Distributed MIMO Multicast With Protected Receivers: A Scalable Algorithm for Joint Beamforming and Nullforming

Amy Kumar, Raghuraman Mudumbai, *Member, IEEE*, Soura Dasgupta, *Fellow, IEEE*, Upamanyu Madhow, *Fellow, IEEE*, and D. Richard Brown III, *Senior Member, IEEE*

***Abstract*— We consider the problem of multicasting a common message signal from a distributed array of wireless transceivers by beamforming to a set of *beam targets*, while simultaneously protecting a set of *null targets* by nullforming to them. We describe a distributed algorithm in which each transmitter iteratively adapts its complex transmit weight using *common* aggregate feedback messages broadcast by the targets, and the *local* knowledge of only its own channel gains to the targets. This knowledge can be obtained using reciprocity without any explicit feedback. The algorithm minimizes the mean square error between the complex signal amplitudes at the targets and their desired values. We prove convergence of the algorithm, present geometric interpretations, characterize initializations that lead to minimum total transmit power, and prescribe designs for such initializations. We show that the convergence speed is nondecreasing in the number of transmitters N if a step size parameter is kept constant. For Rayleigh fading channels, as N goes to infinity: 1) convergence can be made arbitrarily fast and 2) beams and nulls can be achieved with *vanishing total transmit power* even with noise, both with probability one. These results add up to some remarkable scalability properties: the feedback overhead does not grow with the number of transmitters, and with high probability, the algorithm can be configured to converge arbitrarily fast and use vanishingly small total transmit power.**

***Index Terms*— Cooperative communication, distributed beamforming, distributed nullforming, interference management, local channel knowledge, scalability, virtual antenna arrays.**

Manuscript received November 15, 2015; revised June 20, 2016 and October 20, 2016; accepted October 23, 2016. Date of publication November 7, 2016; date of current version January 6, 2017. This work was supported in part by the U.S. NSF under Grant EPS-1101284, Grant ECCS-1150801, Grant CNS-P1329657, Grant CCF-1302456, Grant CCF-1302104, and Grant CCF-1319458 and in part by ONR under Grant N00014-13-1-0202. The associate editor coordinating the review of this paper and approving it for publication was C. Zhong.

A. Kumar and R. Mudumbai are with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: amy-kumar@uiowa.edu; rmudumbai@engineering.uiowa.edu).

S. Dasgupta is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA, and also with the Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center), Jinan 250101, China (e-mail: dasgupta@engineering.uiowa.edu).

U. Madhow is with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: madhow@ece.ucsb.edu).

D. R. Brown, III is with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: drb@wpi.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2016.2625784

I. Introduction

We CONSIDER the *distributed joint beamforming and nullforming* (JBNF) problem shown in Fig. 1, where N single antenna transmitters must broadcast a common message signal by forming beams towards each of M_1 single antenna receivers (*beam targets*), while simultaneously forming nulls at another set of $M - M_1$ receivers (*null targets*). Thus, the transmitters form a *virtual antenna array* and choose phases and amplitudes to shape the array's pattern such that beams and nulls are created at desired locations. By simultaneously transmitting beams and nulls, coherent combining gains can be achieved toward intended receivers while protecting unintended receivers. Some illustrative examples of applications where this capability would be useful are:

- **Electronic Warfare.** A transmit array broadcasts strong jamming signals [2] that disable an enemy's communication infrastructure, while shielding friendly cooperating stations. While enemy nodes are of course hostile, the friendly nodes can cooperate in this process to steer nulls to themselves thus shielding them from the jamming transmission.
- **Cognitive radio.** A transmit array acts as a secondary user of licensed spectrum seeking to communicate with a set of secondary receivers (beam targets) without causing interference at primary receivers (null targets). While this application does require the cooperation of the primary receiver with the secondary transmitters to steer nulls, the cooperation is of a simple kind very similar to methods considered in previous literature [3].

Other possible applications include wireless sensor networks where sensor nodes use beamforming to efficiently transmit observations to data collection nodes and cellular networks where Base Stations form a transmit array and coordinate their transmissions to avoid cochannel interference [4]. The common feature of these applications is the need for interference cancellation at specific locations, an objective at the very core of the JBNF problem. More generally, joint beam and nullforming may be viewed a fundamental building block for increased spatial spectrum reuse [9], and towards achieving MIMO spatial multiplexing gains [10] with distributed arrays. Specifically, a distributed array can send multiple streams of data simultaneously to different receivers without these streams interfering with each other by running multiple JBNF

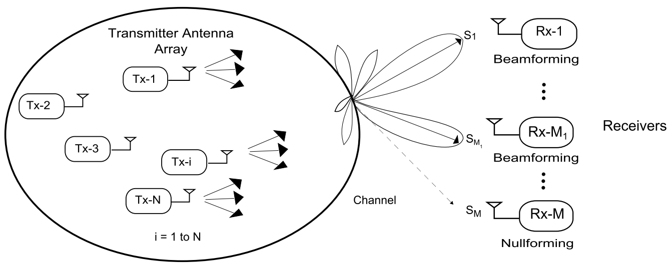


Fig. 1. Problem of joint beam and null-forming using a distributed array.

algorithms in parallel each pointing beams to one receiver while driving nulls to the others.

A. Contributions

We cast the JBNF objective as one of selecting the complex transmit weights of each transmitter to attain specified complex amplitudes that modulate the common message at the beam targets, and formulate it as an unconstrained optimization problem to minimize the mean square error between the achieved and desired modulating amplitudes at the receivers. A gradient descent minimization of this objective leads to a distributed, iterative algorithm with attractive scalability and robustness properties, a detailed analysis of which is the main focus of this paper. In this algorithm, each receiver broadcasts to all transmitters a common feedback message consisting of a single complex number representing the amplitude of the aggregate (total) baseband received signal in the previous iteration. The fact that this feedback is common to all the transmitters, rather than individualized, ensures scalability. Each transmitter adjusts its own complex weight (the magnitude and phase of its RF transmission) using only the feedback from the receivers and the knowledge of its own complex channel gains to each receiver.

While there is a rich literature on the analysis of gradient descent for quadratic cost functions in the adaptive control [29]–[36] and signal processing [38]–[40] literature, our rigorous analysis goes well beyond known results. These include a complete characterization of initializations that achieve power efficient solutions, scalability with respect to N, and almost sure asymptotic energy efficiency as $N \rightarrow \infty$. These results, are summarized as follows:

(a) Characterizing limit points: The JBNF has an entire affine subspace $\mathcal{H}$ of global minima as the optimization is underconstrained because there are more transmitters than receivers. Such a situation is very common in adaptive control and is often linked to the lack of persistent excitation (pe) [30]–[32]. Barring an analysis in a low dimensional case in [41], the literature only shows convergence to $\mathcal{H}$ without pinpointing the limit point further. *We are able to show (Theorem 1) that, in the noise-free setting, our algorithm converges to the projection of the initial iterate on $\mathcal{H}$.* With noise, convergence is to this same point in the mean with bounded variance (Theorem 2). *Demonstration of bounded steady state variance is a novelty in the adaptive systems literature that addresses similar underconstrained problems (see (c) below).*

(b) Power efficiency of the limit point: We characterize conditions under which the limit point corresponds to the solution with minimum transmit power: specifically, the initial iterate of the gradient descent algorithm must lie in the range space of a matrix comprising the complex channel gains. We prescribe practical methods for obtaining such an initialization.

(c) The absence of drift due to noise: If without noise an adaptive algorithm converges to an affine subspace like $\mathcal{H}$ rather than to an isolated point, then the possibility arises that noise may cause the adaptations to drift along $\mathcal{H}$ [42], [45], [46]. Such drift can cause unbounded residual variance. In adaptive control, this causes serious problems like bursting [43] and instability, and leakage is used to combat it [34], [44]. In the JBNF problem,such drift represents wasted transmit power. *However, we show that our algorithm avoids this problem entirely, as noise has no effect along $\mathcal{H}$.*

(d) Scalability: For a given set of beam and null targets, if we keep a step size parameter of the algorithm fixed, then the convergence rate of our algorithm increases with the number of transmitters N. Since each receiver sends a fixed amount of feedback in each iteration, this implies that the *total* feedback overhead of our algorithm does not grow with N, if we use reciprocity to obtain the additional local channel knowledge [17]. This is a significant improvement over methods that require global channel information where the number of channel coefficients to be learned grows proportionally with N.

(e) Asymptotic scalability and power efficiency: For i.i.d. Rayleigh fading channels as $N \rightarrow \infty$ and a fixed number of receivers, even stronger scalability properties can be established: with probability one, the convergence speed can be made arbitrarily fast. Further, if the algorithm is properly initialized, then with probability one the iterates converge in the mean to zero with zero error covariance. This implies that *the total transmit power across all transmitters becomes vanishingly small* (by virtue of coherent combining at the beam targets).

This latter fact has one important implication. An alternative formulation of the JBNF problem is: minimize the net transmit power, while achieving desired power levels (rather than specific complex amplitudes) at the beam targets. Conceptually, this non-convex problem allows the implicit selection of the phases of the target complex amplitudes by the optimization process and may lead to a more power efficient solution than our quadratic formulation. Our results show that our algorithm matches the non-convex alternative for large N almost surely.

B. Background and Related Work

Special cases of the JBNF problem have been considered in the recent literature, notably beamforming alone, or nullforming alone, from a transmit array to a single receiver. In distributed beamforming, the coherence gain is known to be robust to moderate errors in channel estimation [16]. In contrast, nullforming [18], [19] requires that the signals from all the transmitters cancel each other precisely at the receiver, making nullforming highly sensitive to errors, [19], and JBNF a much more challenging problem. The literature on the multicast problem [26]–[28] considers the nonconvex problem of minimizing the total transmit power required by an array to form only beams (no nulls) with a specified SNR at a number of receivers. Further, [26] considers a setting with more receivers than transmitters, whereas we consider the opposite scenario where a *large transmit array* communciates with a *small number of receivers*. Nullforming using *global channel knowledge* is studied in [18] and [19].

An iterative algorithm for nullforming to a single receiver is presented in [20] which, like our algorithm, also requires knowledge at each transmitter only of its own complex channel gain to the receiver in addition to common feedback from the receiver. The algorithm of [20] uses phase-only adaptation yielding a non-convex optimization. By allowing amplitude and phase adaptation and by targeting received complex amplitudes rather than power, we are able to consider a much simpler quadratic optimization framework while generalizing to multiple beam and null targets.

Interference avoidance techniques based on implicitly learning the nullspace to multiple receivers are studied in [3] and [4]. These probe the MIMO channel with different precoding vectors and observe indirect measures of the SINR at the users to which the nulls are steered to determine a nullforming precoding vector.

Our iterative, aggregate feedback based approach to JBNF is novel in conception and solution. Previous JBNF type work relies on a non-iterative, one-shot approach, assumes full CSI and calculates the weights directly from the full channel matrix [8]. A preliminary conference version of this paper is [1].

II. Formulation of the JBNF Problem

A. The Broad Problem, Approach and System Architecture

We consider a distributed array of N transmitters and M receivers. Receivers, labeled $1, \cdots, M_1$ are *beam targets*, and the rest labeled M_1+1 to M are *null targets*. In the k-th time slot all transmitters broadcast after precoding, a common complex baseband signal m_k. At the null targets the received signal must be zero. At the j-th beam target the received baseband signal must be $m_k b_j$, for specified complex $b_j \neq 0$. Knowing b_j, the beam target can recover the transmitted message form $m_k b_j$. Thus the common message m_k is *multicast* to the beam targets, while *protecting the null targets*. Our goal is to design a distributed iterative algorithm that achieves this objective asymptotically in time.

To this end at the k-th time slot or iteration the i-th transmitter selects a complex transmit weight $x_i^*[k]$, and transmits the precoded message signal $x_i^*[k]m_k$. Assume the *aggregate* complex baseband received signal at the j-th receiver in this time slot is $r_j[k]m_k$. The goal is to iteratively drive $r_j[k]$ to b_j at the beam targets and to zero at the null targets. To achieve this goal, at the beginning of time slot $k+1$, the i-th transmitter

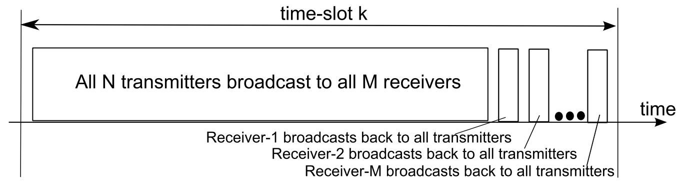


Fig. 2. Description of transmission in time-slot k.

adjusts $x_i[\cdot]$ to iteratively minimize the quadratic cost function:

$$|m_k|^2 \left\{ \sum_{j=1}^{M_1} \left| r_j[k] - b_j \right|^2 + \sum_{j=M_1+1}^{M} \left| r_j[k] \right|^2 \right\}. \quad (1)$$

The minimization of this cost function is equivalent to achieving $r_j[k] = b_j$ at the j-th beam target and $r_j[k] = 0$ at the j-th null target, thereby achieving our stated multicasting objective. There are two novel aspects in our approach to this minimization.

1) Aggregate Feedback: At the end of time slot k, the receivers broadcast back in a time interlaced fashion the conjugate of a sample of the aggregate received baseband signal $m_k^* s_j[k] = \left(m_k r_j[k]\right)^*$. Thus in each time slot the j-th receiver broadcasts this single complex number. The transmitters use this feedback to adjust the transmit weights at the next time slot $k+1$. We call the samples $m_k^* s_j[k]$ the *aggregate feedback* from the j-th receiver. In our theoretical analysis we assume these are of infinite precision. In our simulations in Section V, we represent the real and imaginary parts as double-precision floating point numbers; this requires 128 bits to encode one complex number, which represents a negligible fraction of the payload sizes used in most modern packet wireless networks e.g. WiFi. While we defer a detailed analysis of the effects of quantized feedback to future work, our preliminary numerical studies suggest that the feedback can be encoded with substantially reduced precision without performance loss.

The underlying architecture is depicted in Figure 2, where the k-th time slot, is itself subdivided into $M+1$ subslots. The long slot at the beginning represents the actual multicast transmission from the array. The receivers take turns in the remaining M shorter time slots to send their aggregate feedback information to the array nodes. At the end of each time slot, the nodes update their precoding gains in the manner described in the sequel.

2) Distributed Updates: Adjustments of the precoding gains $x_i[k]$ are effected through the *distributed gradient descent minimization* of (1). Specifically, the i-th transmitter adjusts $x_i[k]$, using the aggregate feedback samples $m_k^* s_j[k]$ broadcast by the receivers, and *local channel information* i.e. the knowledge of its *own* channel gains to the M receivers. Unlike say [8] it does not need the channel gains of the other transmitters. Item (III) below describes how this can be done, though the acquisition of this information is beyond the scope of this paper.

We now elaborate on some implicit assumptions.

(I) **Time-slotting.** We assume a synchronous time-slotted network with the time-division multiplexed schedule of transmissions shown in Fig. 2. This can be achieved using standard network synchronization protocols such as [22]. In addition, the transmit nodes require timing synchronization accurate enough to guarantee that at any time instant all the array nodes are transmitting the same message symbols; this requires timing errors to be much smaller than a symbol duration.

(II) **Slowly varying phase offsets.** We assume that channel gains and oscillator offsets are roughly constant over several iterations to allow the JBNF algorithm to track any channel variations. With time-slots of duration ≈ 50 ms, this requires that the channels and offsets are roughly constant over several seconds. This can be assured by using standard filtering techniques [7] that dynamically track Doppler and clock dynamics.

(III) **Local channel state information.** We assume that each transmitter knows its own complex channel gains to each of the receivers. Means of obtaining this information is not the subject of this paper. However, we do observe that one particularly appealing way of acquiring this local channel knowledge is for each transmitter to observe incoming transmissions from the receivers (e.g., the packets carrying the global aggregate feedback) and then use channel reciprocity. Detailed algorithms to obtain local channel knowledge using reciprocity have been reported in our recent work [17].

B. Technical Problem Statement

Let $\boldsymbol{H} = [\boldsymbol{h}_1\ \boldsymbol{h}_2\ \ldots\ \boldsymbol{h}_M]$ be the $N \times M$ channel matrix whose ij-th entry h_{ij} is the complex channel gain from the i-th transmitter to the j-th receiver, and $\boldsymbol{h}_j$ be the j-th column of $\boldsymbol{H}$, corresponding to the channel vector to receiver j. Define $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$ as the $N \times 1$ vector of the transmit precoding weights. Thus, the noiseless complex baseband signal received at the jth receiver at the k-th slot is $r_j[k] = m_k\left(\boldsymbol{x}^H \boldsymbol{h}_j\right)$.

As $|m_k|^2$ is common to each summand in (1), without loss of generality, we set $m_k \equiv 1$; i.e. the total complex baseband signal seen by receiver j in time slot k is $r_j[k] = \boldsymbol{x}^H[k]\boldsymbol{h}_j$. The complex number $s_j[k] = r_j^*[k] = \boldsymbol{h}_j^H \boldsymbol{x}[k]$ is broadcast by receiver j to all the transmitters. Recall that these samples $s_j[k]$ constitute the common *aggregate feedback* used by each transmitter to implement the JBNF algorithm given in Section II-C, in a distributed fashion.

The vector of feedback signals broadcast by all of the receivers in time slot k is

$$\boldsymbol{s}[k] = \begin{bmatrix} s_1[k] & \cdots & s_M[k] \end{bmatrix}^T = \boldsymbol{H}^H \boldsymbol{x}[k] + \boldsymbol{w}[k] \tag{2}$$

where $\boldsymbol{w}[k] = [w_1[k], \ldots, w_M[k]]^T \sim \mathcal{CN}\left(0, \sigma_w^2 I\right)$, $\forall k$ represents complex Gaussian noise assumed to be i.i.d. across receivers and time slots.

We wish to adapt $\boldsymbol{x}[k]$ in a *distributed fashion* so that $\boldsymbol{x}^H[k]\boldsymbol{h}_j$ are driven towards specified nonzero values b_j for beam targets $1 \le j \le M_1$, and towards zero for null targets $M_1 + 1 \le j \le M$. Call $b_j = 0$ for all $j \in \{M_1 + 1, \cdots, M\}$, and $\boldsymbol{b} = [b_1, \ldots, b_M]^T$. Then, (1) is just $\|\boldsymbol{s}[k] - \boldsymbol{b}\|^2 = \left\|\boldsymbol{H}^H \boldsymbol{x}[k] - \boldsymbol{b}\right\|^2$.

To accommodate noise the JBNF problem can be recast as find $\boldsymbol{x}$ to minimize the quadratic cost function

$$\begin{aligned} J_w(\boldsymbol{x}) &= E_{\boldsymbol{w}}\left[\|\boldsymbol{s} - \boldsymbol{b}\|^2\right] = E_{\boldsymbol{w}}\left[\left\|\boldsymbol{H}^H \boldsymbol{x} + \boldsymbol{w} - \boldsymbol{b}\right\|^2\right] \\ &= \left\|\boldsymbol{H}^H \boldsymbol{x} - \boldsymbol{b}\right\|^2 + M\sigma_w^2. \end{aligned} \tag{3}$$

The minimization of J_w is equivalent to that of

$$J(\boldsymbol{x}) = \|\boldsymbol{H}^H \boldsymbol{x} - \boldsymbol{b}\|^2. \tag{4}$$

Section II-C shows how this minimization can be achieved in a scalable, distributed fashion.

We contrast our approach both to that of [26], which has only beam tragets, and no null targets, and to traditional beamforming, to expose another novelty of our approach. Traditional beamforming as in [26], simply imposes a minimum *power* constraint at a beam target. Such an approach leads to nonconvex optimization problems. While $|b_j|$ implicitly specifies the desired beam power at the j-th beam target, by specifying the phase of the desired b_j rather than just its magnitude, we have over-constrained the problem. This, however, leads to a key advantage: the resulting cost in (4) is *convex* with all its associated advantages. In addition, as discussed in Section III-A, and verified by simulations in Section V, the loss of optimality in setting target phases to arbitrary values vanishes as the number of cooperating nodes increases.

We would also like to note that the objective function in our formulation is the aggregate Mean Squared Error (MSE) which represents the sum of squared deviations of the achieved signal power level (RSS) at each receiver from the desired value. Zero aggregate MSE would mean all individual MSEs are zero and QoS at each individual receiver is deterministically guaranteed. We will show that our algorithm achieves this in the absence of noise. In the presence of noise, while a deterministic QoS guarantee at each individual receiver is no longer possible, very good statistical guarantees can still be obtained. For example, consider a 10 node array with one beam and one null target with noise power at the receivers as -40 dB compared to the desired signal level of 0 dB at the beam target. Our results in Section V show that our algorithm reduces the aggregate MSE to close to the noise floor i.e. -40 dB, which serves as an upper bound for the individual MSEs at both receivers. This means that at the null target, the actual received power from the array fluctuates close to the noise floor which effectively makes the signal indistinguishable from noise. At the beam target, the received signal fluctuates around the desired power level of 0 dB with a variance of -40 dB. Assuming the fluctuations are normally distributed, with greater than 99% probability, the received power at the beam target is in the range of -0.2 dB to 0.2 dB. Thus at the beam target, with 99% probability, the received signal is strong enough to achieve a SNR level between 39.8 dB and 40.2 dB, and similar guarantees apply at every individual receiver in all JBNF problems.

C. A Distributed Algorithm

Using (2), the gradient of $J(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ can be written as

$$\nabla \boldsymbol{J}^{\boldsymbol{H}}(\boldsymbol{x}) = \left(\boldsymbol{H}(\boldsymbol{H}^H\boldsymbol{x} - \boldsymbol{b})\right)^H = (\boldsymbol{H}(\boldsymbol{s} - \boldsymbol{b}))^H .$$

Thus, the gradient descent minimization of $J(\boldsymbol{x})$ is for a suitably small step size $\mu > 0$

$$\begin{aligned}\boldsymbol{x}[k+1] &= \boldsymbol{x}[k] - \mu \nabla \boldsymbol{J}(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}[k]} \\ &= \boldsymbol{x}[k] - \mu \boldsymbol{H}(\boldsymbol{s}[k] - \boldsymbol{b}). \end{aligned} \quad (5)$$

This is similar to the LMS algorithm which too is a gradient descent algorithm. The key difference is that for LMS the role of $\boldsymbol{H}$ is played by a time varying regression vector. The adaptation (5) for transmitter i can be written as

$$x_i[k+1] = x_i[k] - \mu \sum_{j=1}^{M} h_{ij} \left(s_j[k] - b_j\right). \quad (6)$$

Thus to adapt its weight at slot $k+1$, transmitter i only requires from each receiver j the aggregate feedback sample $s_j[k]$, i.e. the total received baseband sample at receiver j, and knowledge of its own channel gains to the receivers h_{ij}, $\forall j$. In particular, transmitter i does *not* need the channel gains of other transmitters h_{mj}, $m \neq i$.

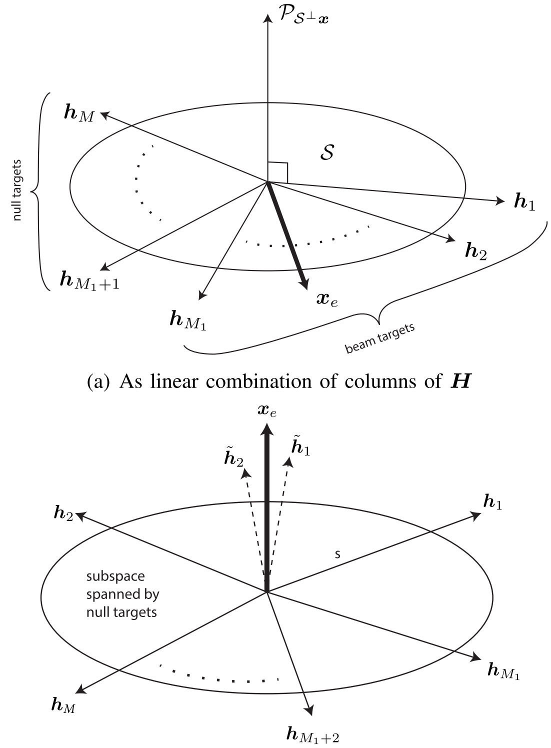
III. Analytical Characterization

In this section, we discuss the geometric structure of the problem and investigate the convergence properties of the iterative algorithm (5). Section III-A provides the geometrical perspective, focusing on the power efficient solution i.e. the choice of transmit precoding weights $\boldsymbol{x}$ that satisfies the JBNF constraints with the minimum total transmit power.

Section III-B analyzes convergence in the noiseless case. As $J(\boldsymbol{x})$ is convex, and in the absence of noise (5) is an exact gradient descent, the convergence of $\boldsymbol{x}[k]$ follows trivially. However, in this case there is an affine subspace $\mathcal{H}$ (see (9)) of minimizing solutions. Precise characterization of the limit point is not in general available in the literature. Theorem 1 precisely characterizes the limit point as being the projection of $\boldsymbol{x}[0]$ on $\mathcal{H}$. This is used to provide a design prescription that makes this limit point power efficient.

Section III-C is on the effect of noise. Theorem 2 shows that the limit point of the mean of the iterates is identical to those in Theorem 1. We also provide a clean expression for the limiting error covariance. We note in subsequent discussions that similar expressions for the error covariance for gradient descent laws like LMS make simplifying assumptions that we are able to avoid. As importantly, even though there is a nontrivial affine subspace $\mathcal{H}$ on which $J(\boldsymbol{x})$ is minimized the residual noise covariance is still bounded. This does not occur without an additional leakage term in algorithms like LMS in applications that have affine subspaces of minimizing solutions. To study the issues of scalability, we discuss in Section III-D convergence rates and their relation to μ.

Since we are interested in scaling to a large number of transmitters N, we focus on the regime $N > M$ (distributed array size larger than the number of receivers), and indeed,



(a) As linear combination of columns of $\boldsymbol{H}$

(b) As linear combination of columns of $\tilde{\boldsymbol{H}}_{beam}$

Fig. 3. Geometric interpretations of the power efficient solution.

on $N \gg M$. With high probability, therefore, the $N \times 1$ channels $\{\boldsymbol{h}_j, 1 \leq j \leq M\}$ are linearly independent. For most of our analysis, we make the latter assumption, stated formally below.

Assumption 1: The $N \times M$ channel matrix $\boldsymbol{H}$ has full column rank.

This assumption on the tall matrix $\boldsymbol{H}$ implies that the $M \times M$ correlation matrix $\boldsymbol{H}^H\boldsymbol{H}$ is full rank and positive definite. We denote its ordered eigenvalues by $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_M > 0$.

A. Geometric Interpretation of Optimum Solution

Under Assumption 1, with $\mathcal{R}(\boldsymbol{H}^H)$ denoting the range space of $\boldsymbol{H}^H$, one has $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{H}^H)$, which guarantees the existence of $\boldsymbol{x}$ such that (4) is zeroed out:

$$\boldsymbol{H}^H\boldsymbol{x} = \boldsymbol{b}. \quad (7)$$

Since this is an underdetermined system in the typical regimes of interest ($N > M$), there exists an entire *affine subspace* of vectors satisfying (7), from which we would like to choose the power efficient solution, $\boldsymbol{x}_e = \arg\min_{\boldsymbol{x}\in\mathcal{H}} \|\boldsymbol{x}\|$ with $\mathcal{H} = \left\{\boldsymbol{x} \in \mathbb{C}^N \middle| \boldsymbol{H}^H\boldsymbol{x} = \boldsymbol{b}\right\}$.

Under Assumption 1, the $M \times M$ matrix $\boldsymbol{H}^H\boldsymbol{H}$ is invertible, and the unique power efficient solution is given by

$$\boldsymbol{x}_e = \boldsymbol{H}\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{b} = \boldsymbol{H}\boldsymbol{a}. \quad (8)$$

Observe that the power efficient solution $\boldsymbol{x}_e$ must lie in the *signal space* $\mathcal{S} = \mathcal{R}(\boldsymbol{H})$ spanned by the columns $\boldsymbol{h}_1, \cdots, \boldsymbol{h}_M$ of $\boldsymbol{H}$. Fig.3-(a) illustrates this geometric interpretation of $\boldsymbol{x}_e$

We can now completely characterize the affine subspace $\mathcal{H}$ of solutions as

$$\mathcal{H} = \{\boldsymbol{x} : \boldsymbol{x} - \boldsymbol{x}_e \in \mathcal{S}^{\perp}\} \tag{9}$$

where $\mathcal{S}^{\perp} = \mathcal{N}(\boldsymbol{H}^H)$ is the null space of $\boldsymbol{H}^H$ [49]. To see this, note that for any $\boldsymbol{x} \in \mathcal{H}$, $\boldsymbol{H}^H(\boldsymbol{x} - \boldsymbol{x}_e) = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$. Thus, the affine subspace of solutions to (7) is the translation of the "undesired" subspace $\mathcal{S}^{\perp}$ by the power efficient solution $\boldsymbol{x}_e$ (or indeed, by any solution $\boldsymbol{x}$ of (7)). We will characterize how the particular solution in this affine subspace that the iteration (5) converges to depends on the initial condition. Before that, we provide an alternative geometric characterization of the power efficient solution $\boldsymbol{x}_e$, working within the signal space $\mathcal{S}$.

1) Alternative Characterization of Power Efficient Solution: This approach treats the beam and null targets separately. Since $\boldsymbol{h}_i^H \boldsymbol{x} = 0$ for $i = M_1 + 1, \ldots, M$, we must have $\boldsymbol{x}$ orthogonal to the subspace $\mathcal{S}_n$ spanned by $\boldsymbol{h}_{M_1+1}, \ldots \boldsymbol{h}_M$. For $j = 1, \ldots, M_1$, let $\tilde{\boldsymbol{h}}_j = \boldsymbol{P}_{\mathcal{S}_n}^{\perp} \boldsymbol{h}_j$ denote the projection of beam target vector $\boldsymbol{h}_j$ orthogonal to this subspace $\mathcal{S}_n$. We can therefore write the beam target equations as $\boldsymbol{h}_j^H \boldsymbol{x} = \tilde{\boldsymbol{h}}_j^H \boldsymbol{x} = b_j$, $j = 1 \ldots M_1$. This can be written in vector form as

$$\tilde{\boldsymbol{H}}_{beam}^H \boldsymbol{x} = \boldsymbol{b}_{beam} \tag{10}$$

where $\tilde{\boldsymbol{H}}_{beam}$ is an $N \times M_1$ matrix with columns $\tilde{\boldsymbol{h}}_1, \ldots, \tilde{\boldsymbol{h}}_{M_1}$, and $\boldsymbol{b}_{beam} = [b_1, \ldots, b_{M_1}]^T$ corresponds to the complex conjugates of the desired complex amplitudes at the beam targets.

Reasoning as before, the minimum norm solution lies in $\mathcal{R}(\tilde{\boldsymbol{H}}_{beam}^H)$, and can be written as $\boldsymbol{x}_e = \tilde{\boldsymbol{H}}_{beam} \boldsymbol{a}_{beam,opt}$ with

$$\boldsymbol{a}_{beam,opt} = \left(\tilde{\boldsymbol{H}}_{beam}^H \tilde{\boldsymbol{H}}_{beam}\right)^{-1} \boldsymbol{b}_{beam}$$

where we have used the fact that $\tilde{\boldsymbol{h}}_1, \ldots, \tilde{\boldsymbol{h}}_{M_1}$ are linearly independent under Assumption 1. Fig. 3-(b) illustrates this alternative geometric interpretation of $\boldsymbol{x}_e$ for an example case of two beam targets. It shows the projections of the beam target vectors, $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$, orthogonal to the subspace $\mathcal{S}_n$, as $\tilde{\boldsymbol{h}}_1$ and $\tilde{\boldsymbol{h}}_2$, and how the power efficient solution $\boldsymbol{x}_e$ is a linear combination of $\tilde{\boldsymbol{h}}_1$ and $\tilde{\boldsymbol{h}}_2$.

Remark: As long as Assumption 1 holds, a solution exists for *any* choice of target complex amplitudes $\boldsymbol{b}_{beam}$. The minimum transmit power, corresponding to the minimum norm solution $\boldsymbol{x}_e$, is

$$P_{TX} = \boldsymbol{x}_e^H \boldsymbol{x}_e = \boldsymbol{b}_{beam}^H \left(\tilde{\boldsymbol{H}}_{beam}^H \tilde{\boldsymbol{H}}_{beam}\right)^{-1} \boldsymbol{b}_{beam}$$

This depends on the target complex amplitudes $\boldsymbol{b}_{beam}$, hence in principle, we could choose these complex amplitudes to further optimize the value of the minimum transmit power. Fixing $||\boldsymbol{b}_{beam}||$, the minimum possible value is attained by choosing $\boldsymbol{b}_{beam}$ along the eigenvector corresponding to the minimum eigenvalue of $\left(\tilde{\boldsymbol{H}}_{beam}^H \tilde{\boldsymbol{H}}_{beam}\right)^{-1}$. However, such a choice of $\boldsymbol{b}_{beam}$ might not be admissible, because we may wish to constrain the magnitudes of the entries of $\boldsymbol{b}_{beam}$ to some values based on the desired SNR at each beam location, and hence may only have control on the phases of the entries. Furthermore, the resulting possible reduction in transmit power is upper-bounded by the condition number of $\tilde{\boldsymbol{H}}_{beam}$ which in the regime $N \gg M$ is likely to be minor since we expect the eigenvalue spread of $\tilde{\boldsymbol{H}}_{beam}^H \tilde{\boldsymbol{H}}_{beam}$ (and hence that of its inverse) to be small. We explore this issue further in Section V with numerical simulations (see Figure 7 and associated discussion), and generally focus on the fixed, arbitrary choice $\boldsymbol{b}_{beam} = \boldsymbol{1}$ for the rest of the paper.

B. Convergence in Noiseless Regime

Behavior of the updates (5) in the noiseless case is provided in Theorem 1 below. A few features are instructive. First observe that, while it is well known that convergence must occur to the affine subspace $\mathcal{H}$, existing analyses of such algorithms fail to characterize the the precise limit point $\boldsymbol{x}_\infty$ on $\mathcal{H}$. In contrast this theorem proves that this limit point is in fact *the projection of the intial iterate* $\boldsymbol{x}[0]$ *on* $\mathcal{H}$. Design implications of this fact is described after the theorem.

The second fact that impacts subsequent noise analysis is as follows. Consider the *error vector* defined as $\boldsymbol{\Delta}[k] \doteq \boldsymbol{x}[k] - \boldsymbol{x}_\infty$. It is shown in the proof that this vector evolves according to:

$$\boldsymbol{\Delta}[k+1] = \boldsymbol{\Delta}[k] - \mu \boldsymbol{H}\boldsymbol{H}^H \boldsymbol{\Delta}[k] = \left(\boldsymbol{I} - \mu \boldsymbol{H}\boldsymbol{H}^H\right) \boldsymbol{\Delta}[k] \tag{11}$$

Observe as $\boldsymbol{H}$ is tall, the transition matrix $\left(\boldsymbol{I} - \mu \boldsymbol{H}\boldsymbol{H}^H\right)$ has eigenvalues at 1, representing modes that are orthonormal to $\mathcal{H}$, and do not decay. To facilitate the convergence analysis, the theorem in fact proves that that the $N \times 1$ vector $\boldsymbol{\Delta}[k]$ lies in the signal subspace, and can therefore be expressed in terms of a lower dimensional $M \times 1$ vector $\boldsymbol{\delta}[k]$ as

$$\boldsymbol{\Delta}[k] = \boldsymbol{H}\boldsymbol{\delta}[k] \tag{12}$$

This vector on the other hand evolves as

$$\begin{aligned} \boldsymbol{H}\boldsymbol{\delta}[k+1] &= \left(\boldsymbol{I}_N - \mu \boldsymbol{H}\boldsymbol{H}^H\right) \boldsymbol{H}\boldsymbol{\delta}[k] \\ &= \boldsymbol{H}\left(\boldsymbol{I}_M - \mu \boldsymbol{H}^H \boldsymbol{H}\right) \boldsymbol{\delta}[k] \end{aligned}$$

where we put subscripts on the identity matrices to specify their dimension. As $\boldsymbol{H}$ has full column rank, this becomes

$$\boldsymbol{\delta}[k+1] = \left(\boldsymbol{I}_M - \mu \boldsymbol{H}^H \boldsymbol{H}\right) \boldsymbol{\delta}[k] \tag{13}$$

and the *de facto* transition matrix $\boldsymbol{I}_M - \mu \boldsymbol{H}^H \boldsymbol{H}$ does not have eigenvalues at 1. Note (13) reflects movement along the signal subspace. We explain later why this reduced state space has important implications to the convergence analysis in the presence of noise.

Theorem 1: Consider (5) under (2) and Assumption 1. With $\boldsymbol{w}[k] \equiv 0$*, and*

$$|1 - \mu\lambda_i| < 1, \quad i \in \{1, \cdots, M\}. \tag{14}$$

the weight sequence exponentially converges to $\boldsymbol{x}_\infty$*:*

$$\begin{aligned} \boldsymbol{x}_\infty &= \lim_{k\to\infty} \boldsymbol{x}[k] = \boldsymbol{x}_e + \mathbf{P}_{\mathcal{S}}^{\perp} \boldsymbol{x}[0] = \boldsymbol{x}_e + \boldsymbol{x}[0] - \mathbf{P}_{\mathcal{S}} \boldsymbol{x}[0] \\ &= \boldsymbol{x}_e + \boldsymbol{x}[0] - \boldsymbol{H}(\boldsymbol{H}^H \boldsymbol{H})^{-1} \boldsymbol{H}^H \boldsymbol{x}[0] \end{aligned} \tag{15}$$

Equivalently, we can express this limit as the projection of $\boldsymbol{x}[0]$ *onto the affine subspace* $\mathcal{H}$*:*

$$\boldsymbol{x}_\infty = Proj_{\mathcal{H}}(\boldsymbol{x}[0]) \tag{16}$$

Proof: Recall that any vector $\boldsymbol{x} \in \mathbb{C}^N$ can be expressed as the sum of its projection onto $\mathcal{S} = \mathcal{R}(\boldsymbol{H})$ and its orthogonal complement, i.e., $\boldsymbol{x} = \mathbf{P}_{\mathcal{S}}\boldsymbol{x} + \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}$. Observe that the update term in (5) lies in the signal space $\mathcal{S}$, for any values of the feedback vector $\boldsymbol{s}[k]$ and desired complex amplitudes $\boldsymbol{b}$. Thus, the component of $\boldsymbol{x}[0]$ orthogonal to $\mathcal{S}$ is unaffected by the iterations. Decomposing the weight sequence into its projection in the signal space and orthogonal to it, as $\boldsymbol{H}(\boldsymbol{s}[k]-\boldsymbol{b}) \in \mathcal{S}$, we may rewrite the iteration as follows:

$$\begin{aligned} \mathbf{P}_{\mathcal{S}}\boldsymbol{x}[k+1] &= \mathbf{P}_{\mathcal{S}}\boldsymbol{x}[k] - \mu\boldsymbol{H}(\boldsymbol{s}[k]-\boldsymbol{b}) \\ \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[k+1] &= \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[k] \equiv \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0] \end{aligned} \tag{17}$$

Thus, the component orthogonal to the signal space remains unchanged at $\mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0]$. On the other hand, the component restricted to the signal space is follows gradient descent on a quadratic cost function with a unique global minimum $\boldsymbol{x}_e$, and therefore converges to $\boldsymbol{x}_e$. To see this, without presuming the existence of a limit point, and treating $\boldsymbol{x}_\infty$ as the well defined vector on the right hand side of (15), consider $\boldsymbol{\Delta}[k] = \boldsymbol{x}[k]-\boldsymbol{x}_\infty$. From (17), $\boldsymbol{x}[k]-\boldsymbol{x}[0] = \mathbf{P}_{\mathcal{S}}(\boldsymbol{x}[k]-\boldsymbol{x}[0])$. Thus,

$$\begin{aligned} \boldsymbol{\Delta}[k] &= \boldsymbol{x}[k] - \boldsymbol{x}_\infty \\ &= \boldsymbol{x}[k] - (\boldsymbol{x}_e + \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0]) \\ &= \boldsymbol{x}[k] - \boldsymbol{x}_e - \boldsymbol{x}[0] + \mathbf{P}_{\mathcal{S}}\boldsymbol{x}[0] \\ &= \mathbf{P}_{\mathcal{S}}(\boldsymbol{x}[k]-\boldsymbol{x}[0]) - \boldsymbol{x}_e + \mathbf{P}_{\mathcal{S}}\boldsymbol{x}[0] \\ &= \mathbf{P}_{\mathcal{S}}\boldsymbol{x}[k] - \boldsymbol{x}_e. \end{aligned} \tag{18}$$

Note that this $N \times 1$ vector lies in the signal space $\mathcal{S}$, and can therefore be written as in (12) for an $M \times 1$ error vector $\boldsymbol{\delta}[k]$. In the absence of noise,

$$\boldsymbol{s}[k] - \boldsymbol{b} = \boldsymbol{H}^H\boldsymbol{x}[k] - \boldsymbol{H}^H\boldsymbol{x}_\infty = \boldsymbol{H}^H\boldsymbol{\Delta}[k] = \boldsymbol{H}^H\boldsymbol{H}\boldsymbol{\delta}[k]$$

Thus (5) becomes (11), or in terms of the $M \times 1$ error vector $\boldsymbol{\delta}[k]$ as in (13). Under (14) and Assumption 1, all eigenvalues of $\boldsymbol{I}_M - \mu\boldsymbol{H}^H\boldsymbol{H}$ are strictly smaller than one in magnitude. Hence $\boldsymbol{\delta}[k]$ exponentially converges to 0. Due to (12) and Assumption 1, so must $\boldsymbol{\Delta}[k]$. Then (18) proves that $\mathbf{P}_{\mathcal{S}}\boldsymbol{x}[k]$ exponentially converges to $\boldsymbol{x}_e$ and hence to $\boldsymbol{x}_\infty$ in (15).

Finally, we derive (16) for the limiting weight. The projection of any N-vector $\boldsymbol{z}$ onto the affine subspace $\mathcal{H} = \boldsymbol{x}_e + \mathcal{S}^{\perp}$ is $\boldsymbol{x}_e + \boldsymbol{y}$, where $\boldsymbol{y} \in \mathcal{S}^{\perp}$ minimizes the distance of $\boldsymbol{z}$ from $\mathcal{H}$:

$$\begin{aligned} &\min_{\boldsymbol{y}\in\mathcal{S}^{\perp}} ||\boldsymbol{z} - (\boldsymbol{x}_e + \boldsymbol{y})||^2 \\ &\qquad = \min_{\boldsymbol{y}\in\mathcal{S}^{\perp}} ||\mathbf{P}_{\mathcal{S}}\boldsymbol{z} - \boldsymbol{x}_e||^2 + ||\mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{z} - \boldsymbol{y}||^2 \end{aligned}$$

where we have decomposed the squared distance across $\mathcal{S}$ and $\mathcal{S}^{\perp}$. We cannot change the first term on the right hand side, but can set the second term to zero by setting $\boldsymbol{y} = \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{z}$, so that

$$\text{Proj}_{\mathcal{H}}(\boldsymbol{z}) = \boldsymbol{x}_e + \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{z} \tag{19}$$

Plugging in $\boldsymbol{z} = \boldsymbol{x}[0]$ completes the proof. □

The proof of exponential convergence of LMS is much more complicated and requires that a sequence of regressors satisfy a persistent excitation (p.e.) condition, [31]. The eigenvalues of certain outer product sums of regressors play the role of λ_i in (14).

1) Design Prescription for Minimizing Transmit Power: A key implication of the theorem is that, as long as the initial condition $\boldsymbol{x}[0]$ is in the signal space (i.e., it can be written as $\boldsymbol{x}[0] = \boldsymbol{H}\boldsymbol{\eta}$ for some $M \times 1$ vector $\boldsymbol{\eta}$), the iterations converge to the power efficient (minimum norm) solution $\boldsymbol{x}_e$. To see this, substitute $\boldsymbol{x}[0] = \boldsymbol{H}\boldsymbol{\eta}$ into (15) and verify that $\boldsymbol{x}_\infty \equiv \boldsymbol{x}_e$. For example, the initialization $\boldsymbol{x}[0] = \mathbf{0}$, or to a spatial matched filter to one of the beam targets, say $\boldsymbol{x}[0] = \boldsymbol{h}_1$, guarantees convergence to the power efficient solution. When initialization in the signal space is not feasible, then leakage-type mechanisms can be introduced to dissipate the $\mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0]$ term that our present algorithm is unable to perturb. This is explored further in Section IV-C.

2) Effect of Linear Dependence: If Assumption 1 is violated (i.e., the channel vectors $\{\boldsymbol{h}_i\}$ are linearly dependent), then $\boldsymbol{b}$ may not be in the range space of $\boldsymbol{H}^H$. In that case $J(\boldsymbol{x})$ has a nonzero minimum. Nonetheless under (14), without noise the gradient asymptotically vanishes

$$\lim_{k\to\infty} \boldsymbol{H}(\boldsymbol{H}^H\boldsymbol{x}[k] - \boldsymbol{b}) = 0 \tag{20}$$

though $\boldsymbol{H}^H\boldsymbol{x}[k]$ need not converge to $\boldsymbol{b}$. As $J(\boldsymbol{x})$ is convex, the is a global minimum.

C. The Effect of Noise

We now extend the preceding arguments and analyze the impact of noise in the feedback:

$$\boldsymbol{s}[k] = \boldsymbol{H}^H\boldsymbol{x}[k] + \boldsymbol{w}[k], \quad \boldsymbol{w}[k] \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\boldsymbol{I}). \tag{21}$$

We characterize the means and covariances of the weight vectors $\{\boldsymbol{x}[k]\}$ in Theorem 2 below. The noisy version of (13) is used to show in the theorem that noise does not cause the $\{\boldsymbol{x}[k]\}$ to drift along the affine subspace $\mathcal{H}$, and the error covariance is bounded with a limit point.

We first note that even with noise, the update term in (5) lies in $\mathcal{S}$, hence we still have

$$\mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[k] \equiv \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0]$$

Define the error vectors $\boldsymbol{\Delta}[k]$ and $\boldsymbol{\delta}[k]$ as before, using (18) and (12). Then

$$\begin{aligned} \boldsymbol{s}[k] - \boldsymbol{b} &= \boldsymbol{H}^H\boldsymbol{x}[k] + \boldsymbol{w}[k] - \boldsymbol{H}^H\boldsymbol{x}_\infty = \boldsymbol{H}^H\boldsymbol{\Delta}[k] + \boldsymbol{w}[k] \\ &= \boldsymbol{H}^H\boldsymbol{H}\boldsymbol{\delta}[k] + \boldsymbol{w}[k] \end{aligned}$$

Thus we obtain,

$$\begin{aligned} \boldsymbol{\Delta}[k+1] &= \boldsymbol{\Delta}[k] - \mu\boldsymbol{H}\left(\boldsymbol{H}^H\boldsymbol{\Delta}[k] + \boldsymbol{w}[k]\right) \\ &= \left(\boldsymbol{I} - \mu\boldsymbol{H}\boldsymbol{H}^H\right)\boldsymbol{\Delta}[k] - \mu\boldsymbol{H}\boldsymbol{w}[k] \\ \boldsymbol{H}\boldsymbol{\delta}[k+1] &= \left(\boldsymbol{I}_N - \mu\boldsymbol{H}\boldsymbol{H}^H\right)\boldsymbol{H}\boldsymbol{\delta}[k] \\ &= \boldsymbol{H}\left(\boldsymbol{I}_M - \mu\boldsymbol{H}^H\boldsymbol{H}\right)\boldsymbol{\delta}[k] - \mu\boldsymbol{H}\boldsymbol{w}[k] \end{aligned} \tag{22}$$

As $\boldsymbol{H}$ has full column rank the last equation becomes:

$$\boldsymbol{\delta}[k+1] = \left(\boldsymbol{I}_M - \mu \boldsymbol{H}^H \boldsymbol{H}\right) \boldsymbol{\delta}[k] - \mu \boldsymbol{w}[k] \tag{23}$$

There are two key points to make about these equations. First with λ_i as in Theorem 1, the transition matrix in (23) is asymptotically stable, though that in (22) is not. More importantly, in both equations, the effect of noise is masked by the channel matrix $\boldsymbol{H}$, precluding the possibility of Brownian motion orthogonal to the signal space.

Theorem 2: Consider (5) under (2) and Assumption 1, with noisy feedback modeled as in (21).

Assume that the adaptation gain μ satisfies (14). Then the mean of the weight sequence converges to the same limit as in the noiseless setting:

$$\begin{aligned}\lim_{k\to\infty} E\left[\boldsymbol{x}[k]\right] &= \boldsymbol{x}_\infty = \boldsymbol{x}_e + \mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0] = \boldsymbol{x}_e + \boldsymbol{x}[0] - \mathbf{P}_{\mathcal{S}}\boldsymbol{x}[0] \\ &= \boldsymbol{x}_e + \boldsymbol{x}[0] - \boldsymbol{H}(\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H\boldsymbol{x}[0] \\ &= Proj_{\mathcal{H}}\left(\boldsymbol{x}[0]\right) \end{aligned}\tag{24}$$

The covariance

$$\boldsymbol{\Sigma}_{\boldsymbol{x}}[k] = E\left[\left(\boldsymbol{x}[k] - E[\boldsymbol{x}[k]]\right)\left(\boldsymbol{x}[k] - E[\boldsymbol{x}[k]]\right)^H\right]$$

converges to

$$\boldsymbol{\Sigma} = \lim_{k\to\infty} \boldsymbol{\Sigma}_{\boldsymbol{x}}[k] = \mu\sigma_w^2 \boldsymbol{H}(2\boldsymbol{H}^H\boldsymbol{H} - \mu(\boldsymbol{H}^H\boldsymbol{H})^2)^{-1}\boldsymbol{H}^H. \tag{25}$$

Proof: Observe (22) and (23) hold. Define the mean vectors $\boldsymbol{m}_{\boldsymbol{\Delta}}[k] = E\left[\boldsymbol{\Delta}[k]\right]$ and $\boldsymbol{m}_{\boldsymbol{\delta}}[k] = E\left[\boldsymbol{\delta}[k]\right]$ and the corresponding covariance matrices

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}[k] &= E\left[\left[\boldsymbol{\Delta}[k] - E[\boldsymbol{\Delta}[k]]\right]\left[\boldsymbol{\Delta}[k] - E[\boldsymbol{\Delta}[k]]\right]^H\right], \\ \boldsymbol{\Sigma}_{\boldsymbol{\delta}}[k] &= E\left[\left[\boldsymbol{\delta}[k] - E[\boldsymbol{\delta}[k]]\right]\left[\boldsymbol{\delta}[k] - E[\boldsymbol{\delta}[k]]\right]^H\right]\end{aligned}\tag{26}$$

Taking expectations on both sides of (22) and (23), it is easy to see that the means follow the same trajectories as in the noiseless setting, and therefore converge to zero under the assumptions of Theorem 1. Subtracting the means from (22 and (23) and taking outer products yields:

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}[k+1] &= \left(\boldsymbol{I} - \mu\boldsymbol{H}\boldsymbol{H}^H\right)\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}[k]\left(\boldsymbol{I} - \mu\boldsymbol{H}\boldsymbol{H}^H\right) \\ &\quad + \mu^2\sigma_w^2\boldsymbol{H}\boldsymbol{H}^H \end{aligned}\tag{27}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\delta}}[k+1] = \left(\boldsymbol{I} - \mu\boldsymbol{H}^H\boldsymbol{H}\right)\boldsymbol{\Sigma}_{\boldsymbol{\delta}}[k]\left(\boldsymbol{I} - \mu\boldsymbol{H}^H\boldsymbol{H}\right) + \mu^2\sigma_w^2 \tag{28}$$

As the magnitudes of all eigenvalues of $\boldsymbol{I}_M - \mu\boldsymbol{H}^H\boldsymbol{H}$ are less than 1, the limiting covariance $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$ in (28) exists and is the unique solution of the Lyapunov equation, [49],

$$\boldsymbol{\Sigma}_{\boldsymbol{\delta}} = \left(\boldsymbol{I} - \mu\boldsymbol{H}^H\boldsymbol{H}\right)\boldsymbol{\Sigma}_{\boldsymbol{\delta}}\left(\boldsymbol{I} - \mu\boldsymbol{H}^H\boldsymbol{H}\right) + \mu^2\sigma_w^2\boldsymbol{I}. \tag{29}$$

We verify below that this solution is in fact:

$$\boldsymbol{\Sigma}_{\boldsymbol{\delta}} = \mu\sigma_w^2(2\boldsymbol{H}^H\boldsymbol{H} - \mu(\boldsymbol{H}^H\boldsymbol{H})^2)^{-1}. \tag{30}$$

Set $\boldsymbol{A} = \boldsymbol{H}^H\boldsymbol{H}$, and note that all quantities like $(a_1\boldsymbol{A}+a_2 I)^{-1}$ commute with all polynomials in $\boldsymbol{A}$, if the a_i are scalar. Then to verify (29), substitute $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}/\sigma_w^2 = \mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)^{-1}$ into $(\boldsymbol{I} - \mu A)\,\boldsymbol{\Sigma}_{\boldsymbol{\delta}}\,(\boldsymbol{I} - \mu A)/\sigma_w^2 + \mu^2\boldsymbol{I}$ to see if it equals $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}/\sigma_w^2$. Indeed

$$\begin{aligned}&(\boldsymbol{I} - \mu\boldsymbol{A})\,\mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)^{-1}(\boldsymbol{I} - \mu\boldsymbol{A}) + \mu^2\boldsymbol{I} \\ &\quad= \mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)^{-1}\left(\boldsymbol{I} - 2\mu\boldsymbol{A} + \mu^2\boldsymbol{A}^2\right) + \mu^2\boldsymbol{I} \\ &\quad= \mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)^{-1}\left(\boldsymbol{I} - \mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)\right) + \mu^2\boldsymbol{I} \\ &\quad= \mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)^{-1} - \mu^2\boldsymbol{I} + \mu^2\sigma_w^2\boldsymbol{I} \\ &\quad= \mu(2\boldsymbol{A} - \mu\boldsymbol{A}^2)^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\delta}}/\sigma_w^2\end{aligned}$$

This proves (30). From (12) we have $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}[k] = \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{\delta}}[k]\boldsymbol{H}^H$. Plugging in (30) yields (25). □

Theorem 2 shows that, even with noise in the feedback, the mean of the weight vector $\boldsymbol{x}[k]$ converges to the same limit as in the noiseless setting, and the limiting covariance is finite. It is worth highlighting the structure of the error revealed through the proof. As before, the N-dimensional error vector $\Delta[k] = \boldsymbol{x}[k] - \boldsymbol{x}_\infty$ is constrained to the M-dimensional signal space $\mathcal{S}$, and can therefore be described in terms of an M-dimensional error vector $\boldsymbol{\delta}[k]$. The limiting M-dimensional covariance $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$ is positive definite under our assumptions, whereas the limiting N-dimensional covariance $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}} = \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{\delta}}\boldsymbol{H}^H$ is positive semidefinite, with M positive eigenvalues, and $N - M$ zero eigenvalues.

We also emphasize that the component of $\boldsymbol{x}[k]$ orthogonal to the signal space remains fixed at $\mathbf{P}_{\mathcal{S}}^{\perp}\boldsymbol{x}[0]$ throughout the iterations, whether or not there is noise in the feedback. *The implication of this is that noise in the feedback cannot cause drift in $\boldsymbol{x}[k]$.* This is in stark contrast to standard adaptive filtering, e.g. with LMS, [43], where noise components orthogonal to the signal space induce drift, leading to *unbounded residual covariance*. This happens in LMS when the p.e. condition alluded to after Theorem 1 is violated and is typically remedied by mechanisms such as tap leakage. The key difference in our setting is that the effect of noise at the *receivers,* when used for *transmitter adaptation,* is seen through the channel matrix $\boldsymbol{H}$, and hence is restricted to the M-dimensional signal space. A comparable rigorous characterization of the residual error covariance is not available for LMS, [39], without certain simplifying independence assumptions [38], [50] that approximately hold for sufficiently small μ.

We note that, even if Assumption 1 is violated (i.e., the channel vectors $\{\boldsymbol{h}_i\}$ are not linearly independent), noise does not induce drift. To see this, note that drift must occur along the null space of $\boldsymbol{H}^H$. This is so as should $\boldsymbol{x} = \boldsymbol{x}^*$ minimize $J(\boldsymbol{x})$ and $\boldsymbol{\eta}$ be in the null space of $\boldsymbol{H}^H$ then as $\boldsymbol{H}^H\boldsymbol{\eta} = 0$, $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{\eta}$ must also minimize $J(\boldsymbol{x})$. Now for any such $\boldsymbol{\eta}$ there holds:

$$\begin{aligned}\boldsymbol{\eta}^H\boldsymbol{x}[k+1] &= \boldsymbol{\eta}^H\left(\boldsymbol{x}[k] - \mu\boldsymbol{H}(\boldsymbol{s}[k] - \boldsymbol{b})\right) \\ &= \boldsymbol{\eta}^H\left(\boldsymbol{x}[k] - \mu\boldsymbol{H}(\boldsymbol{H}^H\boldsymbol{x}[k] + \boldsymbol{w}[k] - \boldsymbol{b})\right) \\ &= \boldsymbol{\eta}^H\boldsymbol{x}[k].\end{aligned}$$

Thus, the noise has no impact along the null space of $\boldsymbol{H}^H$. This argument can be formalized further to show that even when $\boldsymbol{H}$ does not have full column rank, noise does not induce drift.

D. Convergence Speed vs. Residual Variance

To study convergence speed and residual variance as the number of transmitters N increases, but the number of receivers M is fixed, we first quantify the effect of λ_i and the selection of μ.

Define $\mathbf{\Lambda} =$ diag $(\lambda_1, \cdots, \lambda_m)$ as the eigenvalue matrix of $\boldsymbol{H}^H\boldsymbol{H}$ under the ordering $\lambda_i \geq \lambda_{i+1} > 0$. Then, with $\boldsymbol{U} \in \mathbb{C}^{N\times N}$, $\boldsymbol{V} \in \mathbb{C}^{M\times M}$ unitary matrices, one has

$$\boldsymbol{H} = \boldsymbol{U}\begin{bmatrix}\mathbf{\Lambda}^{\frac{1}{2}} & 0\end{bmatrix}^T \boldsymbol{V}. \tag{31}$$

Then from (12) and (13) one readily obtains:

$$\mathbf{\Delta}[k] = \boldsymbol{U} \text{ diag } \left((I-\mu\mathbf{\Lambda})^k, 0\right)\boldsymbol{U}^H\mathbf{\Delta}[0]. \tag{32}$$

Thus, the convergence rate is constrained by the largest among $|1-\mu\lambda_i|$. Specifically $\|\mathbf{\Delta}[k]\| \leq |1-\mu\lambda_i|^k\|\mathbf{\Delta}[0]\|$. Subject to (14), the μ that minimizes the largest among $|1-\mu\lambda_i|$ is, [50],

$$\mu^* = \frac{2}{\lambda_1+\lambda_M}. \tag{33}$$

In this case, the largest value of $|1-\mu\lambda_i|$ is given by $-(1-\mu\lambda_1) = 1-\mu\lambda_M > 0$, yielding

$$\max_{i\in\{1,\cdots,M\}} |1-\mu\lambda_i| = \frac{C-1}{C+1} \equiv \left(1-\frac{2}{C+1}\right) \tag{34}$$

where $C = \frac{\lambda_1}{\lambda_M}$ is the condition number of $\boldsymbol{H}$. The convergence rate improves as C decreases (i.e., as the eigenvalue spread shrinks), and $C = 1$ (no spread) yields deadbeat one step convergence. Note (33) holds also for LMS for suitably defined λ_1 and λ_M, [50]. The recursive least squares algorithm (RLS) is not exponentially convergent in the noise free case, [35], without a forgetting factor. With forgetting factor, (33) applies with λ_i defined similarly to LMS.

A more conservative choice,

$$\mu = \frac{1}{\lambda_1}, \tag{35}$$

ensures that $0 < 1-\mu\lambda_i \leq 1$ for all i. In this case, the convergence rate is given by $1-\mu\lambda_M = 1-\frac{\lambda_M}{\lambda_1} = 1-\frac{1}{C}$. As with the optimal choice, the convergence rate improves with declining C and achieves deadbeat status when $C = 1$. The choice of $\mu = \mu^*$, however, may lead to a larger residual variance. To see this observe that (25) and (31) yield:

$$\mathbf{\Sigma} = \mu\sigma_w^2\boldsymbol{U} \text{ diag } \left((2I-\mu\mathbf{\Lambda})^{-1}, 0\right)\boldsymbol{U}^H. \tag{36}$$

Of course a smaller μ results in a smaller steady state covariance. Under (33) there obtains

$$\frac{\mu^*}{2-\mu^*\lambda_i} \leq \frac{\mu^*}{2-\mu^*\lambda_1} = \frac{2}{2(\lambda_1+\lambda_M)-2\lambda_1} = \frac{1}{\lambda_M}.$$

Thus,

$$0 \leq \mathbf{\Sigma} \leq \frac{\sigma_w^2}{\lambda_M}I. \tag{37}$$

On the other hand if $\mu\lambda_1 \leq 1$, a condition that guarantees convergence, but may not be satisfied by $\mu = \mu^*$, one obtains the smaller bound of

$$\frac{\mu}{2-\mu\lambda_i} \leq \frac{\mu}{2-\mu\lambda_1} \leq \mu = \frac{1}{\lambda_1},$$

leading to

$$0 \leq \mathbf{\Sigma} \leq \frac{\sigma_w^2}{\lambda_1}I. \tag{38}$$

IV. Behavior With Large N: Scalability and Asymptotics

We now study the convergence rate, the power efficient solution $\boldsymbol{x}_e$ and the noise performance of the JBNF algorithm as N, the number of transmitters, becomes large, with the total number of beam and null targets fixed at M. We introduce subscripts to explicitly denote dependence on N. For example, $\boldsymbol{H}_N$ is the corresponding channel matrix, the eigenvalues of $\boldsymbol{H}^H\boldsymbol{H}$, are $\lambda_{i,N}$.

A. Convergence Speed With Deterministic Channels

First suppose that μ_N is fixed at some value μ_0 such that $\mu_0\lambda_{1,N} < 1$. In this case, (14) is satisfied and convergence rate does not decline if $\mu_0\lambda_{M,N}$ does not decline with N. The channel matrix grows from N transmitters to $N+1$ as:

$$\boldsymbol{H}_{N+1} = \begin{pmatrix}\boldsymbol{H}_N \\ \boldsymbol{g}_{N+1}^H\end{pmatrix},$$
$$\boldsymbol{g}_{N+1} = [h_{N+1,1}^*,\ h_{N+1,2}^*,\ \ldots,\ h_{N+1,M}^*]^\top. \tag{39}$$

We then have the following result.

Theorem 3: Consider the family of JBNF algorithms (5) with an increasing number of transmitters $N > M$ while keeping the step size μ_N fixed at μ_0. Then the convergence rate of the algorithm is nondecreasing in the number of transmitters N provided $\mu_0\lambda_{1,N} < 1$.

Proof: As μ is fixed at μ_0 and $\mu_0\lambda_{1,N} < 1$, it suffices to show that for all $l < N$ (a) $\mu_0\lambda_{M,N} < 1$, and (b) $\mu_0\lambda_{M,l}$ is nondecreasing in l. Observe from (39) that $H_{l+1}^H H_{l+1} = H_l^H H_l + g_{L+1}g_{L+1}^H$. Then (a) and (b) follow as for any pair of Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\lambda_{\min}(\boldsymbol{A}+\boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{B})$ and $\lambda_{\max}(\boldsymbol{A}+\boldsymbol{B}) \leq \lambda_{\max}(\boldsymbol{A}) + \lambda_{\max}(\boldsymbol{B})$. □

Note that this result does not depend on any specific channel model and holds for all fading and LoS channels. However, using a fixed step-size μ_0 while increasing the number of transmitters is too restrictive: for large N, the value of μ_0 required is unnecessarily small. Thus, we now consider the setting when μ_N is allowed to change with N. Recall from Section III-D that the convergence rate is bounded from below if and only if the condition number $C_N = \frac{\lambda_{1,N}}{\lambda_{M,N}}$. is upper bounded. To this end we provide a sufficient condition on the $\boldsymbol{h}_i$ that assures the uniform boundedness of C_N. The condition is similar to the p.e. condition described earler, [30]–[32]. It requires that channel submatrices for each new batch of transmitters should be well conditioned. We show later that i.i.d. complex Gaussian channels asymptotically meet the condition.

Theorem 4: Suppose $\boldsymbol{H}_N \in \mathbb{C}^{N\times M}$ and $\boldsymbol{g}_i \in \mathbb{C}^M$ are as in (39). Define $\lambda_{i,N}$ as the eigenvalues of $\boldsymbol{H}_N^H\boldsymbol{H}_N$ and C_N as above. Suppose there exist $0 < \alpha_l$ and an L such that for all i

and g_i defined in (39),

$$0 < \alpha_1 I \leq \sum_{m=i}^{i+L} \boldsymbol{g}_m \boldsymbol{g}_m^H \leq \alpha_2 I.$$

Then for all $N \geq L$, C_N is uniformly bounded in N.

Proof: Follows from the eigen-inequalities in the proof of Theorem 3 and the fact that $\boldsymbol{H}_N^H \boldsymbol{H}_N = \sum_{m=1}^N \boldsymbol{g}_m \boldsymbol{g}_m^H;$. □

B. Asymptotics With Rayleigh Fading Channels

We now derive a variety of results for large N, assuming i.i.d. (across all transmitter-receiver pairs) complex Gaussian channels. All draw upon the following result from [37].

Theorem 5: Suppose the channel coefficients $h_{ij} \sim \mathcal{CN}(0,1)$, $i \in \{1 \ldots N\}$, $j \in \{1 \ldots M\}$ and are i.i.d. Then for any given M, the condition number C_N of the matrix $\boldsymbol{H}_N^H \boldsymbol{H}_N$ satisfies $\lim_{N\to\infty} C_N = 1$ with probability one. Further, with probability one, there holds

$$\lim_{N\to\infty} \frac{\lambda_{1,N}}{N} = \lim_{N\to\infty} \frac{\lambda_{M,N}}{N} = 1 \tag{40}$$

Referring to the discussion in Section III-D, this implies that the optimal choice (33) and the conservative choice (35) of μ_N are asymptotically equivalent, and that we asymptotically obtain arbitrarily fast convergence, both with probability one as $N \to \infty$. Further, the covariance bounds (37) and (38) imply that the covariance tends to zero. These results are summarized below.

Theorem 6: Consider $\boldsymbol{x}_N[k+1] = \boldsymbol{x}_N[k] - \mu_N \boldsymbol{H}_N(\boldsymbol{s}[k]-\boldsymbol{b})$ and suppose the conditions of Theorem 5 hold. Then there exists a sequence of μ_N such that without noise convergence to $\boldsymbol{x}_{N,\infty}$ occurs arbitrarily fast, and with noise $\lim_{N\to\infty} \boldsymbol{\Sigma}_N = 0$, with probability 1.

Finally, recall that any initialization in the range space of $\boldsymbol{H}_N$ causes convergence to the power efficient solution $\boldsymbol{x}_{e,N}$ in the noise free case. In the presence of noise, convergence to the same point occurs in the mean. In view of Proposition 5 and (8) there holds:

$$\lim_{N\to\infty} \boldsymbol{x}_{e,N} = \boldsymbol{H}_N(\boldsymbol{H}_N^H \boldsymbol{H}_N)^{-1} \boldsymbol{b}_N = 0.$$

Further $\boldsymbol{H}_N(\boldsymbol{H}_N^H \boldsymbol{H})_N^{-1} \boldsymbol{H}_N^H = \boldsymbol{U}_N \begin{bmatrix} I_N & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{U}_N^H$. Thus as, $\boldsymbol{U}_N$ is unitary, for every N, we have $\left\| \left(I - \boldsymbol{H}_N(\boldsymbol{H}_N^H \boldsymbol{H})_N^{-1} \boldsymbol{H}_N^H \right) \boldsymbol{x}[0] \right\| \leq \|\boldsymbol{x}[0]\|$. Thus, one obtains with probability one that $\lim_{N\to\infty} \left\| \mathrm{Proj}_{\mathcal{H}_N}(\boldsymbol{x}[0]) \right\| \leq \|\boldsymbol{x}[0]\|$, and the following.

Theorem 7: Suppose the conditions of Theorem 6 hold with $\left|1 - \mu_N \lambda_{i,N}\right| < 1$ for all i. Then in the noise free case for every $\boldsymbol{x}_N[0]$ there holds with probability one, $\lim_{N\to\infty} \|\boldsymbol{x}_{N,\infty}\| \leq \|\boldsymbol{x}_N[0]\|$. Further when $\boldsymbol{x}_N[0]$ is in the range space of $\boldsymbol{H}_N$ then with probability one, $\lim_{N\to\infty} \boldsymbol{x}_{N,\infty} = 0$.

To summarize, for channels that are i.i.d. $\mathcal{CN}(0,1)$, we have established the following results with probability one as N goes to infinity. (i) Convergence is arbitrarily fast. (ii) Residual variance goes to zero. (iii) Initialization in the signal space drives the steady state transmit power to zero. (iv) Regardless of initialization the steady state transmit power is no greater than the initial transmit power. (v) In the presence of noise the last two occur in the mean.

Items (ii), (iii) and (v) together demonstrate the following. As N tends to infinity, should one initialize in the signal space using e.g. the design prescription on Section III-B, then even with noise the limit point approaches almost surely, a *zero transmit power solution in the mean with zero covariance.*

C. Leakage to Minimize the Total Transmit Power

Initialization in the signal space ensures the attainment of the power efficient solution in the mean. But, such initialization may not always be feasible; e.g. if the channel matrix changes, a weight vector that was previously in the subspace $\mathcal{H}$ may no longer be in it. The introduction of *leakage*, a popular device both in adaptive filtering and control [44]–[46] can cope with this.

Leakage involves the addition of a *penalty term* proportional to the total transmit power to the objective function in (4) to get a new objective function: $J_2(\boldsymbol{x}) = \left\| \boldsymbol{H}^H \boldsymbol{x} - \boldsymbol{b} \right\|^2 + \alpha \|\boldsymbol{x}\|^2$ where $\alpha > 0$ is a constant that can be chosen to penalize power inefficiency. This leads to a distributed gradient search implementation, like (5):

$$\boldsymbol{x}[k+1] = (1 - \mu\alpha)\boldsymbol{x}[k] - \mu \boldsymbol{H}(\boldsymbol{s}[k] - \boldsymbol{b}). \tag{41}$$

This achieves the JBNF solution with the *minimum total transmit power* for arbitrary $\boldsymbol{x}[0]$. However, it has a limitation: One can no longer make convergence to be arbitrarily fast as N increases to infinity. Judicious choice of α can be used to gain the benefits of power minimization without compromising convergence speed.

V. Simulation Results

We consider a JBNF system with $N = 20$ transmitters and $M = 5$ receivers of which $M_1 = 2$ are beam targets and the remaining 3 receivers are null targets. All channel gains are modeled as i.i.d. $\sim \mathcal{CN}(0,1)$, and the noise level is taken to be -40 dB at each receiver. We encode the real and imaginary parts of $s[k]$ into double precision floating point numbers with each requiring 64 bits for a total feedback of 128 bits per iteration.

Fig. 4 depicts the variation of the cost function as well as the individual received signal levels at each receiver with $\boldsymbol{x}[0] = 0$. Within about 40 iterations, the cost function as well as the power levels at the null targets converged to a level close to the noise floor of -40 dB. The convergence at the beam targets is even faster. The apparent lack smoothness of the null plots is an artifact of the logarithmic scale of the plots. In fact, the largest fluctuations are about -30 dB (i.e. a factor of a thousand) smaller than the beam power.

If the weights are initialized randomly from a complex Gaussian distribution on the other hand, we expect a non-zero constant component orthogonal to the signal space that leads to wasted transmit power. This is confirmed by Fig. 5-(a) showing that the transmit power does not converge to that for the power efficient solution $\boldsymbol{x}_e$. For zero initialization of weights, Fig. 5-(b) shows that $\boldsymbol{x}_e$ is reached in 50 iterations.

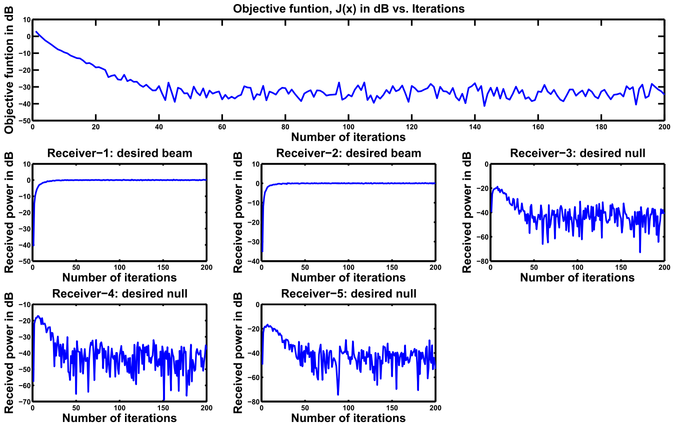
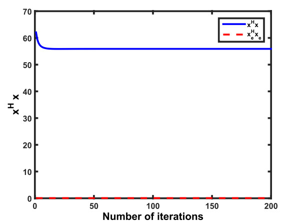
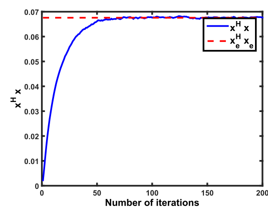


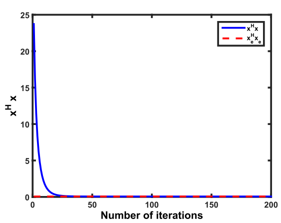
Fig. 4. Convergence of JBNF algorithm with initialization of transmit weights as zeros.



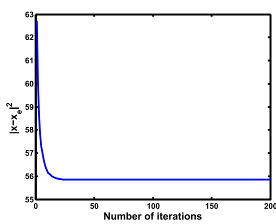
(a) random initialization.



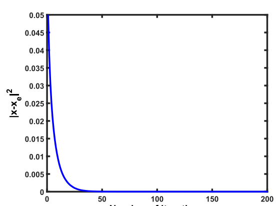
(b) zero initialization.



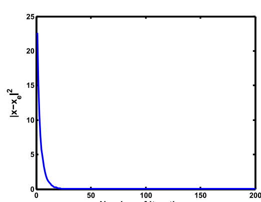
(c) linear combination initialization.



(d) random initialization.



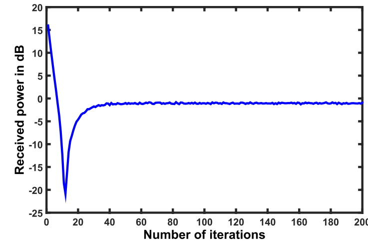
(e) zero initialization.



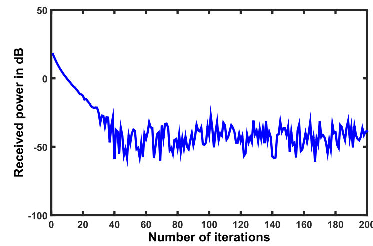
(f) linear combination initialization.

Fig. 5. In (a,b,c), blue line represents the total transmit power, red dashed line represents the power corresponding to power efficient solution. (d,e,f) represent the deviation of transmit weights from optimal weights under the JBNF algorithm.

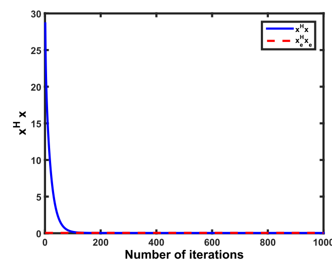
Per our design prescription given after Theorem 1 Fig. 5-(c),(f) initializes its weights $\boldsymbol{x}[0]$ to be a linear combination of the channel vectors to the beam targets i.e. $\boldsymbol{x}[0] \doteq a_1\boldsymbol{h}_1 + a_2\boldsymbol{h}_2$ for randomly chosen scalar constants a_1, a_2. Note that the transmitter i's initial weight is $x_i[0] \doteq a_1 h_{i1} + a_2 h_{i2}$, which only requires knowledge of transmitters i's channel gains h_{i1}, h_{i2}. Converges to $\boldsymbol{x}_e$ is even faster, within about 20 iterations under this initialization. Fig. 6 shows the JBNF algorithm with leakage with random $\boldsymbol{x}[0]$ and the penalty parameter $\alpha = 5$ chosen by a simple trial and error procedure to achieve a



(a) Beam Power.



(b) Null Power.



(c) Total Transmit Power.

Fig. 6. Convergence of JBNF algorithm with leakage. In (c), blue line represents the total transmit power, red dashed line represents the power corresponding to power efficient solution.

good tradeoff between convergence and power minimization; it can be seen that the algorithm effectively achieves the power efficient solution even though $x[0]$ is not in the signal space.

Next, we investigate constraining complex amplitudes rather than powers at the beam targets. If $\boldsymbol{b}_{beam}$ were aligned to the eigenvector corresponding to the minimum eigenvalue ρ_{min} of $\boldsymbol{K} = (\tilde{\boldsymbol{H}}_{beam}^H \tilde{\boldsymbol{H}}_{beam})^{-1}$, then we would obtain a minimum transmit power $P_{TX}^* = \boldsymbol{b}_{beam}^H \boldsymbol{K} \boldsymbol{b}_{beam} = \rho_{min} \|\boldsymbol{b}_{beam}\|^2$. This gives a very conservative lower bound, against which we compare the minimum power P_{TX} obtained for our fixed choice $\boldsymbol{b}_{beam} = \mathbf{1}$. As N increases, the eigenvalue spread of $\boldsymbol{K}$ declines, so that P_{TX} should approach the lower bound. This is confirmed by Fig. 7 which plots the CDF of the ratio $\frac{P_{TX}}{P_{TX}^*}$ (for i.i.d. complex Gaussian channel realizations) for different values of N, for $M = 5$ receivers and $M_1 = 2$ beam targets. For 90% of the channel realizations, the ratio is at most 2.25, 1.35 and 1.14 for $N = 20$, 100, and $N = 500$, respectively. Given the N-fold gain in power efficiency from coherent beamforming, and the rapid convergence, our quadratic approach is clearly an attractive design choice even for relatively small values of N, with the power penalty relative to the lower bound vanishing as N increases.

Fig. 8 considers the performance of JBNF with time-varying channels, modeled by a first order autoregressive Gaussian process. Specifically we model each channel

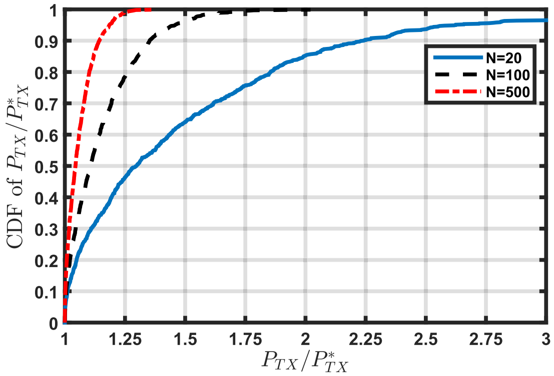


Fig. 7. Cumulative Distribution Function of transmit power corresponding to power efficient solution.

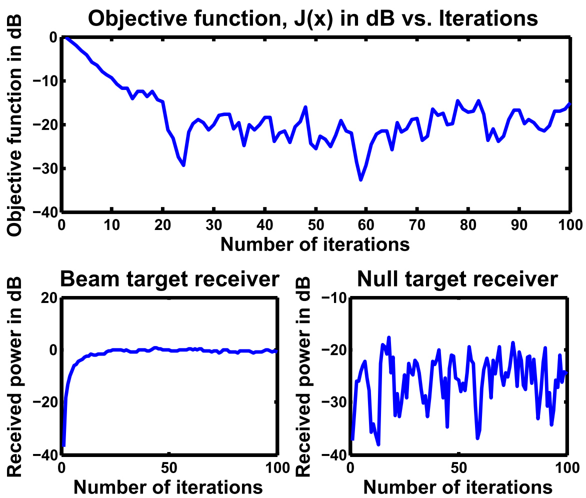
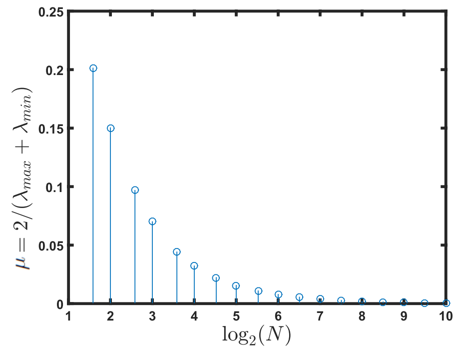


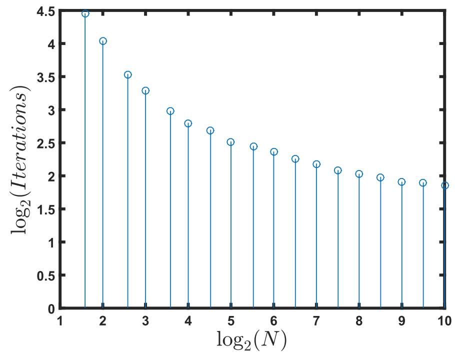
Fig. 8. JBNF algorithm with a time-varying channel.

gain h_{ij} as varying over the iterations k of the JBNF algorithm as $h_{ij}[k+1] = \sqrt{1-\alpha^2}h_{ij}[k] + \alpha w_{ij}[k]$, where w_{ij} are modeled as iid white complex Gaussian processes i.e. $w_{ij}[k] \sim CN(0,1)$. We have $N = 10$, $M = 2$ with one beam target and one null target. The noise level is -40 dB at each receiver, and the channel time-variation rate parameter $\alpha = 0.1$ which makes the channel after 100 iterations to have a 0.6 correlation with the initial channel. The JBNF algorithm adapts to channel variations effectively and converges within approximately 20 iterations which is comparable to the convergence time for static channels. However, it should be noted that variations in the channel inherently limit the quality of the nulls that can be achieved; in Fig. 8, the null power is on average about 10 dB higher than the noise level, and this penalty increases when channel variation rate increases. The JBNF algorithm update rate must be set proportionally with the channel variation rate to allow the array weights to track the varying channels.

Fig. 9 shows improved convergence rate, and thus scalability, as the number of transmitters N increases. The step-size parameter μ varied as per (33), the channels are i.i.d. complex Gaussian, and there is no noise. We set $M = 2$, one beam and one null target. The number of iterations required for the algorithm to drive the cost function to -60 dB decreases with N as predicted.



(a) μ for different number of transmitters (N)



(b) Iterations required for convergence to -60 dB

Fig. 9. Rate of Convergence of JBNF algorithm with different number of transmitters and $M = 2$.

VI. Conclusion

Our algorithm achieves simultaneous beams and nulls in a quadratic framework guaranteeing rapid convergence. It scales to large transmit arrays, as it only requires aggregate broadcast feedback from the receivers and the convergence rate actually improves with the number of transmitters. This work opens up many interesting questions for further inquiry. A non-asymptotic exploration of scalability and generalization to non-Rayleigh channel statistics is one open problem. Studying the effects of quantized feedback, partial CSI and other practical constraints is another interesting topic for future work.

References

[1] A. Kumar, R. Mudumbai, and S. Dasgupta, "Scalable algorithms for joint beam and null-forming using distributed antenna arrays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4042–4047.

[2] L. Dong, Z. Han, A. Petropulu, and H. V. Poor, "Cooperative jamming for wireless physical layer security," in *Proc. IEEE/SP 15th Workshop Statist. Signal Process. (SSP)*, Aug./Sep. 2009, pp. 417–420.

[3] Y. Noam and A. J. Goldsmith, "Blind null-space learning for MIMO underlay cognitive radio with primary user interference adaptation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1722–1734, Apr. 2013.

[4] A. Manolakos, Y. Noam, and A. J. Goldsmith, "Null space learning in cooperative MIMO cellular networks using interference feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3961–3977, Jul. 2015.

[5] R. Mudumbai, U. Madhow, R. Brown, and P. Bidigare, "DSP-centric algorithms for distributed transmit beamforming," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2011, pp. 93–98.

[6] P. Marsch and G. P. Fettweis, Eds., *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[7] D. R. Brown, P. Bidigare, and U. Madhow, "Receiver-coordinated distributed transmit beamforming with kinematic tracking," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 5209–5212.

[8] D. R. Brown and R. David, "Receiver-coordinated distributed transmit nullforming with local and unified tracking," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 1160–1164.

[9] A. Ozgur, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[10] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[11] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[12] R. Mudumbai, D. R. B. Iii, U. Madhow, and H. V. Poor, "Distributed transmit beamforming: Challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 102–110, Feb. 2009.

[13] M. Seo, M. Rodwell, and U. Madhow, "A feedback-based distributed phased array technique and its application to 60-GHz wireless sensor network," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2008, pp. 683–686.

[14] M. M. Rahman, H. E. Baidoo-Williams, R. Mudumbai, and S. Dasgupta, "Fully wireless implementation of distributed beamforming on a software-defined radio platform," in *Proc. 11th Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, New York, NY, USA, 2012, pp. 305–316. [Online]. Available: http://doi.acm.org/10.1145/2185677.2185745

[15] F. Quitin, M. M. U. Rahman, R. Mudumbai, and U. Madhow, "A scalable architecture for distributed transmit beamforming with commodity radios: Design and proof of concept," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1418–1428, Mar. 2013.

[16] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, May 2007.

[17] T. P. Bidigare *et al.*, "Wideband distributed transmit beamforming using channel reciprocity and relative calibration," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2015, pp. 271–275.

[18] D. R. Brown, U. Madhow, P. Bidigare, and S. Dasgupta, "Receiver-coordinated distributed transmit nullforming with channel state uncertainty," in *Proc. 46th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2012, pp. 1–6.

[19] D. R. Brown, P. Bidigare, S. Dasgupta, and U. Madhow, "Receiver-coordinated zero-forcing distributed transmit nullforming," in *Proc. IEEE Statist. Signal Process. Workshop (SSP)*, Aug. 2012, pp. 269–272.

[20] A. Kumar *et al.*, "A scalable feedback mechanism for distributed nullforming with phase-only adaptation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 1, pp. 58–70, Mar. 2015, doi: 10.1109/TSIPN.2015.2442921.

[21] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Scalable feedback control for distributed beamforming in sensor networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Adelaide, SA, Australia, Sep. 2005, pp. 137–141.

[22] J. Elson, L. Girod, and D. Estrin, "Fine-grained network time synchronization using reference broadcasts," *ACM SIGOPS Oper. Syst. Rev.*, vol. 36, no. SI, pp. 147–163, Dec. 2002.

[23] H. Rahul, H. Hassanieh, and D. Katabi, "Sourcesync: A distributed wireless architecture for exploiting sender diversity," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 171–182, 2010.

[24] R. Mudumbai, P. Bidigare, S. Pruessing, S. Dasgupta, M. Oyarzun, and D. Raeman, "Scalable feedback algorithms for distributed transmit beamforming in wireless networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 5213–5216.

[25] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. ISSPA*, Aug. 2005, pp. 403–406.

[26] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[27] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[28] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, Oct. 2014.

[29] B. D. O. Anderson, "Exponential stability of linear equations arising in adaptive identification," *IEEE Trans. Autom. Control*, vol. AC-22, no. 1, pp. 83–88, Feb. 1970.

[30] S. Dasgupta, B. D. O. Anderson, and A. C. Tsoi, "Input conditions for continuous-time adaptive systems problems," *IEEE Trans. Autom. Control*, vol. 35, no. 1, pp. 78–82, Jan. 1990.

[31] B. D. O. Anderson and C. R. Johnson, Jr., "Exponential convergence of adaptive identification and control algorithms," *Automatica*, vol. 18, no. 1, pp. 1–13, Jan. 1982.

[32] S. Dasgupta, Y. Shrivastava, and G. Krenzer, "Persistent excitation in bilinear systems," *IEEE Trans. Autom. Control*, vol. 36, no. 3, pp. 305–313, Mar. 1991.

[33] S. Dasgupta, B. D. O. Anderson, and R. J. Kaye, "Identification of physical parameters in structured systems," *Automatica*, vol. 24, no. 2, pp. 217–225, Mar. 1988.

[34] B. D. O. Anderson *et al.*, *Stability of Adaptive Systems: Passivity and Averaging Analysis*. Cambridge, MA, USA: MIT Press, 1986.

[35] R. M. Johnstone and B. D. O. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor—Adaptive control," *Syst. Control Lett.*, vol. 2, no. 2, pp. 69–76, Aug. 1982.

[36] K. J. Åström and B. Wittenmark, "On self tuning regulators," *Automatica*, vol. 9, no. 2, pp. 185–199, Mar. 1973.

[37] A. Edelman, "Eigenvalues and condition numbers of random matrices," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, 1988.

[38] J. R. Treichler, C. R. Johnson, and M. G. Larimore, *Theory and Design of Adaptive Filters*. New York, NY, USA: Wiley, 1987.

[39] R. Bitmead, "Convergence in distribution of LMS-type adaptive parameter estimates," *IEEE Transactions on Automatic Control*, vol. AC-28, no. 1, pp. 54–60, Jan. 1983.

[40] P. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. New York, NY, USA: Marcel Dekker, 1995.

[41] S. Dasgupta and C. R. Johnson, Jr., "Some comments on the behavior of sign-sign adaptive identifiers," *Syst. Control Lett.*, vol. 7, no. 2, pp. 75–82, Apr. 1986.

[42] S. Dasgupta, E. Bai, and R. Tempo, "Anchored consensus in multiagent systems," in *Proc. 30th Chin. Control Conf.*, Jul. 2011, pp. 4916–4919.

[43] B. D. O. Anderson, "Adaptive systems, lack of persistency of excitation and bursting phenomena," *Automatica*, vol. 21, no. 3, pp. 247–258, May 1985.

[44] P. A. Ioannou and P. V. Kokotovic, "Instability analysis and improvement of robustness of adaptive control," *Automatica*, vol. 20, no. 5, pp. 583–594, Sep. 1984.

[45] G. J. Rey, R. R. Bitmead, and C. R. Johnson, "The dynamics of bursting in simple adaptive feedback systems with leakage," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 476–488, May 1991.

[46] W. Sethares, D. Lawrence, C. Johnson, and R. Bitmead, "Parameter drift in LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 4, pp. 868–879, Aug. 1986.

[47] D. G. Messerschmitt, "A geometric theory of intersymbol interference," *Bell Syst. Tech. J.*, vol. 52, no. 9, pp. 1483–1519, Nov. 1973.

[48] U. Madhow, *Fundamentals of Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[49] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[50] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 680–691, Jul. 1988.

Amy Kumar received the B.E. degree in electrical and electronics engineering from the Birla Institute of Technology, Mesra, India, in 2010. She is currently pursuing the Ph.D. degree in electrical and computer engineering with The University of Iowa, Iowa City, IA, USA. She was with the Research and Development of Maruti Suzuki India Ltd., Gurgaon, India, from 2010 to 2012.

Raghuraman Mudumbai (M'09) received the B.Tech. degree in electrical engineering from IIT Madras, Chennai, India, in 1998, the M.S. degree in electrical engineering from Polytechnic University, Brooklyn, NY, USA, in 2000, and the Ph.D. degree in electrical and computer engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2007. He was with Ericsson Telephone Company for two years. He is currently an Associate Professor of Electrical and Computer Engineering with The University of Iowa.

Soura Dasgupta (M'87–SM'93–F'98) was born in Calcutta, India, in 1959.

He received the B.E. degree (Hons.) in electrical engineering from the University of Queensland, Australia, in 1980, and the Ph.D. degree in systems engineering from Australian National University in 1985. He is currently a Professor of Electrical and Computer Engineering with The University of Iowa, Iowa City, IA, USA, and holds an appointment with the Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center), Jinan, China. His research interests are in controls, signal processing, and communications.

In 1981, he was a Junior Research Fellow with the Electronics and Communications Sciences Unit, Indian Statistical Institute, Calcutta, India. He has held visiting appointments with the University of Notre Dame, The University of Iowa, Universite Catholique de Louvain-La-Neuve, Belgium, Tata Consulting Services, Hyderabad, and Australian National University.

Dr. Dasgupta was a co-recipient of the Gullimen Cauer Award for the best paper published in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS in 1990 and 1991. In 2012, he received the University Iowa Collegiate Teaching Award and was selected by the graduating class for an award on excellence in teaching and commitment to student success. Since 2015, he has been a 1000 Talents Scholar in the People's Republic of China. He served as an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 1988 to 1991, the IEEE Control Systems Society Conference Editorial Board from 1998 to 2009, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II from 2004 to 2007. He is a past Presidential Faculty Fellow, a past Subject Editor of the *International Journal of Adaptive Control and Signal Processing*, and a member of the Editorial Board of the *EURASIP Journal of Wireless Communications*.

Upamanyu Madhow (S'86–M'90–SM'96–F'05) received the bachelor's degree in electrical engineering from IIT Kanpur, Kanpur, India, in 1985, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana–Champaign in 1990. He was a Research Scientist with Bell Communications Research, Morristown, NJ, USA, and a faculty member at the University of Illinois at Urbana–Champaign. He is currently a Professor of electrical and computer engineering with the University of California at Santa Barbara, Santa Barbara, CA, USA. His research interests broadly span communications, signal processing and networking, with current emphasis on millimeter wave communication, and on distributed and bio-inspired approaches to networking and inference. He is the author of the textbooks *Fundamentals of Digital Communication* (Cambridge University Press, 2008) and *Introduction to Communication Systems* (Cambridge University Press, 2014). He was a recipient of the 1996 NSF CAREER Award and co-recipient of the 2012 IEEE Marconi Prize Paper Award in wireless communications. He has served as an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON INFORMATION THEORY, and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

D. Richard Brown III (S'97–M'00–SM'09) received the B.S. and M.S. degrees from the University of Connecticut in 1992 and 1996, respectively, and the Ph.D. degree from Cornell University in 2000, all in electrical engineering. From 1992 to 1997, he was with General Electric Electrical Distribution and Control. He joined as a Faculty Member at the Worcester Polytechnic Institute, Worcester, MA, USA, in 2000. He was a Visiting Associate Professor with Princeton University from 2007 to 2008. Since 2016, He has been with the National Science Foundation as a Program Director in the Computing and Communication Foundations Division.