which implies $\partial g(\mu)/\partial(\mu) \big|_{\mu=0} > 0$. Hence, $g(\mu) > 0$ for sufficiently small μ , i.e., $\|\mathbf{q}_{k+1}\|_4^4 > \|\mathbf{q}_k\|_4^4$. Steps i) and ii) show that for any $\mathbf{q}_k \in \mathcal{D}_i(\mathbf{q})$, by using a gradient ascent algorithm with an arbitrarily small step size, $\mathbf{q}_{k+1} \in \mathcal{D}_i(\mathbf{q})$, and $L(\mathbf{q}_{k+1}) > L(\mathbf{q}_k)$. Since there is one and only one maximum $\mathbf{q}^* = \phi_i \mathbf{e}_i \in \mathcal{D}_i(\mathbf{q})$, and $L(\mathbf{q})$ is closed and bounded on the sphere $\|\mathbf{q}\|_{\Phi}^2 = 1$, $\mathbf{q}_k \rightarrow \mathbf{q}^*$, and $\mathcal{D}_i(\mathbf{q})$ is the DoA of \mathbf{q}^* .

If there exists a point $\hat{\mathbf{q}}$ both in $\mathcal{D}_l(\mathbf{q})$ and in $\mathcal{D}_m(\mathbf{q})$, then $\hat{\mathbf{q}} \in \mathcal{D}_l(\mathbf{q})$ implies $\|\Phi^{-(1/2)}\hat{\mathbf{q}} - \mathbf{e}_l\| < \|\Phi^{-(1/2)}\hat{\mathbf{q}} - \mathbf{e}_m\|$, whereas $\hat{\mathbf{q}} \in \mathcal{D}_m(\mathbf{q})$ implies $\|\Phi^{-(1/2)}\hat{\mathbf{q}} - \mathbf{e}_m\| < \|\Phi^{-(1/2)}\hat{\mathbf{q}} - \mathbf{e}_l\|$, which contradicts each other. Therefore, if $l \neq m$, $\mathcal{D}_l(\mathbf{q}) \cap \mathcal{D}_m(\mathbf{q}) = \emptyset$. Those points satisfying $\|\Phi^{-(1/2)}\hat{\mathbf{q}} - \mathbf{e}_l\| = \|\Phi^{-(1/2)}\hat{\mathbf{q}} - \mathbf{e}_m\|$ for $l \neq m$ are on the common boundary of $\mathcal{D}_l(\mathbf{q})$ and $\mathcal{D}_m(\mathbf{q})$. \square

 APPENDIX C
 PROOF OF THEOREM 1

According to (29), SW receivers have the form $\mathbf{p}^{(i)} = \gamma_i \mathbf{e}_i$. Since $\mathbf{q} = \mathbf{\Pi}^{-1/4} \mathbf{p}$ and $\mathbf{f}^{(i)} = ((\mathbf{H}\mathbf{E}))^t \mathbf{q}^{(i)}$, we have

$$\begin{aligned} \mathbf{f}^{(i)} &= (\mathbf{H}^t)^t \mathbf{E}^{-1} \mathbf{\Pi}^{-1/4} \mathbf{p}^{(i)} = (\mathbf{H}^t)^t \gamma_i (\mathbf{E}^{-1} \mathbf{\Pi}^{-1/4}) \mathbf{e}_i \\ &= \frac{\phi_i}{\sqrt{\mathcal{E}_i}} (\mathbf{H}^t)^t \mathbf{e}_i. \end{aligned} \quad (61)$$

As for the derivation of the DoA for $\mathbf{f}^{(i)}$, we can follow a similar approach. The DoA of $\mathbf{p}^{(i)}$ is defined by

$$\begin{aligned} \mathcal{D}_i(\mathbf{p}) &= \left\{ \mathbf{p} \mid \|\mathbf{p}\|_{\Gamma}^2 = 1, \|\mathbf{\Gamma}^{-1/2} \mathbf{p} - \mathbf{e}_i\| \right. \\ &\quad \left. < \|\mathbf{\Gamma}^{-1/2} \mathbf{p} - \mathbf{e}_j\|, \forall j \neq i, j=1, \dots, N \right\}. \end{aligned} \quad (62)$$

Since

$$\begin{aligned} \mathbf{\Gamma}^{-1/2} \mathbf{p} &= \left(\mathbf{\Pi}^{-1/4} \Phi \mathbf{\Pi}^{-1/4} \right)^{-1/2} \mathbf{\Pi}^{1/4} \mathbf{q} \\ &= \mathbf{\Pi}^{1/2} \Phi^{-1/2} \mathbf{E} \mathbf{H}^t \mathbf{f} \end{aligned} \quad (63)$$

$$\begin{aligned} \left. \frac{\partial g(\mu)}{\partial \mu} \right|_{\mu=0} &= \frac{\left[4 \sum_l (q_l + \mu q_l^3 \phi_l^2)^3 q_l^3 \phi_l^2 \right] \alpha^2(\mu) - \left[\sum_l (q_l + \mu q_l^3 \phi_l^2)^4 \right] 2\alpha(\mu) \left[2 \sum_l \left(\frac{q_l}{\phi_l} + \mu q_l^3 \phi_l \right) q_l^3 \phi_l \right]}{\alpha^4(\mu)} \bigg|_{\mu=0} \\ &= 4 \left[\sum_l q_l^6 \phi_l^2 - \left(\sum_l q_l^4 \right)^2 \right]. \end{aligned} \quad (59)$$

combined with the fact that $\|\mathbf{p}\|_{\mathbf{F}}^2 = \|\mathbf{q}\|_{\mathbf{F}}^2 = \|\mathbf{f}\|_{\mathbf{R}}^2 = 1$, (32) follows. \square

REFERENCES

- [1] A. N. Akansu, P. Duhamel, X. Lin, and M. de Courville, "Orthogonal transmultiplexers in communication: A review," *IEEE Trans. Signal Processing*, vol. 46, pp. 979–995, Apr. 1998.
- [2] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communication," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 385–399, June 1980.
- [3] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*. New York: Springer-Verlag, 1987.
- [4] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
- [5] W. Chung, "Geometrical understanding of the constant modulus algorithm: Adaptive blind equalization and cross-polarized source separation," M.S. thesis, Cornell Univ., Ithaca, NY, 1999.
- [6] W. Chung and C. R. Johnson, "Characterization of the regions of convergence of CMA adapted blind fractionally spaced equalizer," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Asilomar, CA, 1998.
- [7] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Commun.*, pp. 1313–1327, Sept. 1991.
- [8] T. Ericson, "Structure of optimum receiving filters in data transmission systems," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 352–353, May 1971.
- [9] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, Nov. 1980.
- [10] M. Gu and L. Tong, "Geometrical characterizations of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2745–2756, Oct 1999.
- [11] I. Kalet, "The multitone channel," *IEEE Trans. Commun.*, vol. 37, pp. 119–124, Feb 1989.
- [12] P. Lancaster and T. Miron, *The Theory of Matrices*. New York: Academic, 1985.
- [13] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard (CMA) adaptive equalizers," *IEEE Trans. Signal Processing*, pp. 818–826, Apr. 1996.
- [14] P. Boaz, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [15] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Inform. Theory*, vol. IT-36, pp. 312–320, Mar. 1990.
- [16] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.
- [17] L. Tong, M. Gu, and S. Y. Kung, "A geometrical approach to blind signal estimation," in *Signal Processing Advances in Communications*, G. B. Giannakis, P. Stoica, Y. Hua, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2000, ch. 8.
- [18] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 459–472, Apr. 1983.
- [19] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. COMM-19, pp. 628–634, Oct 1971.
- [20] H. Zeng, L. Tong, and C. R. Johnson, "Relationships between CMA and Wiener receivers," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1523–1538, July 1998.



Ming Gu (S'98–M'01) received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree from the University of Connecticut, Storrs, CT, all in electrical engineering, in 1990, 1993, and 2000, respectively.

She joined Voyan Technology, Santa Clara, CA, in 2000. Currently she is engaged in R&D on digital signal processing in DSL and wireless communications.



Lang Tong (S'87–M'91) received the B.E. degree from Tsinghua University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Notre Dame, IN.

He was a Postdoctoral Research Affiliate with the Information Systems Laboratory, Stanford University, Stanford, CA, in 1991. Currently, he is an Associate Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. His areas of interest include statistical signal processing, adaptive receiver design for communication systems, signal processing for communication networks, and information theory.

Dr. Tong received the Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society.